CONSTRUCTION OF OBSTACLE ELEMENT MAP BASED ON INDOOR SCENE RECOGNITION

F. Li [1], H. Wang [1], P. H. Akwensi [1], Z. Kang [1, *]

[1] Department of Remote Sensing and Geo-Information Engineering, School of Land Science and Technology, China University of Geosciences, Xueyuan Road, Beijing, 100083 CN, - foudar@163.com, shaonian1001@126.com, ahtimeless@outlook.com, zzkang@cugb.edu.cn

Commission IV, WG IV/5

KEY WORDS: PointNet, Point cloud, Indoor scenes, Markov Random Field, Semantic recognition, Obstacle element map

ABSTRACT:

Route planning and navigation in indoor space have become a hot topic recently. To accomplish this task, a map and a real-time detection system are needed. Due to Lidar systems' high efficiency in data acquisition, Lidar sensors have become an indispensable part of an object detection system. In this paper, we use Lidar points to generate obstacle maps. The obstacle maps can be used as a reference for route planning and navigation. To identify single objects more precisely, a deep network combined PointNet with Markov Random Field (MRF) is designed in our work to classify Lidar points. Then, single objects are segmented by using the Euclidean clustering method. After that, the prior rules and derived criteria we summarized from large amount images are used to determine objects' kind between Influence Movement Obstacles (IMO) and Non-Influence Movement Obstacles (N-IMO). Finally, objects are projected into a 2D plane to generate obstacle maps. To evaluate the performance of our method, experiments were performed on the S3DIS dataset of Stanford University. The results show that our method greatly improves the overall accuracy compared to the original PointNet model, and can generate high-quality obstacle maps.

1. INTRODUCTION

Semantic scene recognition is a challenging task for robot vision, especially when used for human symbiotic robots living and working collaboratively with humans together in daily life (Kyosuke et al., 2017). Compared with outdoor counterpart, indoor scene annotation is a relatively difficult issue since it usually contains illumination variations, occlusions and overlaps among objects, significant appearance variations and imbalanced representations of object categories (Chu et al., 2017). To data it has become a more noteworthy issue how to connect the indoor scene recognition with the real life. Hence, this paper presented a method of fusion the indoor scene recognition with the real life to construct the obstacle element map.

In recent years, many methods about indoor scene recognition have been presented. Such as Random forests (Fröhlich et al., 2012), Support Vector Machines (SVM) (Chuan et al., 2009), Conditional Random Field (CRF) (Zheng et al., 2015) and Bayesian classifier (Alexander et al., 2012) and so on. With the development of the deep learning, the deep learning -based methods has been increasingly popular in recent years.

Nowadays, numerous deep learning-based architectures are developed, such as Convolutional Neural Networks (CNN) (Girshick, 2015), Recurrent Neural Networks (RNN) (Francesco et al., 2016), Multi-scale Convolutional Neural Networks (MCNN) (Zhao and Du, 2018), Fully Convolutional Network (FCN) (Jonathan et al., 2015), Visual Geometry Group Network (VGGNet) (Simonyan and Zisserman, 2014), Google Inception Network (GoogleNet) (Szegedy et al., 2015), Residual Network (ResNet) (Ren et al., 2012), Recurrent Convolutional Neural Networks (R-CNN) (Ren et al., 2015) and so on, and show superior performance in many applications. However, the data of being used to the models are usually images data or depth images data. And color images data or depth images data have its own limitation that can't describe the real world better. Compared with color images and depth images, the 3D point cloud data not only have RGB information, but also have more comprehensive spatial geometry information. Consequently, the 3D point cloud data have the ability of expressing the real world relatively better. In 2016, the team of Professor Silvio Savarese of the Computer Vision Laboratory at Stanford University in the United States proposed a network model that can apply point cloud data to deep learning, named PointNet (Charles et al., 2017). The availability of PointNet made the point cloud data can be used without voxelization (Dai et al., 2018) or super-voxelization operation directly, and reduce the loss of spatial features of point cloud data in the process. From then on, the utilization of 3D point cloud gets into the new era. However, PointNet has also the boundedness that it could not get a better result at recognition of large scale scene. Therefore, we proposed a method integrating PointNet with Markov Random Field (MRF) to improve the precision of scene recognition.

