Building semantic grid maps for domestic robot navigation

Xianyu Qi¹, Wei Wang¹, Mei Yuan², Yuliang Wang¹, Mingbo Li², Lin Xue² and Yingpin Sun²

Abstract
This article proposes a semantic grid mapping method for domestic robot navigation. Occupancy grid maps are sufficient for mobile robots to complete point-to-point navigation tasks in 2-D small-scale environments. However, when used in the real domestic scene, grid maps are lack of semantic information for end users to specify navigation tasks conveniently. Semantic grid maps, enhancing the occupancy grid map with the semantics of objects and rooms, endowing the robots with the capacity of robust navigation skills and human-friendly operation modes, are thus proposed to overcome this limitation. In our method, an object semantic grid map is built with low-cost sonar and binocular stereovision sensors by correctly fusing the occupancy grid map and object point clouds. Topological spaces of each object are defined to make robots autonomously select navigation destinations. Based on the domestic common sense of the relationship between rooms and objects, topological segmentation is used to get room semantics. Our method is evaluated in a real homelike environment, and the results show that the generated map is at a satisfactory precision and feasible for a domestic mobile robot to complete navigation tasks commanded in natural language with a high success rate.

Keywords
Occupancy grid map, semantic grid map, domestic robot navigation, sonar sensors, stereovision

Date received: 19 April 2018; accepted: 1 December 2019

Introduction
Occupancy grid map is one of the most important environment representation methods in mobile robotics. Unlike topological map modeling the environment as a graph,¹,² the grid map discretizes the space as many grids with unknown, free, and occupied attributions.³,⁴ Recently, the methods used to create occupancy grid maps for 2-D small-scale environments have matured.⁵ Based on these maps, the robots can successfully complete point-to-point navigation tasks.⁶ However, when it comes to using the grid maps in the real domestic scene, they are lack of semantic information for end users to order the robots to complete navigation tasks conveniently. For example, if an end user asks a robot to navigate to the position of a bed, he/she should know where the bed is in the grid map and manually specify a free grid as the goal point. Unfortunately, this approach is not feasible for two reasons: (1) most of the end users have little knowledge of robotics and cannot operate the map easily; (2) it cannot be accepted by the end users to repeat this operation all the time. For that, semantic map,⁷,⁸ which describes not only the geometric arrangement of the

¹ Robotics Institute, Beihang University, Beijing, China
² Beijing Evolver Robotics Technology Co., Ltd, Beijing, China

Corresponding author:
Wei Wang, Robotics Institute, Beihang University, Room 314, Block A, The New Main Building, No. 37 XueYuan Road, HaiDian District, Beijing 100191, China.
Email: wangweilab@buaa.edu.cn

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
In recent years, there are many authors focus on creating semantic grid maps for robot navigation in domestic environments. Our method enhances the occupancy grid map with semantic concepts of objects and rooms. In this way, robots can not only utilize traditional robust navigation algorithms but also make natural interactions with end users.

To build this map, we first create an occupancy grid map with sonar sensors. Point clouds of objects are then built by using object detection with a stereovision. A filtering approach based on the information of the grid map and the size of objects is used to remove outliers, and density-based spatial clustering is adopted to get the final point clouds. We then employ minimum bounding rectangles (MBRs) to approximatively represent the objects and create the topological spaces of objects. The concepts of rooms are acquired by spectral clustering and region growing with domestic common sense. The experimental results show the semantic grid mapping approach is affordable to represent the environment and enable domestic robots with the capability of robust, efficient, and human-friendly navigation skills.

The proposed method has the following contributions:

- A novel method for building domestic semantic maps with low-cost sonar and binocular vision sensors;
- Object and room semantics are contained;
- Make end users order the robots to complete navigation tasks conveniently;
- Take advantage of the information from occupancy grid map and domestic common sense to remove outliers;
- Build topological spaces of objects for robots to autonomously select navigation goals;
- Use domestic common sense to merge segment results.

