HIGH TECHNOLOGY LETTERS | Vol. 24 No. 2 | June 2018 | pp. 125 ~ 133

doi:10.3772/j.issn.1006-6748.2018.02.002

Semantic categorization of indoor places using CNN for mobile robot exploration①

Li Guangsheng (李广胜)②*, Chou Wusheng* **

(* School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, P. R. China)

(** State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, P. R. China)

Abstract

The ability of achieving a semantic understanding of workspaces is an important capability for mobile robot. A method is proposed to categorize different places in a typical indoor environment by using a Kinect sensors for mobile robot exploration. At first, the invariant feature based images stitching approach is adopted to form a panoramic image according to Kinect visual information, and the translation between Kinect depth information and obstacle distance information is performed to obtain virtual LIDAR data. Then, the semantic classifier is designed by using convolutional neural networks (CNN) for indoor place categorization based on Kinect visual observations with panoramic view. At last, a frontier-based exploration method is applied to carry out indoor autonomous exploration of mobile robots, which integrates the CNN-based categorization approach. The proposed method has been implemented and tested on a real robot, and experiment results demonstrate the approach effectiveness on solving the semantic categorization problem for mobile robot exploration.

Key words: exploration, mobile robot, semantic categorization, convolutional neural network (CNN), Kinect

0 Introduction

Exploration is a fundamental problem in the field of mobile robotics, whose goal is to acquire as much information about a given environment as possible[1]. Especially for the indoor environment whose structure is structured, rich semantic information about the place where mobile robot is located allows it to perform high-level tasks more effectively and greatly improve its capabilities in various domains, such as localization, path planning or human-robot interaction[2]. Therefore, an important capability for robot is its ability to categorize different places. In this study, the work mainly focuses on the semantic categorization in indoor environments during autonomous exploration of mobile robots.

In the past, many researchers using different types of sensors addressed the problem of semantic place categorization. In some previous work, 2D laser range finders were popular sensors used for scene classification. For instance, Mozos, et al.[3] employed the AdaBoost-based classifier to categorize indoor places by extracting a large number of geometrical features obtained with a laser range finder. The work in Ref. [4] used a support vector machine (SVM) classifier to estimate the type of indoor places based on above mentioned features. Moreover, Shi, et al.[5] introduced the logistic regression approach as a classifier to classify indoor environments into semantic categories. Other work categorized places using vision sensors. Ref. [6] introduced a novel technique called PLISS for place categorization is in accordance with image streams rather than single image. PLISS used change-point detection to segment image sequences, and subsequently perform labeling using a probabilistic classifier and keeping track of its label uncertainty inside a systematic probabilistic framework. The work in Ref. [7] proposed CENTRIST as a visual descriptor for recognizing topological places and scene classification. The descriptors are later classified by SVM. Furthermore, a novel context-based place recognition method that enables mobile robots to categorize places indoors and outdoors was introduced in Ref. [8], which adopted the Histogram of oriented uniform patterns extracted from images. Finally, the combinations of different sensors were also applied to robot place recognition. Ref. [9] presented a supervised learning approach that combines 2D

① Supported by the National Key Basic Research Program of China (No. 2013CB035503).

② To whom correspondence should be addressed. E-mail: Liguangsheng10@163.com

Received on May 27, 2017

laser scans with visual object detection to robustly classify indoor places by using boosting algorithm. In Ref. [10], a multi-modal place classification system based on SVM was proposed to identify places and recognize semantic categories in an indoor environment by fusing multiple visual cues and laser range data. Besides, the recent work in Ref. [11] presented an approach to categorize typical indoor places using local binary pattern histograms of range and reflectance images from 3D laser scans.

In this paper, a semantic categorization approach is proposed to categorize indoor places using Kinect sensor for mobile robot exploration. The key idea of the approach is to use CNN to classify the indoor places, where mobile robot has been explored based on the frontier-based exploration method, according to the visual and depth information provided by Kinect sensor. Compared with other approaches, the proposed method can effectively categorize indoor places into semantic categories with high precision during mobile robot autonomous exploration, and experiment results have validated this approach.

The remainder of this paper is organized as follows. CNN is introduced in Section 1, the details about the visual and depth information extracted from Kinect are given in Section 2. In Section 3, the application of the CNN-based classifier for place categorization is described. Subsequently, the frontier-based exploration method is introduced in Section 4. In Section 5, experiments are conducted. Finally, concluding remarks are provided in Section 6.

1 CNN algorithm

CNN is a multilayer feed-forward neural networks, which consists of various combinations of convolutional layer, sub-sampling layer and fully connected layer. In practice application, CNN has been demonstrated to be an efficient and stable method, and provides even better categorization performance than AdaBoost and SVM.

