ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 8, Issue 6, December 2018

Overview of RGB-D SLAM map building methods

Tian Yu Zhu, Wei Wang, Tong Wang, Yihao Cui

Abstract*—Real-time mapping and positioning and navigation of robots have always been a hot topic which is the key to autonomous movement of robots in unknown and known environments. In SLAM field, the VSLAM technology based on RGB-D camera has become a hot research topic in the unmanned field. The environment information captured by the visual camera is used for real-time map construction and autonomous exploration and path planning are carried out according to the constructed map. This article mainly includes three parts: image feature extraction, map perception and map building and the combination of map building and depth learning algorithm. It describes RGB-D SLAM map building methods and discusses the research and development direction of SLAM map building navigation technology.

***Index Terms*—VSLAM; Map-Building; Map-Aware; RGB-D; Deep-Learning.**

I. INTRODUCTION

Robots and UAV have the functions of automatic navigation, environment exploration, mapping and path planning according to the environment information collected by sensing network at a location. These functions are called SLAM (simultaneous localization and mapping).SLAM robots have been used in robot vacuum cleaners and driverless vehicles. They are equipped with few sensors and can complete mapping and navigation in one duty cycle. The SLAM technology at early stage was based on 2D maps of single-threaded laser radar and acoustic molding. With the increasing function demand of robot, multi-dimension map models are needed for complex location environment.However, the acoustic modeling is easily disturbed by environment, and laser radar builds maps with rapid speed and wide range, other kinds of sensors need to be used for SLAM navigation. In recent years,3D map model reconstruction has become a research hot spot in the computer image processing.

Manuscript received: 23 November 2018
Manuscript received in revised form: 20 December 2018
Manuscript accepted: 05 January 2019
Manuscript Available online: 10 January 2019

It is for the processing end to rebuild the map model according to the environmental information collected by the image equipment; and even the environmental information is rather complex, the modeling work can also be completed.When3D map model reconstruction technology is introduced to SLAM mapping, the robot can carry deep camera or a multi-threaded laser radar to scan for the map.Then,the environment information needed for map reconstruction is obtained, and the environment information obtained by sensors is classified.Therefore,it is extremely important to change the traditional environment information-collecting method and information processing technology. Facing the complex and changing environment and the development trend of human-computer interaction, the multi-intelligent-device collaborative platform, which is more intelligent, can provide better service for people. By taking visual camera as the sensor, SLAM which based on machine vision technology and computer image processing technology is VSLAM (visual SLAM). VSLAM adopts binocular camera, monocular camera and other depth cameras, such as RGB-D camera. The collected RGB images and environmental depth information are then filtered, matched and reconstructed according to the machine vision algorithm for mapping.

This paper takes RGB-D camera as the primary sensing device of VSLAM and describes its mapping methods. As the best one of visual equipment, RGB-D camera is powerful and simple, and it can obtain the color information and depth information of environment at the same time. Taking KINECT for example, it can collect color images and environment deep information, and can sense surrounding environment information by optical imaging and digital filtering. Kinect is a RGB-D camera and a motion sensing camera produced by Microsoft for Xbox video game consoles, which is used for motion sensing, mode identification and collection of depth information in industry, entertainment and consumer electronics. Kinect, which has high acquisition rate, low price and wide user and platform, has been widely used in computer image processing and machine vision. In the past, laser radar was trended to be taken as main information collection sensing device in the field of SLAM. It had faster speed and higher accuracy than other sensors. However, because it needed to adopt multiple threaded equipment with high cost and is inferior to visual sensors in details, it was not used widely. Now, the visual sensor is taken as the mainstream equipment for information collection. It can be seen from above discussion that the procedure by using RGB-D as the primary sensing device of SLAM system and by using machine vision and computer visual as core algorithm has become the mainstream plan in relevant VSLAM fields. In traditional VSLAM visual navigation, the calibration technology of space and plane map points has adopted monocular or binocular cameras. Based on its characteristics, the RGB-D camera is superior to the former in obtaining depth information and environmental information. When a RGB-D camera(for example, Kinect) is used to collect surrounding environment information by taking the principle of red structured light and fright time as the information-collecting plan, its working principle is analogic with the working mode of laser radar. The RGB-D camera not only inherits the imaging characteristics of the optical camera, but also has the fast and high efficiency characteristic of the radar. Therefore, as the main sensing equipment of the VSLAM technology, the RGB-D SLAM technology has become the mainstream of the development. The mapping framework of RGB-D SLAM is shown in figure 1.Firstly, RGB images and environment depth information are collected by the RGB-D camera, and the 3D spatial coordinates of the descriptor are obtained by the analysis of the improved ORB algorithm on the feature points of the image. Then, the best optimal matching effect is obtained by the registration and matching filtering of descriptors based on coordinates. Next, it comes to the key frame extraction and the initial point cloud image mosaic process through conversion matrix. The image mosaic process improves the efficiency and robustness of the system by the optimized algorithm, and the optimization process is completed by visual odometer, back-end and loop detection. Finally, a complete 3 D color point cloud map is generated. This paper mainly includes image feature extraction, map environment awareness methods and mapping, and combination of map construction and deep learning algorithm, presents the RGB-D SLAM mapping method, and discusses the research development direction of SLAM mapping navigation technology.