The remainder of this article is organized as follows. The “Related work” section presents related work about semantic mapping. Then, the “Grid map with object semantics” section introduces our semantic grid mapping method. Based on the generated object grid map, the “Semantic concepts of rooms” section subsequently describes our method for acquiring room concepts. The suitability of our approach is demonstrated in the “Experimental results” section. Finally, the “Conclusion” section concludes this article.

**Related work**

In recent years, there are many authors focus on creating semantic maps for robot navigation. According to the different goals of semantic mapping methods, they can be divided into two classes: One aims to augment metric/topological models with semantic labels like types of objects or rooms. For example, on the basis of the grid map, Mozos et al.\(^9\) added the semantic labels of the scene through supervised learning. Using a laser sensor to construct a grid map and a visual sensor to identify objects, Meger et al.\(^10\) constructed a map containing both space arrangements and object semantics. Pillai and Leonard\(^11\) combined monocular simultaneous localization and mapping (SLAM) and object recognition to create a point cloud map with object semantics.

The other aims to study semantic map representation and reasoning. These methods use multilayer maps to represent the environment, providing the robot with capability of complex task planning, and implicit knowledge reasoning. For instance, Ruiz-Sarmiento et al.\(^12\) used RGB-D sensors with the conditional random field to construct a multiversal semantic map containing spatial relationships and symbolic grounded uncertainty. Pronobis and Jensefelt\(^13\) fused multiple sensor information and used a probability method to create a multilayer map. Zender et al.\(^14\) conceptualized the environment at different levels of abstraction.

The key part of these methods is to get the semantic information of the environments. For that, object recognition,\(^15,16\) scene recognition,\(^17,18\) database of geometric model,\(^19,20\) human–computer interaction,\(^21,22\) artificial landmark,\(^23,24\) are used to get semantic concepts. In recent years, with the advancement of deep learning, the methods based on neural networks occupy a leading position. The interested reader can refer to the work by Kostavelis and Gasteratos\(^25\) for a comprehensive review of semantic mapping approaches for robotic tasks and the work by Landesdeld et al.\(^26\) for spatial reasoning and interacting.

Although these methods successfully created semantic maps, they only describe the occupied grids in occupancy grid maps and could not be used for the robot navigation as the path planning algorithms need the goal points located on free grids. Moreover, these methods do not consider the computational resources consumption and the price of robot sensors which are of great importance for domestic robots.

As for robot navigation, classical methods, such as AStar,\(^27\) dynamic window approach (DWA)\(^28\) can make the robots complete point-to-point navigation tasks successfully while lacking of semantic information. Inspired by the fact that human beings and higher animals only need significant landmarks and spatial relationships to navigate, some authors try to endow the robots with the capacity of cognitive navigation. For example, Ko et al.\(^29\) first built a topological-semantic-metric map, then proposed a human-like semantic navigation method for mobile robots. Crespo et al.\(^30\) proposed a semantic relational model for robot path planning. However, as these methods do not yet have high robustness and real-time performance, they cannot be applied in the real scene application at the current stage.
proposed by Lee et al., which is based on the approximate laser information, the map quality may be severely degraded. Since region-based fully convolutional networks can maintain or even improve the accuracy of position and classification of objects, as well as has high computational efficiency, we use it to detect the object. Specifically, we use it to detect the left image and get the object bounding boxes. Then, the sub-map \( M_2 \) is created by using the reprocessed data. Finally, the grid map \( M \) is the union of sub-map \( M_1 \) and \( M_2 \).

Object detection

Our method uses object detection to provide semantic information. In domestic environments, there are a large number of object classes. We first have to determine the categories of detected objects. As we extract the semantic information of the rooms based on the common sense of the relationship between rooms and objects, and consider that the objects are only used for robot navigation, we determine to detect 10 classes of objects, including: sofa, coffee table, TV cabinet, TV, bookcase, bed, bedside table, dining table, and wardrobe. Since region-based fully convolutional networks can maintain or even improve the accuracy of position and classification of objects, as well as has high computational efficiency, we use it to detect the object. Specifically, we use it to detect the left image and get the object bounding boxes. We then triangulate the feature points in these regions. In addition, since open source data sets cannot meet our requirements for domestic robot navigation, we train and test the net using our data set and the confusion matrix of our object detection is shown in Figure 2.
Original point clouds with object semantics