When it comes to construction of obstacle element map or construction of indoor navigation map, several methods have been proposed. Nowadays, the most common method is to generate the map manually by software including Arcgis, Auto CAD and so on. With the advancement of technology, the predecessors also have proposed some automatic mapping methods, such as probabilistic mapping (Nüchter and Andreas, 2008), feature based mapping (Hao and Srigrarom, 2016), and the image intensity and shadow based mapping (Pradeep et al.,

* Corresponding author

2018). Except above methods, the more popular now is Simultaneous Localization and Mapping (SLAM), it is one of the main techniques for map generation. (Egodagamage and Tuceryan, 2017).

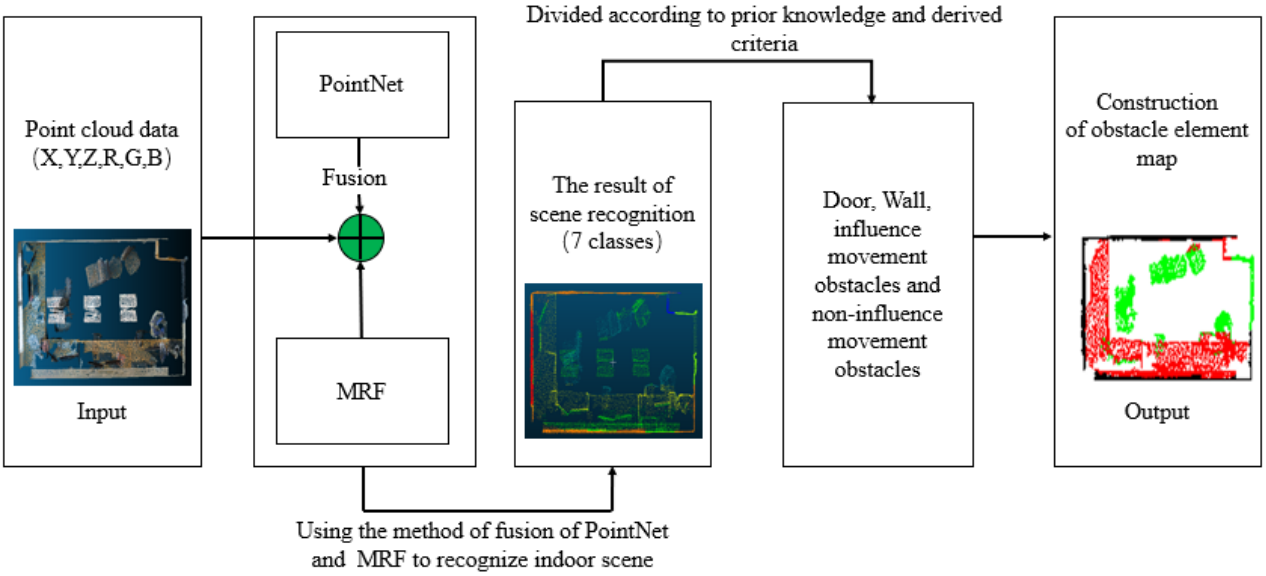


Fig. 1 The construction outline of the obstacle element map

Through the method of constructing obstacle element map based on indoor scene recognition proposed in this paper, the indoor scene recognition and indoor navigation can be combined better, and it has strong practical significance.

The rest of the paper is as follows. Section 2 describes the proposed method in detail. Section 3 presents the experimental results and analysis for evaluating the proposed method. This paper concludes with a discussion of future research considerations in Section 4.

2. METHODOLOGY

In this paper, we proposed a method based on indoor scene recognition to automatically construct obstacle element map. This method combines semantic information from point cloud data, and indoor scene recognition to construct the obstacle element map. Thus: (1) Point cloud data input, and PointNet and MRF model fusion for recognition of indoor scenes. (2) Classification of the point cloud data of objects obtained in step (1) into two groups (i.e., influence movement obstacles and non-influence movement obstacles) and the extraction of semantic information for the four main classes (wall, door, influence movement obstacles and non-influence movement obstacles). (3) Construction of obstacle element map from the semantic information obtained from step (2), Fig. 1 outlines the workflow.