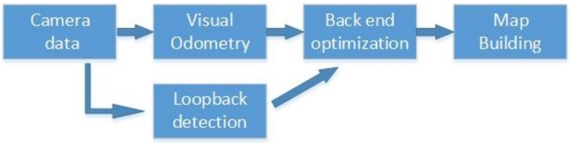

Fig.1.VSLAM Process Diagram

II. IMAGE FEATURE EXTRACTION

Image filtering is an early implementation program of image processing based on monocular vision VSLAM technology [7]-[10]. The extended Kalman filter algorithm is used based on monocular vision VSLAM. It is mainly to represent the map coordinates and depth information captured by sensors through state vector, and the uncertainty is summed up by probability model. When the probability model and observation model are built, the state vector and the information collected by the sensors are processed through state vectors to obtain mean value and variance for improvement, which is the process of extended Kalman filter. Extended Kalman filter is with linear computation, the use of extended Kalman filter algorithm in VSLAM system will lead to uncertainty problem. In reference [5]-[7], an unscented Kalman filter algorithm is introduced. In order to solve the linearization problem in extended Kalman filter algorithm, another method is introduced at the same time-the lossless Kalman filtering algorithm, which is combined with monocular vision. However, the lossless Kalman filtering algorithm increases the computation complexity but solves the uncertainty problem of system uncertainty. In reference [14]-[15], a method of combining particle filter with monocular vision positioning is introduced, which is helpful to the attitude motion calculation of camera. The summary of VSLAM based filtering method is shown in Fig. 2.As the VSLAM system using filtering algorithm will be affected by linear computation and the computational complexity is high, the scheme of combining key frames with VSLAM system is introduced. In reference [19], a VSLAM system based on key frame extraction is proposed, which classifies visual mapping, loop detection and navigation as parallel tasks. In reference [16], an algorithm with key frame extraction is

IJEIT
ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 8, Issue 6, December 2018

presented. And in VSLAM system, mapping and navigation exit at the same time and are taken as two parts to be processed in parallel. In this reference, the improved FAST angular points can be used to extract the same ORB image features, which can be effectively optimized and corrected in loop detection so to ensure the validity and integrity of image feature extraction.

The use of point feature is the mainstream of image feature extraction, and the use of ORB feature [23] and SIFT [25] feature of improved FSAT angular points is the main use scheme. SIFT has been developed from the 1990s to now and has made great achievements. SIFT adopts 128-dimension vector for presentation. So, compared with other methods, the rotation, scale and radiation are invariable. However, the vector dimension is too high, which increases the time complexity, but the algorithm is with robustness when it is affected by surrounding environment (such as optical flow field). In reference [30]-[32], a method of SUFR feature extraction is proposed, which reduces the time complexity of VSLAM because of its special computation. In reference [34]-[35], efficiency of the SIFT feature algorithm is compared with the original one and proves that the efficiency is seven times higher. The improved method of ORB feature descriptor based on FAST angular points [33] is proposed by Ethan Rublee [20]. Because the speed of binary descriptor is faster and more efficient than that of conventional ones, which shortens the work cycle of ORB feature, the combination of binary feature descriptors and BRIEF [21] effectively solves this problem. ORB feature description is the Euclidean distance between feature vectors, which take the matching computation between STFT and SUFR of adjacent frames as the measurement. In reference [36]-[38], a method based on ORB feature description is described.

Table 1. Comparison of image filtering

Literature	Means
SLAM with a single camera	The extended Kalman filter algorithm is adopted based on monocular vision VSLAM.
Real-time simultaneous localisation andmapping with a single camera	
Vision-based SLAMusing the Rao-Blackwellised particle filter	A method of combining particle filter with monocular vision location is introduced, which is helpful for the calculation of camera pose motion.
Novel Rao-Blackwellized particle filter for mobile robot SLAM usingmonocular vision	
Parallel Tracking and Mappingfor Small AR Workspaces	The co-existing mapping and navigation are divided into two parts and are processed in parallel.
ORB-SLAM: A Versatile and Accurate Monocular SLAM System	A VSLAM system based on the extraction of key frames is proposed.
A Brain-inspired SLAM System Based on ORB Features	An unscented Kalman filter algorithm is described to solve the problem of linearization during the using of extended Kalman filter algorithm.

III. MAP ENVIRONMENT AWARENESS METHODS

The core of VSLAM system is the awareness on the surrounding environment. With the development of SLAM technology, semantic maps appear in the field of SLAM map construction.Semantic maps have the properties of connotation, root label and detection spacerange. According to these attributes, the environment can be predicted better,the path of the robot can be optimized better,the performance of human-computer interaction can be improved, and people can better understand the detailed information of the environment.The use of ambient awareness in complex and volatile public places can effectively avoid information redundancy and misjudge. Before the use of scene awareness, many SLAM engineers have adopted location awareness technology,by which extracts the features of the information captured by the sensing devices is extracted and compared with the previous captured information [41]. Thus, the current environmental information is obtained without considering the robustness and stability of the system [40]However, as scene awarenessmeans that the information itself has the interactive learning function,which is more practical than scene recognition in application.In some special locations, there is not much difference between visual awareness and recognition.Visual awareness technology can identity locations by calibrating the surrounding environment with visual information collection devices.Compared with the awareness semantic map,visual recognition needs to obtain the relevant information of the detection environment in

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 8, Issue 6, December 2018

advance for complete recognition.Sometimes,when the scene environment has low complexity and small variation,there is no clear distinction between recognition and awareness, and some recognition techniques may be superior to the awareness technology [40-42]. This paper will discuss different sensing information in the visual information of environment layout,geometric information of environment layout and user-oriented information.