To construct a classifier for categorizing different indoor places, Caffe[12] is chosen as the base framework of CNN. The Caffe-based CNN architecture is similar to the AlexNet described by Krizhevsky, et al.[13], which is shown in Fig. 1.

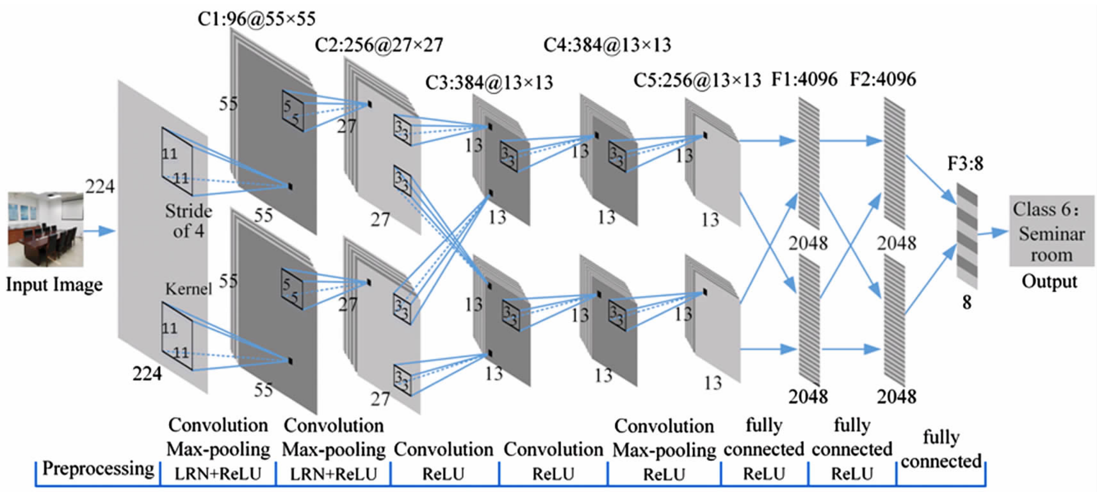


Fig. 1 The overall architecture of the proposed CNN

As depicted in Fig. 1, the proposed CNN consists of five convolution layers, some of which are followed by sub-sampling (max-pooling) layers, and three fully connected layers with a final 8-way softmax. The convolutional layer computes convolutions of the input images with multiple trainable filters (kernels), and generates different feature maps, which are used as input of the next layer. The kernels of the second, fourth, and fifth convolutional layers are connected only to those feature maps in the previous layer, but the kernels of the third convolutional layer are connected to all feature maps in the second layer. The neurons in the fully connected layers are connected to all neurons in the previous layer. Furthermore, response-normalization layers (LRN) are applied to the first and second convolutional layers, and rectified linear unit (ReLU) follows the output of every convolutional and fully-connected layer.

More specifically, each input image has three channels: red, green, and blue. For each channel, the first convolution layer filters the 227×227 (51529-dimensional) input image with 96 kernels of size 11×11 with a stride of 4 pixels. The second convolution layer takes the output of the first convolution layer as input, and filters it with 256 kernels of size 5 × 5. The third convolution layer and the fourth convolution layer has 384 kernels of size 3 × 3 connected to the outputs of the previous convolution layer, respectively. The fifth convolution layer has 256 kernels of size 3 × 3. The fully-connected layers have 4096 neurons, and the output of the final fully-connected layer is a distribution over 8 class labels. In addition, since CNN requires a constant input dimensionality, the image should be first rescaled to a fixed resolution of 227 × 227.

2 Information from Kinect sensor

In this paper, a Kinect is used instead of 2D laser as the main sensor, which can provide both visual information and depth information at high rates. The specific Kinect sensor is ASUS Xtion Pro Live with 58° horizontal field of view and 45° vertical field of view, and the reliable range of depth is approximately 0.8m – 3.5m, as shown in Fig. 2. In the proposed approach, the visual information is applied for the semantic place categorization, while the depth information is used for mobile robot exploration.

Fig. 2 The ASUS Xtion Pro Live Kinect

2.1 Visual information

During the exploration process, a mobile robot takes one observation with panoramic view at some location and assigns the category label to the corresponding place according to the Kinect visual information. However, since Kinect has a restricted working range (field of view), mobile robot does not have a broad perspective to make a 360° vision observation.

To form a complete panoramic scan, mobile robot rotates around its vertical axis and takes 8 snapshots at each target location. In other words, the complete panoramic scan is divided into 8 overlapping partial scans, and the adjacent partial scans has 13° overlapping angle, as shown in Fig. 3. Consequently, each panoramic image consists of 8 sub-images covering the 360° field of view around the robot.

In this section, the ORB[14] based image stitching approach proposed by Adel, et al.[15] is employed to create panoramic image with a high resolution according to the input partial images. As an example, Fig. 4 shows the partial and stitched images.