2.1 Recognition of Indoor Scene

In this section, we consider point cloud data as the experimental data. Compared with other types of data, point cloud data not only has RGB information, but also has more comprehensive spatial geometric information than images or depth images. Conversely, two-dimensional data is compared with three-dimensional data, lacking much valuable information about the geometry and geometric layout of objects (Anand et al., 2012). Hence, in this paper, we proposed a method to recognize indoor scene by fusing PointNet with MRF using point cloud data.

2.1.1 PointNet Model: The PointNet (Charles et al., 2017) has three key parts: the max pooling layer as a symmetric function to aggregate information from all the points, a local and global information combination structure, and two joint alignment networks that align both input points and point features. To further improve on the PointNet model, we added colour features based on the original PiontNet architecture, thus improving recognition accuracy of indoor scene greatly.

2.1.2 Markov Random Field (MRF): Markov Random Field (MRF) (Geman and Geman, 1987) is a widely commended model. It has been used for a lot of meaningful things, such as scene annotation (Ren et al., 2012), scene segmentation (Russell et al., 2009), model reconstruction (Sengupta and Sturgess, 2015), etc. and so on. The MRF model is a weighted undirected graph E with a set of vertices V and a set of undirected edges between neighbouring vertices (Liu et al., 2018). In this paper, when scene recognition of point clouds is performed, many point clouds get distributed in discrete random fields, V represents the set of points in point cloud data, and each random variable has a point associated with it. Our purpose is to infer the label of point cloud data $Y = \{y_1, y_2, \ldots, y_V\}$, where y_i is the label of point cloud i.

2.1.3 Fusion of PointNet and MRF: For the traditional PointNet, when point cloud segmentation or classification is implemented on a small scale, we can get better a result. However, when we implemented the traditional PointNet on a large scale data such as ours, we couldn't get a better result, so, there was a need to improve it. We combined the PointNet with MRF based on the original PointNet architecture. In fusion process, expression of the energy function is the most important. Motivated by idea of deep learning and random field fusion (Liu et al.,2018; Li and Wang, 2016; Zheng et al.,2015), we considered the output layer of the PointNet as the unary term. Also, we can generate the pairwise term by using the feature correlation matrix to find out similar points and restrain the corresponding features to become more similar. So, the energy function is shown by Eq (1). As a result, the point clouds in indoor scene are classified into seven classes through PointNet-MRF integration.

$$\mathrm{En(Y)} = \mathrm{En}_{\mathrm{unary}}(Y) + \lambda_1 \cdot \mathrm{En}_{\mathrm{pairwise}}(Y) \quad (1)$$

where the unary energy term is the output layer of the PointNet, and the pairwise energy item is generated by using feature correlation matrix to find out the more similar points.

2.2 Construction of Obstacle Element Map

This section includes mainly two parts. Section 2.2.1 describes the criteria and method for influence movement obstacles and non-influence movement obstacles determination. The generation and expression of obstacle element map is described in section2.2.2

2.2.1 Criteria and Method: The indoor scenes are generally more complicated. According to the Flexible Space Subdivision (FSS) framework principle (Abdoulaye et al., 2018), objects in indoor environments be divided into three classes, thus: the static objects (walls, ceilings, floors, etc.), semi-moving objects (beds, tables, chairs, etc.), and moving objects (mainly people and robots, etc.). The focus of this study is the recognition of static (specifically wall) and semi-moving object classes.

There are several kinds of objects in an indoor scene. From the perspective of indoor navigation, some of them will obstruct the movement, and others will not. Thus, we can divide the objects in the room into two classes, those that influence movement and those that do not influence movement. Before determining whether the obstacles influence the movement or not, we collected a large number of images of the environment (family room, office, hospital, school classroom and indoor environment of the train station, etc.), as shown in Fig. 2. After summarizing the indoor environment in the form of samples, we made some observations.