A. Location Awareness based on Visual Information of Environment Layout

Large amounts of environment information is obtained by visual sensors,which can provide more rich environment information.Intelligent devices can obtain the 2D and 3D information of surrounding environment by visual sensors.The visual information of environment layout is described in three kinds of spacial distributions with 2D and 3D information.

1) Location Awareness based on Visual Information of Plat Image

Relevant methods of location awareness is mainly based on plat image by taking image feature as a clue.The key problem in the research is to search the image feature(combination) and corresponding processing framework that suitable for location description.For 2D environment information,it is to get the platform image by visual sensing devices,to distract image features,and to classify the features.Therefore,the collection and processing of 2D information emphasizes on the obtaining of feature combination of image information and the algorithm framework used for processing.

Table 2. The summery of awareness methods for plane images

Literature	Means
Recognizing indoor scenes.	A method of the judgment on the global information of map based on the detected local information.
Supervised learning of places from range data using AdaBoost.	A method of the judgement on the global information of map based on the detected local information.
A discriminative approach to robust visual place recognition.	
Detecting and labeling places using online change-point detection.	The video streaming is analyzed for plane awareness. Then, objects with features are mined from the collected plane images, and are classified by Bayesian model.

Under the condition that the image extraction is suitable for indoor scene awareness,Quattoni and more [45] have proposed a method to judge the overall information of the map based on the detected local information.Pronobis and more [48] first usedthe classifier of support vector machine to complete the sensing of environment, and then presented a method to integrate CRFH and SIFT information [49], in order to achieve better awareness effect.Madokoro and more[46] and Luo and more [51] introducedseparately amethod of unsupervised scene information awareness and a method that image information improved the fitness and time validity of scene awareness by using incremental support vector machine. Luo and more [50] linked reference [43] and [48] by support vector machine so to make intelligent devices have the ability of independent judgement. Ranganathan [54] carried out the plane awareness through the analysis of video stream,which was to search objects with features from the collected plane images and then to classify it by Bayesian model,and finally to finish the scene detection by identified and classified information.The summery of awareness methods for plane images in references is shown in Table.2.When RGB-D sensor is used as the main sensing device for plane image collection, RGB information and depth information are obtained by RGB-D.Jung and more [59] have verified the reliability of the awareness from grayscale image and depth information on environmental information by dividing grayscale image and depth information into different locations. Dynamic Bayesian mixture model is used to classify the information collected by RGB-D sensors [56] and to improve the robustness of the system.For the verification on the use of open source database as information source,some VSLAM researchers have proposed to optimize the local awareness information collected by visual sensing devices againsttheoverall information. The plane image information has high collection rateand strong autonomy. However,as the spatial information data is lacked, it is not suitable for 3D spatial

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 8, Issue 6, December 2018

model and is not applicable under the environment with complex and changeable elements.

2) Location Awareness based on 3D Image

Images of plane awareness lack 3D information data in spatial awareness.Now, more and more 3D sensing devices are used for recognition and sensing of intelligent devices. 3D sensing devices,which are easy to operate,can directly obtain 3D information of their surrounding environment.The map model generated from the integration of the RGB information of the surrounding environment and the depth information is called the 3D map model [65],which isshown in Figure 2.The intelligent device can finish the tasks of exploration and path planning according to the 3D map.

Fig.2.VSLAM Process Diagram

In traditional VSLAM system,2D plane map model came after the exploration mapping.With the population of RGB-D SLAM technology,intelligent devices can process the environment information collected by visual sensing devices to generate 3D Point Cloud Map [57-62].3D point cloud map can rebuilt the real 3D map, and express the spatial information intuitively.It is necessary to use PCL point cloud library in the construction of point cloud map. As the freedom degree ofintelligent devices' motion is high,in reference [62], the local description plan is used to consider the global description scheme and a classifier is used to classify the generated local detection environment model. When the data collected by RGB-D camera is processed, the point cloud data is firstly processed to obtain the corresponding point cloud spatial environment in reference [61], and the feature description factor is obtained according to its characteristics;then,the scheme in reference [62] is combined for the identification of the scene classification.In the above scheme,the collected point cloud data is used to generate the local point cloud map for map model planning bythe classifier, which can fully improve the stability of the system and the independent judgment ability of the intelligent equipment.

3) Location Awareness base on Space Distribution of Image Sequence

The space distribution information based on image sequence is classified according to different environmental modes and its unique spatial information, and then the classified results are mapped into the constructed map model.W. Hao and more [63] have divided the map-building into three categories according to the special information of space. It is firstly to classify the spacial areas in plane according to the clustering algorithm,and then to combine with the spatial vertical depth information and the spatial awareness information.Reference [62] has proposed to divide spacial environment according to the clustering method.It is to use large obstacles such as the walls of the room for spacial classification. However,as each part of the space is completely isolated and is not contacted with each other in this method,which is not applicable in wide places.In the large-scale scene environment, reference [64] has suggested to use portable sensors.And multi-mobile terminals are used to explore and sensing in the partitioned space environment,and all terminal sensor devices are clustered to obtain the semantic information of the map environment,so to complete the sensing task of the space.