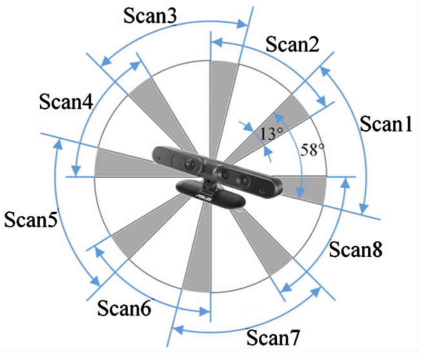


Fig. 3 Partial scans in a panoramic scan with gray areas indicating the overlapping angle

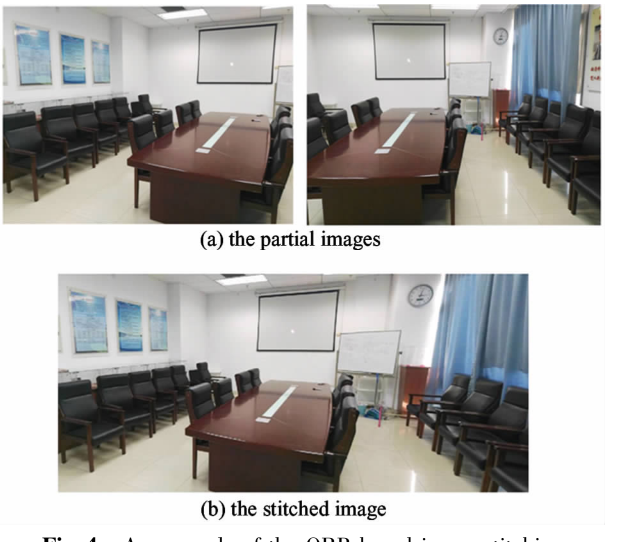


Fig. 4 An example of the ORB based image stitching

By using the ORB feature-based technique, the panoramic image around mobile robot is obtained, and used as input to the CNN-based classifier for categorization purposes.

2.2 Depth information

In this paper, the Kinect sensor is also used as a virtual LIDAR, and the standard laser-based SLAM system implemented in robot operating system (ROS) is adopted for acquiring environment information to build 2D grid map. However, unlike LIDAR scan data, the Kinect's depth information comes in the form of 3D-point cloud because that the depth values have been transformed into points in a 3D-coordinate system. Therefore, in order to use the 2D-based algorithms (e.g. laser-based SLAM), the acquired depth information (3D point cloud data) should be converted into a 2D laser scan.

To create the equivalent LIDAR scan data, ROS provides the package *Pointcloud _ To _ Laserscan*[16], which can convert the 3D-point cloud data into a set of 2D points in a certain plane. The key idea is to read the 3D-point cloud data at a particular height (28cm in our case) and at a certain linear angle, which give us distance information like a laser beam. In addition, the 2D-point data at a specific frequency was republished to match the LIDAR publish frequency. By this way, it is able to make the Kinect appear like a LIDAR for using the Laser-based SLAM and generating real-time grid map.

Fig. 5 shows the Kinect depth information from 3D-point cloud data to the 2D points of LIDAR using *Pointcloud _ To _ Laserscan*.

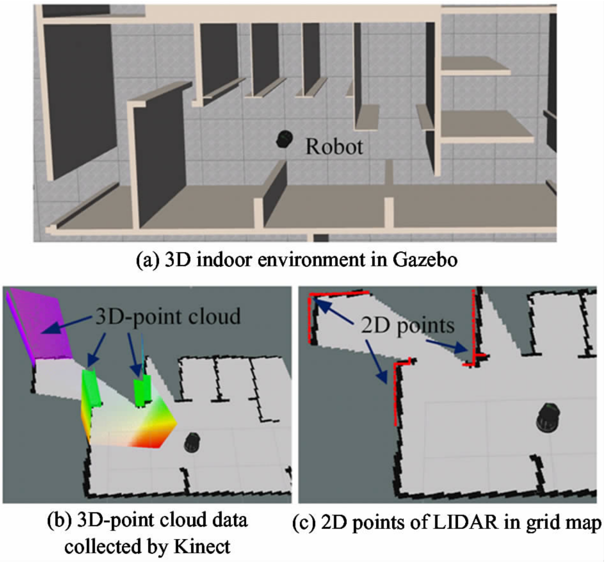


Fig. 5 An example of converting 3D-point cloud data to 2D points in grid map

3 Semantic categorization using CNN through panoramic scans

This paper considers the following types of indoor environment corresponding to eight different place categories. Typical instance of these categories in office environment are shown in Fig. 6. For semantic categorization at indoor places, a classifier learned with CNN is designed based on Kinect visual observations (panoramic scans). The idea behind is to take the panoramic scans as input of CNN-based classifier for each type of place and return the labels according to the position of mobile robot during its exploration phase.