For the hospital indoor environment, the beds and cabinets in the ward are placed against the wall. In the clinic, the doctor's desk and the bed for observing the condition of patient are also placed against the wall. But there are a lot of benches in the waiting room that are not placed against the wall. In the classroom scene, some of the chairs and tables are placed against the wall, and the other part are basically placed far away from wall. And, there is a teaching desk in front of classroom. The desks, sofas etc. are placed against the wall, and some employees' desks are placed away from the wall, and the conference tables are basically placed in the central position of the unit space in office buildings. In shopping mall, most of the rest chairs and the shelves in the store are placed against the wall, in addition, a small number of shelves are placed away from the wall. In the family room, the beds, bookcases, desks, sofas, wardrobes, etc. are placed near the wall, but the objects such as dining tables, coffee tables, etc. are placed away from the wall. In the station, objects such as ticket vending machines are basically placed against the wall, in the waiting area, most of the rest benches are placed away from the wall. Some of the tables and chairs are placed close to the wall, and many others are placed far from the wall inside the restaurant. Therefore, we can summarize the following prior knowledges according to the layout rules of each typical scenario:

(1) In the hospital, the beds of patient, cabinets, tables and benches are IMO.
(2) The students' desks and the teaching table are IMO in the classroom scene.
(3) In the office buildings, the desks, sofas, conference tables, bookcases and so on are the IOM.
(4) For the shopping mall scene, the rest chairs, shelves and the tables are IOM.
(5) At family room, so many objects are IOM, such as, the beds, bookcases, desks, sofas, wardrobes, etc.
(6) There are mainly dining tables and bar counter are the IMO at restaurant.
(7) In the station, objects such as ticket vending machines and the benches are IMO.

If the objects are not included above all in each scene, we will determine them by two criteria. First, we can calculate the shortest distance between wall and object, and consider it as a criterion. Furthermore, some of the objects are especial, and we can determine whether they influence movement or not according to their properties (weight, wheels). However, we usually can't get the weight of an object, so, we use the volume attribute of objects to measure the mobility of them. Moreover, because the wheels are mostly located at the bottom of the object, they are easily blocked by the objects themselves. Unfortunately, point cloud data in these areas sometimes has occlusion, and it is difficult to determine whether the wheels exist or not, hence, the availability of wheels as a criteria is not included in the scope of this study. For the objects that influence movement, we have to avoid them when we carry out indoor path plan. And for the objects that do not influence movement, we need not to avoid them. The detail discussion of possible priori rules are as follows:

First, we have two indicators, which are the shortest distance D from the wall and the volume V_t of the space occupied by the object. (1) All point clouds after indoor scene recognition can be divided into three types. The first type includes doors and walls, and belongs to the two major parts of the obstacle element map. The second type mainly includes tables, chairs, and bookcases etc. that are easily recognizable. The third type is referred to as others, and such objects will not be counted in the production process of the obstacle element map. (2) When classifying according to the shortest distance from the recognized indoor object to the wall, we need to calculate the shortest distance D between the object and the wall, and setting an appropriated threshold D'. (3) For the calculation of the volume V of the object, first, we create a grid on the XOY plane, and the area of each grid is set to S, and then the point clouds are projected onto the XOY plane. For each grid, if it contains points, it will be marked as 1, and if there is no point, it is marked as 0. Finally, the number of grids with the value of 1 is represented as N'. Now, the maximum and minimum values of the Z values in the cloud cluster are determined as $Zmax$ and

Zmin respectively, and the corresponding difference $Z' = Zmax - Zmin$ is calculated. Hence, the formula for calculating the volume of the object is as shown in equation (2). (4) In the z-axis direction, the minimum value z for a point cloud cluster is compared with the minimum value *Zmw* of the wall class point cloud, the difference is recorded as *Zc*, and this operation's formula is as shown in equation (3). Now a threshold *Zy* is set. If $Zc > Zy$, the point cloud cluster will belong to other class and henceforth ignored.