B. Location Awareness base on Geometry Information of Environment Layout

In the field of spatial awareness,the entity restore of surrounding environment elements is the mainstream of the research direction.Besides,it is also an effective method to describe and plan the geometric information of the elements in the space [66,67],and to obtain the corresponding scene classification by analyzing the geometric information of elements. The restore process of actual elements is finished at the last stage,in which the first task is to obtain the geometric information of spatial elements. In reference [66.67], a relatively simple artificial environment than complex and changing environment informationhasbeen introduced,in which spatial information is described by feature vector,and then the awareness of spacial information is realized for element classification according to the basis of vectors.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 8, Issue 6, December 2018

Table 3. The environment awareness guided by users in VSLAM

Literature	Means
Interactive SLAM using laser and advanced sonar	The equipment scans and locates on the surrounding environment with sensing devices, and generates a map with coordinates. Then the semantic segmentation of the map is carried out according to the coordinates.
Enabling a robot to learn new places in a real home environment by integrating spoken dialog and visual localization	When RGB-D visual camera is used for the collection of environmental elements, the segmentation calculation and recognition of the object are carried out. The experimental results show that the complete recognition and judgment can be finished by man-machine interaction.
Living with robots: interactive environmental knowledge acquisition	
Interactive semantic mapping: experimental evaluation	
Bringing together human and robotic environment representations-a pilot study.	The intelligent device can combine the detected information of map environment elements with the information of human-computer for the optimization of results.

C. Location Awareness based on Guidance of User

Human-computer interaction and the improvement of autonomous learning ability is one of the characteristic for semantic SLAM.The intelligent equipment has the autonomous learning ability,which not only can obtain environment elements by sensors,but also can obtain information from human.In the intelligent equipment,the learning database is its brain.The database is composed with large amounts of data,and intelligent devices can get the autonomous characteristic by the training and learning of knowledge information. The method summary in references on the environment awareness guided by users in VSLAM is shown in Table3.In references,an environment awareness scheme with user guidance is firstly proposed.When intelligent equipment enters the environment,the equipment scans and locates the surrounding environment through the sensing device,and generates a map with coordinates,then the map semanteme is segmented according to coordinates[69] for the effect awareness on the whole environment,which depends on the semantic tags established in the scanning and sensing process of sensing devices.In references [68], [73], [74],RGB-D camera is used for environment element collection and the segmenting calculation and identification of objects.The experiment shows that the man-machine interaction can be used for complete recognition and judgment. In reference [72], intelligent devicesare equipped with human-computer interactive devices.Thus,the intelligent devices can combine the information of map environment elements and human-computer interaction information to achieve a better perception effect.And the correct environment elements obtained and the changing information are transmitted to the intelligent equipment for the optimization of results in reference[72].

It can been seen from the above discussion on the awareness methods of three kinds of environment elements that the development direction of VSLAM semantic maps is the artificial intelligence in which intelligent devices have autonomous learning, autonomousdetection and functional imitation.For the optimization of map awareness method,it needs to adopt artificial intelligence algorithm to improve the autonomous learning ability of intelligent devices.

IV. MAPPING AND DEEP LEARNING

The development direction of intelligent devices is to obtain autonomous learning ability and autonomous judgment ability.For complex and changing environment elements, the combination and use of RGB-D SLAM technology and depth learning algorithms is one of the research hotspots.In this chapter,the method of integrating artificial intelligence and RGB-D SLAM technology is discussed from two aspects,which are deep learning-image VO(visualodometer) and loop detection-deep learning,by comparing the VSLAM technology with artificial intelligence with the classical SLAM technology.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 8, Issue 6, December 2018

A. Visual Odometry

VO(visual odometer) is called the front-end system of the VSLAM system.The pose of the visual sensor,the transmission of information to the back- end and the optimization of camera poses and information data through loop detection are all completed by VO [75]. The method summery of the combination of VSLAM and depth learning algorithms in references is shown in the following table. 4.In the discussion of image feature extraction above,it emphasizes on the ORB feature of FAST angular features,and compares the two adjacent frames.However, the artificial intelligence algorithm does not need complicated extraction and matching operation,which reduces the complexity of the algorithm and makes the system more intelligent.

Table 4. The method summery of the combination of VSLAM and depth learning algorithms

Literature	Means
Neural network library for geometric computer vision	The space transformation network is extended, and the regression on the classical computer vision method is carried out during the designing process of network.
Learning visual odometry with a convolutional network	Architecture of end-to-end-based depth neural network is proposed to predict the change of the speed and direction for camera.
Exploring representation learning with CNNs for frame-to-frame ego-motion estimation	The optimal feature representation of image data is studied by convolutional neural network, and is estimated by the visual odometer.Besides, the robustness of the algorithm in dealing with image motion blur and illumination change is described.
Unsupervised learning of depth and motion	The neural network is used to detect the synchronization of sequential stereo images, and the neural network is used for the synchronous detection of time sequence three-dimensional images, wherein, the estimation of space transformation between three-dimensional image sequences is converted into synchronous detection and this network is also called unsupervised synchronization / depth automatic encoder.

B. Loop Detection

Asd Loop detection is also called closed loop detection, which is the judging process in VSLAM system.In classical VSLAM systems,the convolutional neural network is used by some researchers for the extraction of image features. However,the neural network has several levers,and there are differences in the image information description of every lever.For the description of image feature,Hou has used the CAFFA framework of deep learning and combined with convolutional neural network for feature description in reference[80],which can describe images from different angles.In reference [80],a plan of improving the speed rate of image feature extraction based on depth learning framework (caffa) has been introduced by using deep learning framework CAFFA for image feature extraction.In the references [78], [84],the visual location and scene awareness have been analyzed respectively, and the data source has been trained by neural network to improve image retrieval.With this method,the stability of VSLAM system is improved.And in complex and changing environment,loop detection is necessary.When deep learning algorithm is combined,the predicted data can be compared with local information in loop detection so to optimize the mapping effect.