After panoramic scans are obtained, a categorization vector $z(t)$ is constructed representing the t-th panoramic scan in one place, which can be described as

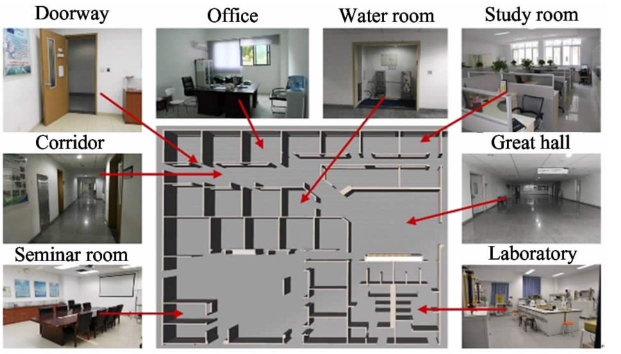


Fig. 6 A typical indoor environment with corridor, doorway, great hall, laboratory, office, seminar room, study room and water room

$$z(t) = \{label_1, \cdots, label_j, \cdots, label_N\} \quad (1)$$

where $label_j$ corresponds to the j-th categories, and N is the total number of categories.

By applying the CNN-based classifier, vector $z(t)$ is transformed into a probabilistic distribution over N categories:

$$\boldsymbol{P}(z) = \{p(label_1 \mid z(t)), \cdots, p(label_j \mid z(t)), \cdots, p(label_N \mid z(t))\} \quad (2)$$

where $p(label_j \mid z(t))$ is the probability of j-th categories for the t-th panoramic scan.

Finally, the final categorization result for the t-th panoramic scan can be obtained by selecting the label with highest probability:

$$Label(t) = \mathrm{argmax}(p(lable_j \mid z(t))), \quad j = 1, \cdots, N \quad (3)$$

An example that panoramic scan taken in a great hall together with its corresponding probability distribution is shown in Fig. 7.

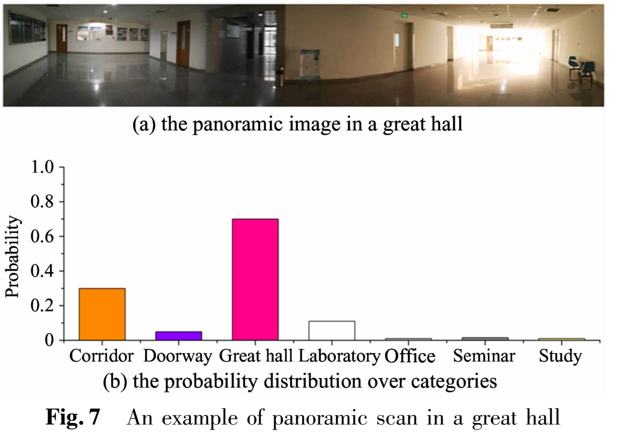


Fig. 7 An example of panoramic scan in a great hall and its probability distribution

In addition, since the categorization method is supervised, training for the CNN-based classifier needs to be performed in advance.

4 Frontier-based exploration method

This section describes the exploration approach to generate navigation goals for acquiring semantic information about indoor environments. To this end, a frontier-based exploration approach is used, which was originally proposed by Yamauchi[17,18].

The frontier-based exploration method provides an efficient way to explore environments, which directs a mobile robot to the frontiers that lie on the boundary of explored space and unexplored space (see Fig. 8). This strategy can make the mobile robot see into the unexplored space laying beyond target frontier. As a result, the mobile robot can constantly extend its knowledge about the environment. Once arriving at target frontier, the mobile robot performs a 360° sensor sweep using Kinect, computes frontiers and updates occupancy grid map based on Kinect perceptions. And after that, the mobile robot exploits semantic information by using the proposed CNN technique according to panoramic scan, and hereby constructs a semantic map. These processes mentioned above are repeated until the overall environment has completely been explored by mobile robot. In addition, the A*[19] planner is used to plan the shortest collision-free path from robot's current location to the target frontier.

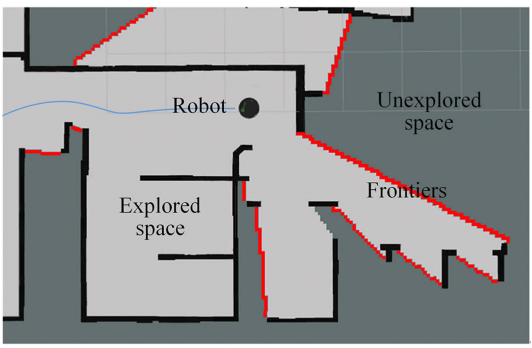


Fig. 8 An example of frontiers in grid map used for exploration

However, the fact that multiple frontiers appear during mobile robot exploration raises the problem of which frontier should move to next for the mobile robot. Therefore, one of the key issues in the context of the frontier-based exploration is how to choose the target frontier from the list of potential candidates. A common approach to determine an appropriate target frontier is to apply a utility-cost strategy that takes into account the distance between robot's current location and frontier position, and the expected information gain.