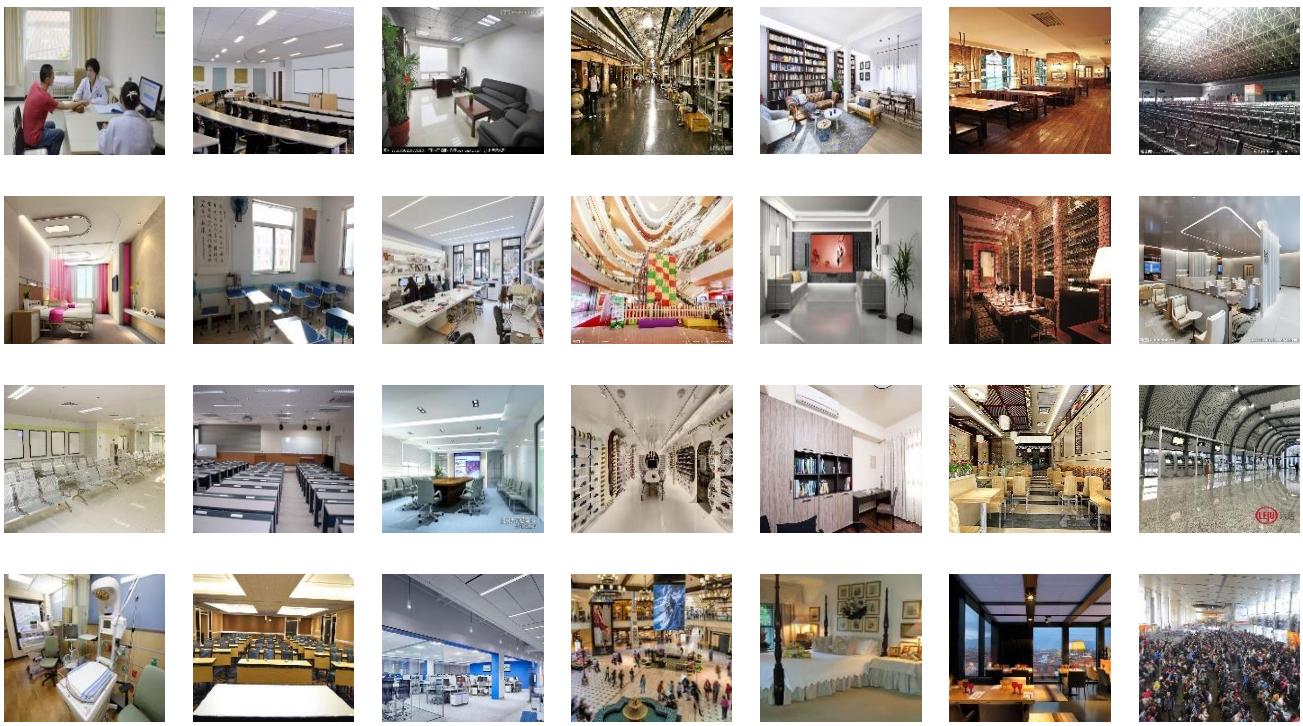
$$V_t = N' * S * Z' \tag{2}$$

$$Z_c = z - Zmw \tag{3}$$

Based on the above two criteria, we conclude four possible cases:

$D<D'$ and $V_t>V_t'$;
$D<D'$ and $V_t<V_t'$;
$D>D'$ and $V_t>V_t'$;
$D>D'$ and $V_t<V_t'$;

Where V_t' is set as the volume threshold. Finally, based on the a priori rules we have obtained, we can make the following judgments. If there is the first to third condition, the object will be classified as an object that influence movement, and the object needs to be circumvented during path planning and navigation. If there is a case of the fourth condition, the object will be classified as non- influence movement. At the same time in path planning and navigation, we do not need to avoid this obstacle. In this paper, the data we use belongs to the office building scene, thus, we could use the third prior knowledge and derived criteria to determine objects.



(a)Hospital (b)Classroom (c)Office building (d)Shopping mall (e)Family room (f)Restaurant (g)Station

Fig. 2 The image data of collecting

Based on the above multiple criteria, the point cloud experimental data finally can be divided into five classes (doors, walls, influence movement obstacle, non-influence movement obstacle and other class), and then complete the final judgment of object classes in complex indoor environments.

2.2.2 Generation of Obstacle Element Map: In this section, we will use the point cloud experimental data that has classified elements to generate obstacle element map. First, we put the point cloud experimental data after finishing the element classification in an XOY plane. Second, the experimental data with doors and walls semantic information are grouped together and completed by RANSAC fitting (Pfister, 2003; Qian and Ye, 2014). Third, we group the data with semantic information of influence movement obstacle and non-influence movement obstacle together, and complete the work of cluster by Euclidean Cluster Extraction (Yu et al., 2015; Sparks and Algorithm, 1973). Finally, we used black lines for walls, green lines for doors, red planes for influence movement obstacles, and green planes for non-influence movement obstacles.

3. EXPERIMENTATION AND ANALYSIS

In this section, the main discussion is the evaluation and analysis of the experiment. To ensure the validity of the experiment, we selected multiple experimental evaluation indicators to evaluate the experiment.