Figure 3 shows the coordinate frames of our system. Since the domestic environment is usually a small-scale scene and the resolution of the grid map is more than 5 cm, we directly employ triangulation method to acquire original point clouds instead of the state estimation method like ORB-SLAM33 or LSD-SLAM.34 In addition, to simplify system complexity and speed up calculations, the robot does not need to visit the complete environment and record all images. So we make the robot navigate to specified points to take images. Specifically, based on the occupancy grid map, the robot moves to the points specified by users, and turns around to record images and corresponding poses. Object detection is then used to acquire the region of interest (ROI) of objects. For the points in ROI, ORB35 is exploited to find corresponding points \((u_L, v_L)\) and \((u_R, v_R)\) in left and right images. The coordinates of the feature point in the camera frame \(P_{CL}^i = (x_{CL}^i, y_{CL}^i, z_{CL}^i) \in \mathbb{R}^3\) are then computed as

\[
\begin{align*}
  x_{CL}^i &= f_x \left( u_L - u_L^0 \right) / b \\
  y_{CL}^i &= f_x \left( u_R - u_R^0 \right) / f_x \left( z_{CL}^i \right) \\
  z_{CL}^i &= b \left( u_L - u_L^0 \right) / \left( u_R - u_R^0 \right) - (u_L - u_L^0) \\
  y_{CL}^i &= 0
\end{align*}
\]

where \(u_L^0\) and \(u_R^0\) are the coordinates of axes \(u\) of principal points in pixels, \(f_x\) is the size of unit length in horizontal pixel, and \(b\) is the length of baseline. Notice that the point clouds are projected onto the grid map, thus \(y_{CL}^i\) is simply set to zero. Thus, the point \(P_{W}^i = (x_{W}^i, y_{W}^i, z_{W}^i) \in \mathbb{R}^3\) with object label in the world frame is computed as

\[
P_{W}^i = W T_R R_{CL}^T P_{CL}^i
\]

As shown in Figure 4, the original points include many outliers and object detection cannot distinguish different instances of the same concept. The information of grid map, size of objects, and density-based clustering are used to address these problems. As the objects are obstacles and they can only locate on occupied grids, we first exploit the free grids to remove some outliers. Then, the occupied grids are abstracted to form a set \(O\) and the shortest distance between point \(P_{W}^i\) and occupied grid point \(O_{W}^j \in O\) is given as

\[
q_i = \min_i \left\| P_{W}^i - O_{W}^j \right\|_2
\]

Next, the minimum length \(m_{oi}\) of the object of \(P_{W}^i\) is used as a prior to defining the outlier

\[
c(P_{W}^i) = \begin{cases} \text{inlier}, & \text{if } q_i \leq m_i \\ \text{outlier}, & \text{otherwise} \end{cases}
\]

Finally, density-based clustering36 is employed to further filter out the outliers and distinct different instances with the same concept.

Object representation

While point clouds use feature points to represent objects, we can take a compact way of representation. Consider the
detected objects can be approximated as rectangles and are always parallel to walls, MBRs are thus used to represent objects. In addition, as the MBRs of objects usually include occupied grids that cannot be goal points for navigation tasks, it needs to define the free spaces nearby the objects to make the robot autonomously select goal points. We thus expand the MBRs to build the topological spaces of objects.

Figure 5 shows our expanding approach uses a three triple \((e, d, l)\) to determine the topological space, where \(e\) represents the expanding edge, \(d\) is the expanding direction, \(l\) represents the expanding length. As many objects are located on unknown grids, it is hard to directly to expand the MBRs of objects. We employ robot positions that can observe the objects as additional information to define topological spaces. Suppose the center of MBR of object \(i\) is \(P^W_{ci}\), the vertexes of MBR used to determine the edge \(r_j\) is defined as \(P^W_{aj}\) and \(P^W_{bj}\), and the robot pose that can observe the object is \(P^W_{ri}\). The line segments \(s_i\) and \(s_{ij}\) are defined as

\[
s_i = P^W_{ci} - P^W_{ri}
\]