C. Comparison between Deep Learning and Traditional Mapping Method

In this paper,the combination of VSLAM and deep learning is discussed in front-end design of VSLAM and loop detection.It can be concluded that,when combined with deep learning algorithm,the intelligent device obtains the autonomous learningability,which greatly strengthens the awareness and judgement of the environment.The artificial interference is gradually reduced,and the intelligence degree is improved qualitatively.Deep learning has the following characteristics: (1) the foundation of deep learning is the massive training and calculation of the data and deep leaning have autonomy; (2) the deep learning does not need complicated classification,and the classification of elements is completed together with the training calculation.(3)deep learning has adopted multi-dimension information for training,which improves the reliability of training results; (4)the use ofdeep learning on intelligent equipment can enhance the ability of improving man-machine interaction; (5) no complex algorithm is needed to calculate results in the awareness of image feature; (6) deep learning plays an important role in the research of semantic map.To sum up,deep learning has great potential in the field of SLAM, and it shows many advantages compared with traditional methods.

V. CONCLUSION

This paper has discussed the RGM-D SLAM mapping method from the aspects of image feature extraction,integration with deep learning and space environment awareness.For image feature extraction,the ORB feature extraction of angular points based on FAST is the mainstream extracting program ofVSLAM image,but,its computation complexity needs to be improved.For the space awareness of semantic maps,it concludes that the spatial sensing scheme based on the combination of human-computer interaction and 3D spatial information is more suitable for the awareness of complex and changing environment elements,and it also needs to be improved in intelligent computation.For the integration of deep learning,the computation complexity and sensing detection capacity have been strengthened by combing with deep learning algorithm.However,because the training process of deep learning relays on database heavily,which brings limitation of intelligent equipment in database,the RGM-D SLAM system are still facing severe challenge in the combination of deep learning for intelligenization.

ACKNOWLEDGMENT

This paper is found by The National Natural Science Foundation of China (No. 61802107); The Open Project of Hebeiot data acquisition and processing engineering technology research center (NO. 2016-2). Jiangsu Postdoctoral Research Grant Program (NO. 1601085C). Science and Technology Research and Development Program in Handan (NO. 1625202042-1).

REFERENCES

[1] Y. Xiaofeng, Z. Wenjuan, L. Xu, L. JiangGuo. A novel visual inertial monocular SLAM. International Symposium on Multispectral Image Processing and Pattern Recognition, 2018.

[2] Z. Haijiang, W. Zhicheng, Z. Jinglin, W. Xuejing. Loose fusion based on SLAM and IMU for indoor environment. International Conference on Graphic and Image Processing, 2018.

[3] S. Dan, X. Xingyu, J. Bin, W. Zhonghai, C. Genshe, E. Blasch, K. Pham. A robotic orbital emulator with lidar-based SLAM and AMCL for multiple entity pose estimation. Defense + Security, 2018.

[4] R. WanPing, W. De Cheng. Numerical study of the wave-induced slamming force on the elastic plate based on MPS-FEM coupled method. Journal of Hydrodynamics,vol. 30, no. 01, pp.70-78,2018

[5] Z. SunChun, Y. Rui, L. JiaXin, C. Ying-Ke , T. Huajin. A Brain-inspired SLAM System Based on ORB Features. International Journal of Automation and Computing, vol.14, no. 05, pp.564-575, 2017.

[6] L. Yunhui,Z. Fan, G. Ruibin, W. Jiangliu,N. Qiang,W. Xin,W. Zerui.Robot Intelligence for Real World Applications. Chinese Journal of Electronics, vol. 7, no. 03, p.446-458, 2018.

[7] A. J. DAVISON. SLAM with a single camera //Proceedings of Workshop on Concurrent Mapping an Localization for Autonomous Mobile Robots in Conjunction with ICRA. Washington, DC, USA, pp. 18-27, 2002.

[8] A. J. DAVISON. Real-time simultaneous localization and mapping with a single camera Proceedings of the Ninth IEEE International Conference on Computer Vision. Washington, DC, USA, pp. 1403-1410, 2013.

[9] A. J. DAVISON, I. D. REID, N. D. MOLTON N D, . Mono‐SLAM: real-time single camera SLAM. IEEE transactions on pattern analysis and machine intelligence, vol. 29, no. 6, pp. 1052-1067, 2007.

[10] J. CIVERA, A. J. DAVISON, J. M. M. MONTIEL. Inverse depth parameterization for monocular SLAM. IEEE transactions on robotics, vol 24, no. 5, pp. 932-945, 2008.

[11] R. M. CANTIN, J. A. CASTELLANOS. Unscented SLAM for large-scale outdoor environments Proceedings of 2005 IEEE RSJ International Conference on Intelligent Robots and Systems. Edmonton, Alberta, Canada, pp. 3427-3432, 2008.

[12] D. CHEKHLOV, M. PUPILLI, W. M. CUEVAS.Realtime and robust monocular SLAM using predictive multiresolution descriptors Proceedings of the Second

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 8, Issue 6, December 2018

International Conference on Advances in Visual Computi. Lake Tahoe, USA, 276-285, 2006.

[13] S. HOLMES, G. KLEIN, D. W. MURRAY. A square root unscented kalman filter for visual mono SLAM Proceedings of 2008 International Conference on Robotics and Automation, ICRA. Pasadena, California, USA, 3710 -3716, 2008.