For each frontier t, the expected utility $U(t)$ is computed in the exploration system, which is defined as

$$U_{utility}(t) = U_{info}(t) - C_{cost}(t) \quad (4)$$

where $U_{info}(t)$ is the approximated information gain related to the number of frontiers that fall within the Kinect sensing range at the frontier t. $C_{cost}(t)$ is the optimal path (the shortest distance) that the robot has to travel from robot's current position to frontier t.

The main goal of the exploration strategy is to navigate mobile robot to the accessible and unvisited frontier with the maximum utility $U_{utility}(t)$. In this way, the frontier-based exploration system can direct mobile robot to the areas that are likely to provide the most new information about the environment, and eventually explore all of the accessible space until all the target frontiers are covered.

According to the aforementioned descriptions, the overall flowchart of the proposed exploration system combined with the CNN-based classifier for mobile robot can be shown in Fig. 9.

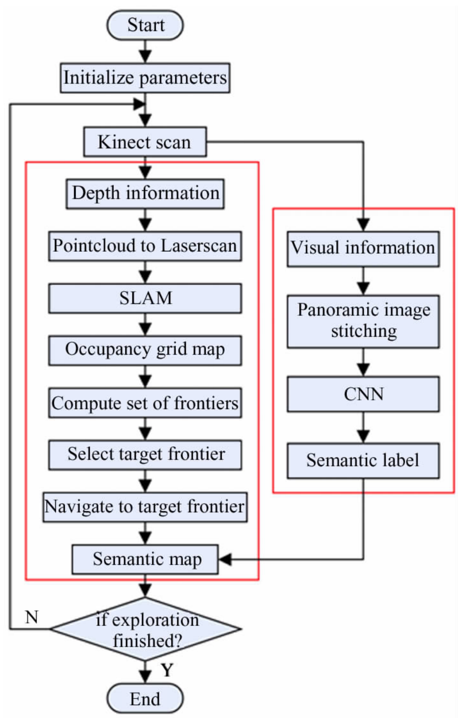


Fig. 9 The flowchart of the proposed exploration system combined with the CNN-based classifier

5 Experiments

To evaluate the performance of the proposed approach described above, several experiments are conducted using a real robot in an office environment.

The robot used in the experiments is Turtlebot2 equipped with odometer, gyroscope and Kinect, as shown in Fig. 10. It should be noted that the exploration system is realized based on open-source exploration package named *frontier _ exploration*[20] in ROS environment[21], and the robot has the means for solving localization and mapping problems by using the laser-based SLAM given Kinect and odometer data. Furthermore, the 3D visualization tool RVIZ[22] is used to visualize the grid map published from Turtlebot2.

Fig. 10 The Turtlebot2 robot equipped with a Kinect

5.1 Data-set of indoor places

In this section, a data-set of indoor places have been created by collecting data in the New Main Building of Beihang University, where the environment contains eight different place categories, namely corridor, doorway, great hall, laboratory, office, seminar room, study room and water room. Each category contains multiple sets of data coming from different locations at one place pertaining to that category, and each set of data contains one panoramic image and 8 partial images.

To collect the data for the data-set, the mobile robot is steered through the indoor environment, and situated in different locations inside each place. At each place, mobile robot takes 64 panoramic scans with Kinect. It should be noted that the locations are spatially distributed inside each place covering the most of the possible situations. Consequently, the data-set contains a total of 512 panoramic images and 4096 partial images of indoor places, which can be used for training and testing the CNN-based classifier.

5.2 CNN setup and training

In this work, CNN are initialized with the default parameters proposed by Krizhevsky, et al.[13]. Aside from replacing the AlexNet-specific 1000-way classification layer (F3) with an 8-way classification layer, the CNN parameter settings are unchanged. Furthermore, the parameters configuration of CNN's solver should be also done for training CNN. Fig. 11 shows an example configuration of solver in the convnet style.

test_iter:10	test_interval:500
base_lr:0.0001	lr_policy: "step"
gamma:0.1	stepsize:100
display:1000	max_iter:100000
momentum:0.9	weight_decay:0.0005
snapshot:50000	solver_mode:GPU

Fig. 11 An example of solver protobuf files in the convnet style

In practical applications, training CNN needs much experience and skill, and it consumes time very much. Furthermore, due to insufficient images, training CNN on the dataset created in this paper may suffer from overfitting. To overcome this problem, therefore, the fine-tuning method is used for training CNN based on the Girshick's implementation of training[23]. In Caffe, the fine-tuning is a very effective way to the adaptation of an existing model to new architectures or data. In this study, training CNN is done through the fast and standard stochastic gradient descent algorithm, and it is carried out on a PC with Intel i7 processor, 64GB memory and a NVIDIA GeForce GPU. And the total training time is approximately 2 days.