(7)

\[
s_{ij} = P^W_{ri} - P^W_{bj}
\]

(8)

Based on the result of expanding edge \(e_{oi}\), the expanding direction \(d_{oi}\) is given as

\[
e_{oi} = \begin{cases} 
  r_j, & \text{if } \frac{s_{ij} \times r_j}{r_j \times s_i} \in (0, 1) \\
  \text{null}, & \text{otherwise}
\end{cases}
\]

(9)
Semantic concepts of rooms

In order to acquire concepts of rooms, we first use spectral clustering to coarsely segment the grid map, and then use the topological spaces of objects and domestic common sense to merge segment results.

Normalized graph cut and spectral clustering

Suppose an undirected weight graph $G(V,E)$ is composed of nodes $V = \{V_1, V_2, \ldots, V_n\}$, edges $E = \{E_1, E_2, \ldots, E_m\}$ and the weight between $V_i$ and $V_j$ is $W(i,j)$, we want to divide $G$ into two subgraphs $G_1$ and $G_2$, where $G_1 \cup G_2 = G$ and $G_1 \cap G_2 = \emptyset$. To get the best division, the norm cut criteria is given as

$$Ncut(G_1, G_2) = \arg \min_{G_1, G_2} \text{cut}(G_1, G_2)$$

where $Ncut(G_1, G_2)$ is objective function

$$Ncut(G_1, G_2) = \frac{\text{cut}(G_1, G_2)}{\text{asso}(G_1, G)} + \frac{\text{cut}(G_2, G_1)}{\text{asso}(G_2, G)}$$

$\text{cut}(G_1, G_2)$ represents the dividing cost function

$$\text{cut}(G_1, G_2) = \sum_{u \in G_1, v \in G_2} W(u,v)$$

and $\text{asso}(G_1, G)$ is the weighted sum of edges of $G_1$

$$\text{asso}(G_1, G) = \sum_{u \in G_1, v \in G} W(u,v)$$

However, it is an NP-hard problem to find the minimum value of $Ncut$. Therefore, spectral clustering is used to get an approximate solution as following steps:

1. Down sampling the grid map into a graph $G(V,E)$.
2. Compute affinity matrix $W$ by

$$W(i,j) = \begin{cases} \frac{1}{|x_i - x_j|} & \text{if } e_{ai}(y) = 0 \\ \frac{1}{|y_i - y_j|} & \text{if } e_{ai}(x) = 0 \\ 0 & \text{otherwise} \end{cases}$$

3. Define $D = D^{-\frac{1}{2}}WD^{-\frac{1}{2}}$ and $D$ is a diagonal matrix with $D(i,i) = \sum_j W(i,j)$.
4. Select a desired number of subgraphs $k$.
5. Find $x_1, \ldots, x_k$ the largest eigenvectors of $L$, and form the matrix $X = [x_1, \ldots, x_k] \in \mathbb{R}^{n \times k}$.
6. Set $Y \in \mathbb{R}^{k \times k}$ to be $X$ with rows renormalized to have unit length, such that $Y_{ij} = X_{ij}/(\sum_j X_{ij}^2)$.
7. Use k-means clustering on each row of $Y$.
8. Assign $V_i$ to cluster $j$ if and only if row $i$ of $Y$ is assigned to cluster $j$.
9. Assign other grids to cluster $k$ with $n$-nearest neighbor if and only if the most neighbors are in the cluster $k$.

Using common sense to improve the results of segmentation

Based on the results of spectral clustering, the grid map can be segmented into $k$ clusters. Figure 6 shows that some of them contain object semantics, and some do not. As spectral clustering is one of unsupervised learning methods, the number of clusters $k$ is a predefined parameter need to be tuned in different domestic environments. In addition, although the number $k$ is equal to the number of rooms, the spectral clustering is not able to get the correct results. It thus needs to tune many magic parameters manually. To overcome this limitation, we choose a larger number $k$ to coarsely segment grid map and use object semantic map and domestic common sense to merge the segment results.