3.1 Experimental Data and Evaluation Indicators

In this paper, we considered S3DIS data set of Stanford university (Dai et al., 2017) as the experimental data, and the data set was collected using the Matterport Camera. The whole data set can be divided into 6 areas covering over 6000 square meters, and including a total of 695,878,620 colour points. The entire dataset mainly includes 13 object classes (structural elements: ceiling, floor, wall, beam, column, door, window and movable elements: table, chair, sofa, bookcase, board and other elements) and 11 scene categories (Office, conference room, hallway, auditorium, open space, lobby, lounge, pantry, copy room, storage room, and toilet). In this paper, we used an office point cloud data as the test data in the experiment, and took nine rooms as training data to perform indoor scene recognition. During the indoor scene recognition experiment, we selected a total of five indicators to evaluate the experimental results, i.e., global accuracy, classification accuracy, average classification accuracy, IoU and mean IoU.

$$\text{Globalaccuracy} = \frac{TP}{GT} \tag{4}$$

$$\text{Classaccuracy}_i = \frac{TP_i}{GT_i} \tag{5}$$

$$\text{Mean classaccuracy} = \frac{\sum_{i=1}^{C} \text{Classaccuracy}_i}{C} \tag{6}$$

$$\text{IoU}_i = \frac{TP_i}{GT_i + FP_i} \tag{7}$$

$$\text{Mean IoU} = \frac{\sum_{i=1}^{c} \text{IoU}_i}{C} \tag{8}$$

where TP and GT denote the total number of points of true positive and ground true respectively, TP_i , GT_i and FP_i denote the number of points of true positive, ground truth and false positive in a class i respectively. C denotes the number of class.

3.2 Experimental analysis

Due to the limitations of the network, the traditional PointNet not being able to yield better recognitional accuracy when performing large-scale recognition and coupled with MRF being a widely accepted model, we proposed the idea of combining MRF with the traditional PointNet model. Fig. 3 shows the raw data in our experiment, and Fig. 4 shows the result of classification point data. From Table I and Table II, we can infer that the recognition accuracies of some classes are lower than those of the PointNet, but the accuracies of most classes are greatly improved. From a holistic point of view, compared to global accuracy, mean class accuracy and mean IoU of PointNet, our proposed method has improved significantly. We are able to obtain the elements (wall, door) of the obstacle element map by classifying the point cloud data. In addition, our classification approach is robust in discriminating the different types of point clouds even when they are really close together.

We combined the point cloud data after completed scene recognition, with the actual application to generate an obstacle element map for indoor path planning and navigation. From the final obstacle element map, we can infer that the generation of this map does not require particularly any better recognition accuracy, so long as the average accuracy of the recognition is above 80%. From the perspective of indoor path planning and navigation, the elements in this element map are simple and avoid many messy and unwanted elements.

Class	Class accuracy (%)		IoU (%)	
	PointNet	Proposed method	PointNet	Proposed method
bookcase	**51.47**	34.44	**37.78**	26.51
chair	79.88	**94.18**	47.92	**86.23**
clutter	61.41	**82.01**	52.99	**66.48**
door	77.40	**90.07**	74.18	**85.67**
table	**62.26**	56.40	**56.72**	51.26
wall	**98.97**	96.79	70.34	**82.30**
window	87.42	**89.92**	**84.87**	84.46

Table I Comparison of classification accuracy of each class, the best performance is marked with **BOLD** fonts.