In order to evaluate the performance of CNN, both the overall loss and accuracy are calculated during training and testing. Based on the solver setting, the training loss at each iteration and the test loss every 500 iterations are printed and the accuracy every 500 iterations are calculated. Fig. 12 shows the results of training and testing CNN on the created data-set. From

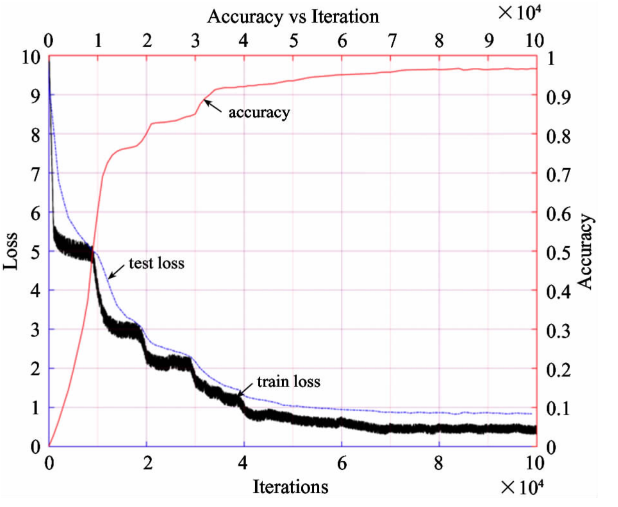


Fig. 12 The results of loss and accuracy during training and testing

the reslults, it can be found that using CNN provides higher accuracy (about 0.974) and lower loss (about 0.325).

Once CNN have been trained, the semantic categorization task can be performed by it.

5.3 Error analysis of the CNN-based classifier for semantic categorization of indoor places

In order to verify the performance of the CNN-based classifier for solving semantic categorization problem, a series of comparative experiments have been conducted using the created data-set. As the baseline, the standard SVM and AdaBoost are used, which classify indoor places based on the geometrical features (2D structure) extracted from LIDAR scans. These features are standard geometrical features often used in shape analysis, which can be found in Ref. [24].

For performing this experiment, the whole data-set has been divided into training set and testing set. The testing set is created by randomly selecting some panoramic images from each category in the data-set, while the rest of panoramic images in the same category are used as training data. In this way, the testing set does not contain any images of a place in the training set.

In this research, the CNN-based classifier is trained by using 320 training examples from the data-set, i.e. each training set contains always eight categories and each category contains 40 panoramic images, and the rest of data are used for test experiments. The test experiments are done repeatedly 20 times, and the categorization results of the three methods with respect to their average correct categorization rate for each category are summarized in Table 1, in which the best results are typed in bold.

As shown in Table 1, the CNN-based classifier performs much better performance of successfully classifying the indoor places than SVM and AdaBoost. The highest accuracy rates of the CNN-based classifier are achieved for the 'Doorway' and 'Laboratory' categories. This can be due to the fact that 'Doorway' and 'Laboratory' categories have distinctive features than the other categories, which can be easily distinguished from panoramic scans. In contrast, the lowest accuracy rates are obtained for the 'Office room' and 'Study room' categories. Although the CNN-based classifier behaves with the worst performance in the accuracy rates of 'Office room' and 'Study room' categories, it presents competitive results compared with SVM and AdaBoost.

Table 1 Comparison of correct categorization rates using CNN, SVM and AdaBoost

Category	CNN	SVM	AdaBoost
Corridor	**94.83 ± 0.83**	89.16 ± 1.66	88.33 ± 2.07
Doorway	**97.35 ± 0.42**	88.33 ± 1.23	89.60 ± 1.75
Great hall	**96.66 ± 0.58**	90.00 ± 1.41	86.66 ± 1.52
Laboratory	**97.33 ± 0.78**	86.63 ± 1.27	84.41 ± 1.90
Office	**91.50 ± 1.26**	84.16 ± 2.49	82.63 ± 2.76
Seminar room	**96.50 ± 0.42**	84.94 ± 1.25	80.00 ± 1.36
Study room	**90.60 ± 1.11**	85.00 ± 2.24	80.83 ± 2.49

From the results, it can be concluded that the CNN-based classifier performs significantly better than SVM and Adaboost with the advantage of providing high precision, which clearly demonstrates the effectiveness of the proposed CNN for addressing semantic categorization problems of indoor environments.