According to the clusters $k$, a region adjacency graph $G'(V', E')$ is formed with $V' = \{V'_1, \ldots, V'_q\}$, and $E' = \{E'_1, \ldots, E'_q\}$, and the weights $W'(i,j)$ between node $V'_i$ and $V'_j$ is computed as

$$W(i,j) = \begin{cases} \frac{1}{\sigma_i} ||V_i - V_j||^2 & \text{if } i \neq j \text{ and } l_{ij} \text{ in free space} \\ 0 & \text{otherwise} \end{cases}$$

where $\sigma_i$ is defined as the median of the $n$ nearest neighbors.
$W(i,j) = \begin{cases} 
\# \text{ of adjacent grids} & V_i, V_j \text{ are adjacent} \\
0 & \text{otherwise} 
\end{cases}$

Figure 7. Experimental setup. (a) A FABO robot with 13 sonar sensors and a stereovision. (b) Experimental environment.

Figure 8. Experimental results of occupancy grid maps. (a) Sonar-based grid map. The blue dashed area indicates that the map does not model the environment completely because the size of FABO robot is larger than TurtleBot. The occupied grids in the red line rectangle represent robot charging pile. (b) Laser-based grid map. The black, white, and grey colors represent occupied, free, and unknown areas, respectively. The red dashed areas in both figures indicate that the sonar sensors can completely model the area but the laser cannot.

Experimental results

This section presents the experimental results of the proposed semantic grid mapping method and navigation.
performance using the generated map. Figure 7 shows that our experiments have been conducted using a FABO robot in a homelike environment containing three rooms and some domestic objects. The robot is a differential drive robot equipped with 13 sonar sensors and a stereovision. In addition, it also has the capacity of speech recognition for end users to command navigation tasks in natural language to evaluate the navigation performance.

The robot was first manually controlled at an average speed of about 0.2 m/s while acquiring sensor data at a rate of 4 Hz and recording the coordinates of designated positions used to take images. Then, an occupancy grid map is built and the grid size of the generated grid map was 12 cm x 12 cm. Based on the grid map, the robot autonomously went to each recording positions to collect images and built a semantic grid map. Finally, the robot was asked to six different places by an end user in natural language.

**Semantic grid map**

The performance of sonar grid mapping was compared with an open source laser-based SLAM method called Cartographer. Since the FABO does not have a laser sensor, a TurtleBot equipped with a SLAMTEC RPLIDAR A2 was used to build the laser-based grid map. Figure 8 shows the results of occupancy grid maps. Although the laser-based grid map captures the overall shape of the environment more successfully and provides the environment information more detailedly, the sonar-based grid map is able to make the robot navigation successfully. In addition, the 2-D laser sensor scans in a two-dimensional plane, while the sonar projects a three-dimensional cone. Thus, as shown the dashed red areas in Figure 8(a) and (b), the laser sensor cannot detect obstacles that are above or below the laser plane, but still detectable by the sonar. These results further provide the evidence of the advantages of sonar sensors in domestic environments. In addition, we use the sonar maximum sound pressure map to evaluate the sonar sensor errors and its improvement. There are 6975 sonar readings and 22.29% of them are correct. Namely, the method can filter out 77.71% sonar readings, which are incorrect. Based on the sonar-based grid map, Figure 9 shows the process of the proposed object semantic mapping method. Using object detection and triangulation, the original object point clouds are generated as described in Figure 9(a). Most of the object points are locating on the unknown areas for the reason that objects are obstacles in the occupancy grid map.

Based on our filtering and clustering method, the outliers are easily filtered out and the different instances of the same concepts can be distinct. Note that the size of the robot should be considered in the grid map, so the two free grids that near the occupied grids are treated as the occupied grids as shown in Figure 9(b). Figure 9(c) shows that MBRs and expanding rectangles are used to represent the object compactly. Although some objects such as bed1 and sofa2 slightly deviate from the corresponding positions in the grid map, it does not make a significant impact on the generated topological space of these objects. Therefore, the experimental results of object semantic mapping show that the proposed method can successfully add the object semantics on the grid map.