[14] R. SIM, P. ELINAS, M. GRIFFIN. Vision based SLAM using the Rao‑Blackwellised particle filter. IJCAI workshop on reasoning with uncertainty in robotics, 9(4): 500-509, 2005.

[15] L. Maohai, H. Bingrong, C. Zesu. Novel Rao-Blackwellized particle filter for mobile robot SLAM using monocular vision. International journal of intelligent technology, 1(1): 63-69, 2006.

[16] G. KLEIN, D. MURRAY. Parallel Tracking and Mapping for Small AR Workspaces IEEE and ACM International Symposium on Mixed and Augmented Reality. Nara, Japan, 225-234, 2007.

[17] G. KLEIN, D. MURRAY. Improving the agility of key frame based SLAM European Conference on Computer Vision. Marseille, France, 802-815, 2008.

[18] R. MUR‑ARTAL R, J. D. TARDÓS. Fast relocalisation and loop closing in key frame based SLAMIEEE International Conference on Robotics and Automation. New Orleans, LA, 846-853, 2014.

[19] R. MUR‑ARTAL,J. M. M. MONTIEL, TARDOS J D. ORB‑SLAM: A Versatile and Accurate Monocular SLAM System. IEEE transactions on robotics, 31(5), 1147-1163, 2015.

[20] E. RUBLEE, V. RABAUD, K. KONOLIGE. ORB: an efficient alternative to SIFT or SURF Proceedings of 2011 IEEE International Conference on Computer Vision. Barcelona, Spain,, 2564–2571, 2011.

[21] J. S. GUTMANN, K. KONOLIGE. Incremental mapping of large cyclic environments Proceedings of 1999 IEEE International Symposium on Computational Intelligence in Robot ics and Automation. 318–325, 1999.

[22] HENRYP, KRAININM, HERBSTE. RGB‑Dmap‑ping: using depth cameras for dense 3D modeling of indoor environments KHATIBO, KUMARV, PAP‑PAS G J. Experimental Robotics. Berlin Heidelberg: Springer, 647-663, 2014.

[23] D. G. Lowe. Distinctive image features from scale invariant key points. International journal of computer vision, 60(2):91-110, 2004.

[24] BAYH, TUYTELAARST, VANGOOLL.SURF: speeded up robust features Berlin Heidel‑berg, Springer, 2006.

[25] RUBLEEE, RABAUDV, KONOLIGE K. ORB: An efficient alternative to SIFT or SURF Barcelona, Spain, 2564-2571, 2011.

[26] ALIAM, JANNORDINM.SIFT based monocular SLAM with multi‑clouds features for indoor navigation 2010 IEEE Region 10 Conference TENCON. Fukuoka, 2326-2331, 2010.

[27] E. Y. WU, L. K. ZHAO, Y. P. GUO.Monocular vision SLAM based on key feature points selection 2010 IEEE International Conference on Information and Automation (ICIA)Harbin, China,1741-1745, 2010.

[28] C. H. CHEN, Y. P. CHAN. SIFT based monocluar SLAM with inverse depth parameterization for robot localization IEEE Workshop on Advanced Robotics and Its Social Impacts ,Hsinchu,China,2007:1-6.

[29] D. X. Zhu. Binocular Vision SLAM Using Improved SIFT Algorithm 2010 2nd International Workshop on Intelligent Systems and Applications (ISA).Wuhan, China, 2010:1-4.

[30] Z. Y. ZHANG, Y. L. HUANG, C. LIC. Monocular vision simultaneous localization and mapping using SURF WCICA 2008. 7th World Congress on Intelligent Control and Automation. Chongqing, China, 1651-1656, 2007.

[31] YEY. The research of SLA Mmonocular vision based on the improved surf feather International Conference on Computational Intelligence and Communication Net‑ works. Hongkong, China, 344-348, 2014.

[32] Y. T. WANG, Y. C. FENG. Data association and map management for robot SLAM using local invariant features 2013 IEEE International Conference on Mechatronics and Automation .Takamatsu, 2013.

[33] ROSTENE,DRUMMONDT.Machine Learning for High‑Speed Corner Detection LEONAR -DISA , BISCHOF H,PINZA, etal. European Conference on Computer Vision. Berlin Heidel- berg: Springer,, 430-443, 2006.

[34] ENGELJ, SCHÖPST, CREMERSD. LSD‑SLAM: Large‑Scale Direct Monocular SLAM FLEETD, PAJDLA T, SCHIELEB,eds. Computer Vision‑ECCV2014. Switzerland: Springer International Publishing, 834 -849, 2014.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 8, Issue 6, December 2018

[35] N. ERA, IZADIS, HILLIGESO. Kinect- Fusion: Real-time dense surface mapping and tracking IEEE International Symposium on Mixed and Augmented Reality. Basel, Switzerland, 127-136.2014.

[36] X. FEN, W. ZHEN.Anembedded visual SLAM algorithm based on Kinect and ORB features201534th Chinese Control Conference. Hangzhou, China, 6026- 6031, 2015.

[37] X. XINGX, X. T. ZHANG, X. WANGX. ARGBD SLAM algorithm combining ORB with PROSAC for indoor mobile robot 20154th International Conference on Computer Science and Network Technology (ICCSNT). Harbin, China, 2015:71-74.

[38] J. LI, S. PANT, K. TSENGK. Design of a monocular simultaneous localization and mapping system with ORB feature International Conference on Multimedia and Expo (ICME), SanJose, California, USA, 2013:1-4.

[39] A. Swadzba, S. Wachsmuth. Indoor Scene Classification Using Combined 3D and Gist Features Computer Vision – ACCV 2010. Springer Berlin Heidelberg, 2010.