5.4 Semantic categorization of indoor places with CNN for mobile robot exploration

To further demonstrate the performance of the proposed exploration method, a real experiment was carried out with Turtlebot2 robot in the New Main Building of Beihang University. In this experiment, the exploration method with integration of the CNN-based classifier has been tested to categorize the indoor environment.

The experiment result is depicted in Fig. 13, where the developed semantic grid map is viewed with the help of RVIZ. The different color areas on the map indicate different categorization of the corresponding indoor places. Fig. 14 shows the parts (excluding robot path) of the enlarged versions in Fig. 13.

According to the experimental results, it can be concluded that the proposed method with integration of the CNN-based classifier and the frontier-based exploration approach can successfully categorize the different places in indoor environment by using Kinect during mobile robot autonomous exploration.

6 Conclusion

This work presents an approach to classify different places in indoor environments by using Kinect for mobile robot autonomous exploration. This approach applies CNN to form a classifier to perform semantic categorization based on panoramic scans. Moreover, the frontier-based exploration approach is developed to explore the indoor environment and build a semantic map. The experiment results show that the proposed

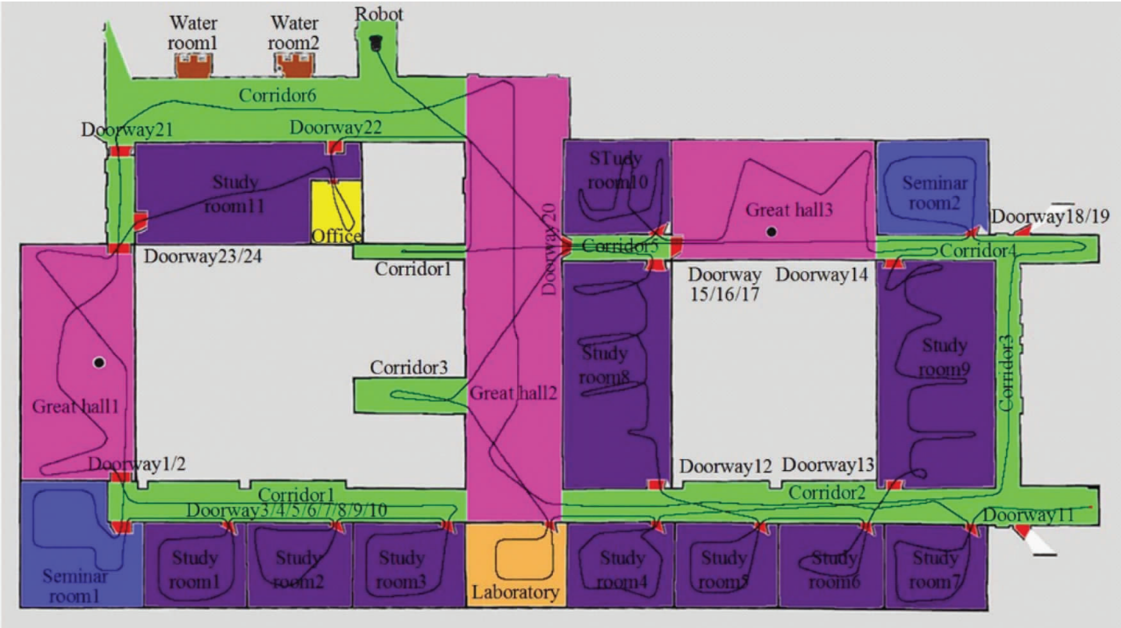
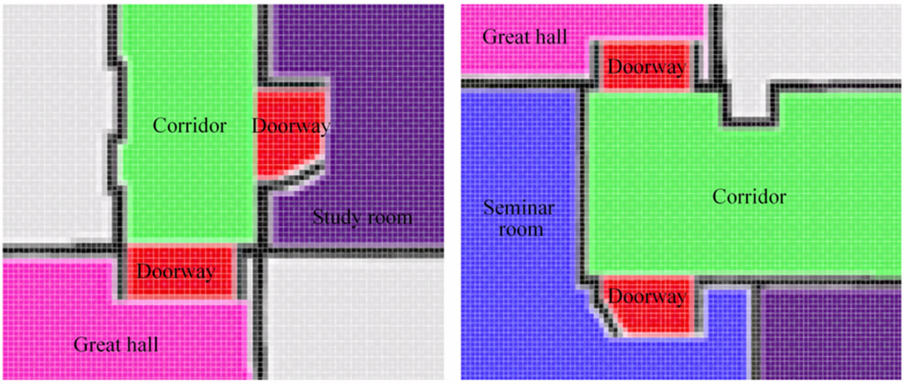


Fig. 13 The semantic map of indoor environment explored by Turtlebot2 robot in RVIZ



(a) Great hall-corridor-study room (b) Seminar room-corridor-great hall

Fig. 14 Parts of the enlarged version of Fig. 13

approach with integration of the CNN-based classifier can effectively categorize indoor places with high reliability during mobile robot autonomous exploration.