**Figure 9.** The process of object semantic grid mapping method. (a) Original object point clouds. (b) The object point clouds generated by using filtering and clustering. (c) Representation of objects and the corresponding topological space.
Therefore, the experimental results of object semantic mapping show that the proposed method can successfully add the object semantics on the grid map.

**Room concepts**

Figure 10 shows the results of topological segmentation of rooms. Even if we know the correct result is that the environment contains three rooms, Figure 10(a) and (b) indicates it is hard for the topological segmentation to divide the map into correct rooms. We thus need to tune the magic parameters to get the correct results. However, based on the object grid map and domestic common sense, we only need to tune the cluster number \( k \). The domestic common sense used in our experiments includes: the living room contains a TV, TV cabinet, sofa, tea table; the bedroom contains a bed, desk, and wardrobe. One example of our segment and merging results is shown in Figure 10(c) and (d). To prove the effectiveness of our method, we test two different parameters and varied the cluster number \( k \) from 3 to 20. The former test uses the parameters of \( k_{nn} = 3 \), scaling parameter \( k_s = 7 \) and the latter is \( k_{nn} = 3 \) and \( k_s = 3 \). Figure 11 shows that the topological segmentation based on the object semantic grid map can successfully get the correct room concepts by only changing the cluster number \( k \).

**Navigation**

To evaluate the effectiveness of proposed semantic grid map, a user in turn commanded the robot go to TV cabinet, sofa1, desk, wardrobe, bed1, sofa2. As shown in Figure 12(a), the user asked the robot to go to the TV cabinet. According to the generated semantic grid map and consider the factor of radius of FABO, we inflated the occupied grids with two free grids. In order to increase the
diversity of navigation behavior, the goal points were generated using random selection in topological spaces of objects except the inflated grids. AStar was then used to generate navigation paths. Figure 12(b) shows the final pose for the first navigation task. The whole goal points and paths for the navigation tasks are described in Figure 12(c).

Since we only used the odometry to build the semantic grid map and navigate in it, we couldn’t get the ground truth. We thus adopted the maximum pose error to evaluate the navigation perforation. We used the manual loop closing to make the robot start from the charging pile and finally return it to get an optimal trajectory. Specifically, we recorded the odometry readings and use g2o\(^3\) to optimize the initial trajectory. Please note that since the pose graph only has one loop position, the error will propagate far away from the charging pile. However, we compare the maximum error mapped onto trajectory pose error and the moving distance is not long, this method can get a relatively accurate result.

As can be seen from Figure 13, the maximum error of the robot is 9.6 cm. The resolution of grid map is 12 cm \(\times\) 12 cm and the error is no more than 1 grid size. Additionally, from the perspective of indoor applications, as long as the position of the robot near the specified object, and there is no collision during the navigation, it can be regarded as successful navigation. Thus, the FABO successfully completed the navigation task. However, the uncertainty of the robot becomes larger as the moving distance increases. This may make the robot cannot complete other navigation tasks. To solve this problem, we will use the charging pile as a global landmark to reduce the uncertainty of the robot in our future work.

Table 2 presents the time evaluation of this navigation task, which includes planning time and execution time. In terms of planning time, planning the third path is the most time consuming, which costs 0.5121 s. The sixth path costs 0.3677 s, and other paths are around 0.1 s. As the domestic environment is small scale, it can meet the requirements of actual applications. In terms of execution time, it is much longer. One reason is that we set the average moving speed
of the robot to 0.2 m/s. If we want to reduce the execution time, we can increase the moving speed of the robot. However, for indoor service robots, security is first considered.

In general, the results indicate that the robot can successfully complete the navigation tasks as the robot was in 2-D small-scale environment. At the same time, the users can conveniently command the robot in natural language.