[40] E. Fazl-Ersi, J. K. Tsotsos. Histogram of Oriented Uniform Patterns for robust place recognition and categorization. International Journal of Robotics Research, 468-483, 2012, 31(31).

[41] A. Pronobis. Semantic mapping with mobile robots [Ph.D. dissertation], KTH Royal Institute of Technology, Sweden, 2011.

[42] I. Ulrich, I. Nourbakhsh. Appearance-based place recognition for topological localization. In: Proceedings of the 2000 IEEE International Conference on Robotics and Automation. San Francisco, CA, USA: IEEE, 1023−1029, 2000.

[43] O. M. Mozos, C. Stachniss, W. Burgard. Supervised learning of places from range data using AdaBoost. In: Proceedings of the 2005 IEEE International Conference on Robotics and Automation. Barcelona, Spain: IEEE, 1742−1747, 2005.

[44] C. Premebida, D. R. Faria, F. A. Souza, U. Nunes. Applying probabilistic mixture models to semantic place classification in mobile robotics. In: Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems. Hamburg, Germany: IEEE, 4265−4270, 2015.

[45] A. Quattoni, A. Torralba. Recognizing indoor scenes. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA: IEEE, 413−420, 2009.

[46] H. Madokoro, Y. Utsumi, K. Sato. Scene classification using unsupervised neural networks for mobile robot vision. In: Proceedings of the 2012 SICE Annual Conference. Akita, Japan: IEEE, 1568−1573, 2012.

[47] J. Wu, J. M. Rehg. Where am I: place instance and category recognition using spatial PACT. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, Alaska, USA: IEEE, 2008. 1−8.

[48] A. Pronobis, B. Caputo, P. Jensfelt, H.Christensen. A discriminative approach to robust visual place recognition. In: Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems. Beijing, China: IEEE, 3829−3836, 2006.

[49] A. Pronobis, B. Caputo. Confidence-based cue integration for visual place recognition. In: Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems. San Diego, California, USA: IEEE, 2394−2401, 2007.

[50] J. Luo,A. Pronobis, B. Caputo, P. Jensfelt. Incremental learning for place recognition in dynamic environments. In: Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems. San Diego, California, USA: IEEE, 721−728, 2007.

[51] A. Pronobis, O. M. Mozos, B. Caputo, P. Jensfelt. Multi-modal semantic place classification. The International Journal of Robotics Research, 29(2−3): 298−320, 2010.

[52] F. F. Li, P. Pietro. A Bayesian hierarchical model for learning natural scene categories. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA: IEEE, 524−531, 2005.

[53] A. Ranganathan. Pliss: Detecting and labeling places using online change-point detection. In: Proceedings of the 2010 Robotics: Science and Systems. Zaragoza, Spain: MIT Press, 185−192. 2010.

[54] J. X. Wu, H. I. Christensen, J. M. Rehg. Visual place categorization: problem, dataset, and algorithm. In: Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. St. Louis, MO, USA: IEEE, 4763−4770, 2010.

[55] O. M. Mozos, H. Mizutani, R. Kurazume, T. Hasegawa. Categorization of indoor places using the Kinect sensor. Sensors, 2012, 6695−6711, 2012.

[56] I. Kostavelis, K. Charalampous, A. Gasteratos, J. K. Tsotsos. Robot navigation via spatial and temporal coherent semantic maps. Engineering Applications of Artificial Intelligence, 48: 173−187, 2016.

[57] H. Jung, O. M. Mozos, Y. Iwashita, R. Kurazume. Local N-ary Patterns: a local multi-modal descriptor for place categorization. Advanced Robotics, 30(6): 402−415, 2016.

[58] H. Jung, O. M. Mozos, Y. Iwashita, R. Kurazume. Indoor place categorization using co-occurrences of LBPs in gray and depth images from RGB-D sensors. In: Proceedings of the 5th International Conference on Emerging Security Technologies. Alcala de Henares, Spain: IEEE, 40−45, 2014.

[59] N. Jie, B. Xiong-Zhu, Q. Kun, L. Zhong. An indoor scene recognition method combining global and saliency region features. Robot, 37(1), 122−128, 2015.

[60] C. RomeroGonz´alez, J. Mart´ınezGo´mez, I. Garc´ıa-Varea, L. Rodrıguez-Ruiz. 3D spatial pyramid: descriptors generation from point clouds for indoor scene classification. Machine Vision and Applications, 27(2), 263−273, 2016.

[61] Z. Zivkovic, O. Booi, B. Krose. From images to rooms. Robotics and Autonomous Systems, 411−418, 2007.

[62] W. Hao, T. Guo-Hui, C. Xi-Bo, Z. Tao-Tao, Z. Feng-Yu. Map building of indoor unknown environment based on robot service mission direction. Robot, 32(2), 196−203, 2010.

[63] A. Rituerto, A. C. Murillo, J. J. Guerrero. Semantic labeling for indoor topological mapping using a wearable catadioptric system. Robotics and Autonomous Systems, 62(5): 685−695, 2014.

[64] J. P. Laumond. Model structuring and concept recognition: two aspects of learning for a mobile robot. In: Proceedings of the 8th International Joint Conference on Artificial Intelligence. Karlsruhe, West Germany: Morgan Kaufmann Publishers Inc, 839−841, 1983.

[65] A. Swadzba, S. Wachsmuth. Indoor scene classification using combined 3D and gist features. In: Proceedings of the 10th Asian Conference on Computer Vision. Queenstown, New Zealand: Springer, 201−215, 2010.