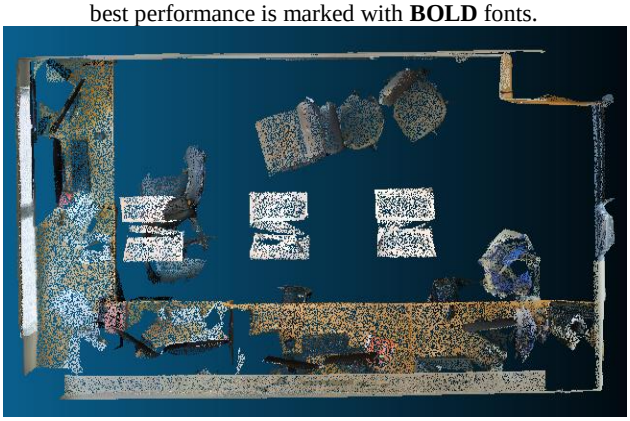


Fig. 3 Experimental raw data scene of point cloud

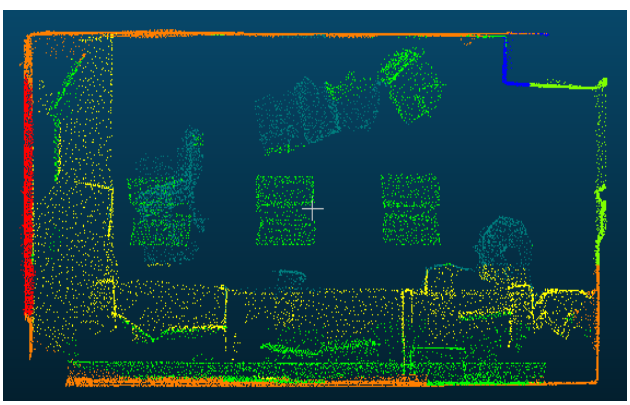


Fig. 4 Classification result of experimental data

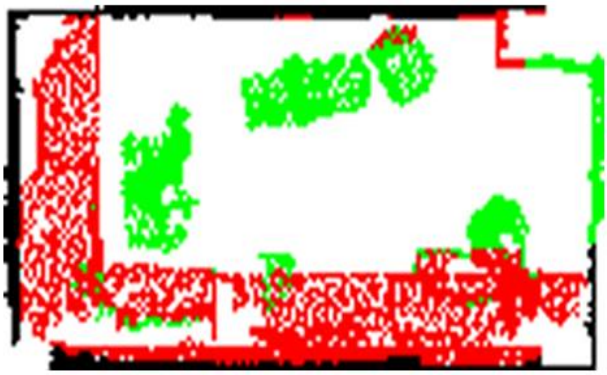


Fig. 5 The final obstacle element map

Method	Data	Global accuracy	Mean classaccuracy	Mean IoU
PointNet	Point cloud	75.09	77.33	60.68
Proposed method	Point cloud	**84.92**	**82.84**	**68.99**

Table II Accuracy comparison of overall classification accuracy, the best performance is marked with **BOLD** fonts.

4. CONCLUSION

We proposed a method for constructing an obstacle element map based on indoor scene recognition. In this paper, we used point cloud data as experimental data, and using PointNet network combined with MRF method to perform scene recognition. This new method is robust and can improve the accuracy of scene recognition effectively. The result of the scene recognition is used to determine whether the semi-moving objects influence movement or not. After the determination of the required elements in the obstacle element map is complete, we could construct the obstacle element map and obtain the desired results. The final obstacle element map, presented in a two-dimensional plannimetric view is as shown in Figure 5. This indoor obstacle element map can be mainly used for indoor path planning and navigation, and its integration into our real lives has great practical significance.

References

Abdoulaye, A., Diakité., and Zlatanova, S., 2018. Spatial subdivision of complex indoor environments for 3D indoor navigation, International Journal of Geographical Information Science. 32(2), pp. 213-235.

Anand, A., Koppula, H. S., and Joachims, T., 2013. Contextually guided semantic labeling and search for three-dimensional point clouds. The International Journal of Robotics Research. 32(1), pp. 19-34.

Chang, C.Y., Wang H. J., and Li, C. F., 2009. Semantic analysis of real-world images using support vector machine. ISSN 0957-4174, Vol 36, pp. 10560-10569.

Chu, J., Xiao, X., Meng, G., Wang, L., and Pan, C., 2017. Learnable Contextual Regularization for Semantic Segmentation of Indoor Scene Images. IEEE International Conference on Image Processing. pp.1267-1271.

Dai, A., Chang, A. X., and Savva, M., 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. /Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5828-5839.

Dai, A., Ritchie, D., and Bokeloh, M., 2018. Large-scale scene completion and semantic segmentation for 3d scans. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4578-4587.

Egodagamage, R., and Tuceryan, M., 2017. Distributed monocular SLAM for indoor map building. Journal of Sensors.

Fröhlich, B., Rodner, E., and Denzler, J., 2012. Semantic Segmentation with Millions of Features: Integrating Multiple Cues in a Combined Random Forest Approach. Asian conference on computer vision. Springer, Berlin, Heidelberg, pp. 218-231.