Following this work, the CNN-based place categorization method will be further investigated for speeding up mobile robot autonomous exploration.

References

[1] Thrun S, Burgard W, Fox D. Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)[M]. Massachusetts: The MIT Press, 2005. 120-135

[2] Galindo C, Fernández-Madrigal J A, González J, et al. Robot task planning using semantic maps[J]. *Robotics & Autonomous Systems*, 2008, 56(11): 955-966

[3] Mozos O M, Stachniss C, Burgard W. Supervised learning of places from range data using AdaBoost[C]. In: Proceedings of the IEEE International Conference on Robotics and Automation, Barcelona, Spain, 2005. 1730-1735

[4] Sousa P, Araújo R, Nunes U. Real-time labeling of places using support vector machines[C]. In: Proceedings of the IEEE International Symposium on Industrial Electronics, Vigo, Spain, 2007. 2022-2027

[5] Shi L, Kodagoda S, Dissanayake G. Laser range data based semantic labeling of places[C]. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Taipei, China, 2010. 5941-5946

[6] Ranganathan A. PLISS: detecting and labeling places using online change-point detection[J]. *Robotics: Science and Systems VI*, 2010, 32(4): 351-368

[7] Wu J, Rehg J M. CENTRIST: a visual descriptor for scene categorization[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2011, 33(8): 1489-1501

[8] Fazl-Ersi E, Tsotsos J K. Histogram of oriented uniform patterns for robust place recognition and categorization [J]. *International Journal of Robotics Research*, 2012, 31(31): 468-483

[9] Rottmann A, Mozoső M, Stachniss C, et al. Semantic place classification of indoor environments with mobile ro-

bots using boosting[C]. In: Proceedings of the National Conference on Artificial Intelligence, Pittsburgh, USA, 2005. 1306-1311

[10] Pronobis A, Martinez Mozos O, Caputo B, et al. Multimodal semantic place classification [J]. *International Journal of Robotics Research*, 2010, 29(23): 298-320

[11] Mozos O M, Mizutani H, Jung H, et al. Categorization of indoor places by combining local binary pattern histograms of range and reflectance data from laser range finders[J]. *Advanced Robotics*, 2013, 27(18): 1455-1464

[12] Jia Y Q, Shelhamer E, Donahue J, et al. Caffe: convolutional architecture for fast feature embedding [C]. In: Proceedings of the 22nd ACM International Conference on Multimedia, Florida, USA, 2014. 675-678

[13] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]. In: Proceedings of the International Conference on Neural Information Processing Systems, Lake Tahoe, USA, 2012. 1097-1105

[14] Rublee E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to SIFT or SURF[C]. In: Proceedings of the International Conference on Computer Vision, Barcelona, Spain, 2011. 2564-2571

[15] Adel E, Elmogy M, Elbakry H. Image stitching system based on ORB feature-based technique and compensation blending[J]. *International Journal of Advanced Computer Science & Applications*, 2015, 6(9): 55-62

[16] Bovbet P. pointcloud _ to _ laserscan[EB/OL]. http://www.ros.org/wiki/p- ointcloud _ to _ laserscan: Ros.org, 2015

[17] Yamauchi B. A frontier-based approach for autonomous exploration[C]. In: Proceedings of the 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation, Monterey, USA, 1997. 146-151

[18] Yamauchi B. Frontier-based exploration using multiple robots[C]. In: Proceedings of the Second International Conference on Autonomous Agents, Minneapolis, USA, 1998. 47-53

[19] Li X, Wei R X, Wang Z K. Three-dimension path planning for UAV using improved A* algorithm in complicated threat environment[J]. *High Technology Letters*, 2011, 17(1): 13-18

[20] Bovbet P. frontier _ exploration[EB/OL]. http://wiki.ros.org/frontier _ ex-ploration: Ros.org, 2014

[21] Quigley M, Conley K, Gerkey B P, et al. ROS: an open-source robot operating system[C]. In: Proceedings of the IEEE International Conference on Robotics and Automation, Kobe, Japan, 2009. 5

[22] Hershberger D, Gossow D, Faust J. rviz[EB/OL]. http://wiki.ros.org/rviz/: Ros.org, 2014

[23] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014. 580-587

[24] Óscar M M. Semantic Learning of Places from Range Data [M]. Berlin/Heidelberg: Springer-Verlag, 2010. 15-34

Li Guangsheng, born in 1985. He is currently a Ph. D candidate at Beihang University. He received his M. S. and B. E. degrees in Mechatronics Engineering at Beihang University in 2010 and 2013 respectively. His current research interests include embedded robot control and multi-robot coordinated technology.