[66] A. Swadzba, S. Wachsmuth. A detailed analysis of a new 3D spatial feature vector for indoor scene classification. Robotics and Autonomous Systems, 62(5): 646−662, 2014.

[67] T. Spexard, S. Y. Li, B. Wrede, J. Fritsch, G. Sagerer, O. Booij, Z. Zivkovic, B. Terwijn, B. Krose. BIRON, where are you? Enabling a robot to learn new places in a real home environment by integrating spoken dialog and visual localization. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Beijing, China: IEEE, 934−940, 2006.

[68] A. Diosi, G. Taylor, L. Kleeman. Interactive SLAM using laser and advanced sonar. In: Proceedings of the 2005 IEEE International Conference on Robotics and Automation. Barcelona, Spain: IEEE, 1103−1108, 2015.

[69] M. Milford, R. Schulz, D. Prasser, G. Wyeth, J. Wiles. Learning spatial concepts from RatSLAM representations. Robotics and Autonomous Systems, 55(5), 403−410 2007.

[70] C. Nieto-Granda, J. G. Rogers, B. Trevor J, Christensen H I. Semantic map partitioning in indoor environments using regional analysis. In: Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems. Taipei, China: IEEE, 1451−1456, 2010.

[71] E. A. Topp, H. Huettenrauch, Christensen H I, Eklundh K S. Bringing together human and robotic environment representations-a pilot study. In: Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems. Beijing, China: IEEE, 4946−4952, 2006.

[72] G. Gemignani, R. Capobianco, E. Bastianelli, D. D. Bloisi, L. Iocchi, D. Nardi. Living with robots: interactive environmental knowledge acquisition. Robotics and Autonomous Systems, 78: 1−16, 2016.

[73] G. Gemignani, D. Nardi,D. D. Bloisi, R. Capobianco, L, Iocchi. Interactive semantic mapping: experimental evaluation. In: Proceedings of the 14th International Symposium on Experimental Robotics. Tokyo, Japan: Springer, 339−355,2016.

[74] B. Kitt, A. Geiger, H. Lategahn. Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme Intelligent Vehicles Symposium. Piscataway, USA: IEEE, 486-492, 2010.

[75] K. Konda, R. Memisevic. Unsupervised learning of depth and motion [EB/OL], 2016-11-10.

[76] M. Jaderberg, K. Simonyan, A. Zisserman. Spatial transformer networks Advances in Neural Information Processing Systems. San Francisco, USA: Morgan Kaufmann, 2015: 20172025.

[77] Z. T. Chen, O. Lam, A. Jacobson. Convolutional neural network based place recognition [EB/OL], 2014-09-06.

[78] M. Cummins, P. Newman. FAB-MAP: Probabilistic localization and mapping in the space of appearance.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 8, Issue 6, December 2018

International Journal of Robotics Research, 27(6): 647-665, 2008.

[79] Y. Hou, H. Zhang H, S. L. Zhou S L. Convolutional neural network based image representation for visual loop closure detection IEEE International Conference on Information and Automation. Piscataway, USA: IEEE, 2238-2245, 2015.

[80] N. Sunderhauf, S. Shirazi S, A. Jacobson A. Place recognition with ConvNet landmarks: Viewpoint-robust, condition-robust, training-free[C/OL] Robotics: Science and Systems, 201606-27.

[81] R. Gomez-Ojeda R, M. Lopez-Antequera M, N.. Petkov. Training a convolutional neural network for appearance invariant place recognition [EB/OL], 2015-3-27.

[82] N. Sunderhauf, S. Shirazi, F. Dayoub. On the performance of ConvNet features for place recognition/IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 4297-4304, 2015.

[83] R. Arandjelovic,P. Gronat,A. Torii.NetVLAD:CNN architecture for weakly supervised place recognition IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 5297-5307, 2016.

[84] X. Li, R. Belaroussi. Semi dense 3D semantic mapping from monocular SLAM [EB/OL], 2016-11-13.

[85] A. Handa, M. Bloesch, V. Patraucean. Gvnn: Neural network library for geometric computer vision Lecture Notes in Computer Science, vol. 9915. Berlin, Germany: Springer Verlag, 67-82, 2016.

[86] K. Konda, R. Memisevic. Learning visual odometry with a convolutional network Proceedings of the 10th International Conference on Computer Vision Theory and Applications. Lisbon, Portugal: SCITCC Press, 2015: 486-490, 2015.

[87] G. Costante, M. Mancini, P. Valigi. Exploring representation learning with CNNs for frame-to-frame ego-motion estimation. IEEE Robotics and Automation Letters, 1(1): 18-25, 2016.

AUTHOR BIOGRAPHY

Tian-Yu Zhu is a master student in Hebei University of Engineering, major in computer technology，research area is image processing.

Wei Wang is a doctor, lecturer, graduated from University of Science & Technology Beijing (USTB) in control science and engineering. Have long engaged in theory and technology of Internet of things, man-machine interaction, affective computing and computational intelligence. Participated in the National High-tech R&D Program (863 Program) and The National Natural Science Foundation of China, and a number of horizontal topics. More than 20 academic papers have been published, 11 have been retrieved by SCI and EI, and 3 have been authorized and published patents. Member of CCF。

Tong Wang is a master student in Hebei University of Engineering, major in computer technology，research area is image processing.

Yi-Hao Cuiis a master student in Hebei University of Engineering, major in computer technology, research area is image processing.