Geman, S., and Geman, D., 1987. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. Readings in Computer Vision. pp. 564-584.

Girshick, R., 2015. "Fast r-cnn," in Proceedings of the IEEE international conference on computer vision. pp. 1440–1448.

Hao, E. Z., and Srigrarom, S., 2016. Development of 3D Feature Detection and on Board Mapping Algorithm from Video Camera for Navigation. Journal of Applied Science and Engineering. Vol. 19, No. 1, pp. 23-39.

Li, C., and Wang, M., 2016. Combining markov random fields and convolutional neural networks for image synthesis. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2479-2486.

Liu, Z., Lin, G., and Yang, S., 2018. Learning markov clustering networks for scene text detection. arXiv preprint arXiv, 1805.

Liu, Z., Li, X., Luo, P., Loy, C. C., and Tang, X., 2018. Deep Learning Markov Random Field for Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 40(8), pp. 1814-1828.

Long, J., Shelhamer, E., and Darrell, T., 2015. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431-3440.

Nüchter., and Andreas., 2008. 3D robotic mapping: the simultaneous localization and mapping problem with six degrees of freedom. Springer, Vol. 52.

Pfister, S. T., Roumeliotis, S. I., and Burdick, J. W., 2003. Weighted line fitting algorithms for mobile robot map building and efficient data representation. 2003 IEEE International Conference on Robotics and Automation. vol.1, pp. 1304-1311.

Qian, X., and Ye, C., 2014. NCC-RANSAC: a fast plane extraction method for 3-D range data segmentation. IEEE transactions on cybernetics. 44(12), pp. 2771-2783.

Qi, C. R., Su, H., and Mo, K., 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 652-660.

Ren, S., He, K., and Girshick, R., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems. pp. 91-99.

Ren, X., Bo, L., and Fox, D., 2012. RGB-(D) scene labeling: Features and algorithms. 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, pp. 2759-2766.

Russell, B., Efros, A., and Sivic, J., 2009. Segmenting scenes by matching image composites. Advances in Neural Information Processing Systems. pp. 1580-1588.

Sengupta, S., and Sturgess, P., 2015. Semantic octree: Unifying recognition, reconstruction and representation via an octree constrained higher order MRF. IEEE International Conference on Robotics and Automation (ICRA). pp. 1874-1879.

Simonyan, K., and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409, pp.1556.

Sparks, D. N., and Algorithm, A. S., 1973. Euclidean cluster analysis. Journal of the Royal Statistical Society. Series C (Applied Statistics). 22(1), pp. 126-130.

Szegedy, C., Vanhoucke, V., and Ioffe, S., 2016. Rethinking the inception architecture for computer vision. /Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2818-2826.

Tokuhara, K., 2017. Semantic Indoor Scenes Recognition Based on Visual Saliency and Part-Based Features. IEEE/SICE International Symposium on System Integration. IEEE, pp.662-667.

Vezhnevets, A., Ferrari, V., and J. M. Buhmann., 2012. Weakly supervised structured output learning for semantic segmentation, 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, pp. 845-852.

Visin, F., Ciccone, M., and Romero, A., 2016. Reseg: A recurrent neural network-based model for semantic segmentation. /Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp.41-48.

Weingarten, J. W., Gruener, G., and Siegwart, R., 2004. A state-of-the-art 3D sensor for robot navigation. 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566). 3, pp. 2155-2160.

Xie, S., Girshick, R., and Dollár, P., 2017 Aggregated residual transformations for deep neural networks. Aggregated residual transformations for deep neural networks. pp. 1492-1500.

Yu, Y., Li, J., and Guan, H., 2015. Semiautomated extraction of street light poles from mobile LiDAR point-clouds. IEEE Transactions on Geoscience and Remote Sensing. 53(3), pp. 1374-1386.

Zhao, W., and Du, S., 2016. Learning multiscale and deep representations for classifying remotely sensed imagery. ISPRS Journal of Photogrammetry and Remote Sensing, Vol:113. pp.155-165.

Zheng, S., Jayasumana, S., and Romera, P. B., 2015. Conditional random fields as recurrent neural networks. Proceedings of the IEEE international conference on computer vision. pp. 1529-1537.