**Conclusion**

As more and more robots are coming into domestic environments, they need semantic knowledge to make human beings operate conveniently. In this work, we present a novel semantic mapping method for domestic navigation application. Our method enhances the occupancy grid map with semantics of objects and rooms, endowing the robots with robust navigation skills and the semantic knowledge about environments. Concretely, the proposed method first builds occupancy grid map and object point clouds respectively and then fuse them correctly to get an object grid map. Based on the generated map and domestic common sense, topological segmentation is used to get room concepts. The experiments of semantic mapping and navigation were conducted with a FABO robot in a homelike environment. The experimental results prove the mapping method has a satisfactory precision to represent the environment and makes the robot can complete domestic navigation tasks in a human-friendly operation way.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is supported by the National Natural Science Foundation of China (nos 91748101).

**ORCID iD**

Xianyu Qi https://orcid.org/0000-0002-6636-374X

**References**

1. Remolina E and Kuipers B. Towards a general theory of topological maps. *J Artif Intell* 2004; 152(1): 47–104.
2. Ranganathan A, Menegatti E, and Dellaert F. Bayesian inference in the space of topological maps. *IEEE Trans Robot* 2006; 22(1): 92–107.
3. Elfes A. Sonar-based real-world mapping and navigation. *IEEE J Robot Autom* 1987; 3(3): 249–265.
4. Thrun S. Learning occupancy grid maps with forward sensor models. *Auton Robot* 2003; 15(2): 111–127.
5. Kostavelis I and Gasteratos A. Semantic maps from multiple visual cues. *Expert Syst Appl* 2017; 68: 45–57.
6. Galindo C, Saffiotti A, Coradeschi S, et al. Multi-hierarchical semantic maps for mobile robotics. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Edmonton, Canada, 2–6 August 2005, pp. 3492–3497.
7. Nuchter A and Hertzberg J. Towards semantic maps for mobile robots. *Robot Auton Syst* 2008; 56(11): 915–926.
8. Mozos OM, Triebel R, Jensfelt P, et al. Supervised semantic labeling of places using information extracted from sensor data. *Robot Auton Syst* 2007; 55(5): 391–402.
9. Pillai S and Leonard J. Monocular SLAM supported object recognition. In: *Proceedings of robotics science and systems*, Rome, Italy, 13–17 July 2015.
10. Ruiz-Sarmiento JR, Galindo C, and Gonzalez-Jimenez J. Building multiversal semantic maps for mobile robot operation. *Knowl Based Syst* 2017; 119: 257–272.
11. Pronobis A and Jensfelt P. Large-scale semantic mapping and reasoning with heterogeneous modalities. In: *2012 IEEE international conference on robotics and automation (ICRA)*, St Paul, MN, USA, 14–19 May 2012, pp. 3515–3522.
15. Rusu RB. Semantic 3D object maps for everyday manipulation in human living environments. *Künstl Intell* 2010; 24(4): 345–348.

16. Yang J, Zhang J, Wang G, et al. Semantic map building based on object detection for indoor navigation. *Int J Adv Robot Syst* 2015; 12(12): 178–194.

17. Sunderhauf N, Dayoub F, McMahon S, et al. Place categorization and semantic mapping on a mobile robot. In: *IEEE international conference on robotics and automation (ICRA)*, Stockholm, Sweden, 16–21 May 2016, pp. 5729–5736.

18. Liu Z and Von Wichert G. Extracting semantic indoor maps from occupancy grids. *Robot Auton Syst* 2014; 62(5): 663–674.

19. Strasdat H, Newcombe RA, Salas-Moreno RF, et al. SLAM++: simultaneous localization and mapping at the level of objects. In: *IEEE conference on computer vision and pattern recognition (CVPR)*, Portland, Oregon, 25–27 June, 2013, pp. 1352–1359.

20. Gunther M, Wiemann T, Albrecht S, et al. Model-based furniture recognition for building semantic object maps. *J Artif Intell* 2017; 247: 336–351.

21. Gemignani G, Capobianco R, Bastianelli E, et al. Living with robots: interactive environmental knowledge acquisition. *Robot Auton Syst* 2016; 78: 1–16.

22. Walter MR, Hemachandra SM, Homberg BS, et al. A framework for learning semantic maps from grounded natural language descriptions. *Int J Robot Res* 2014; 33(9): 1167–1190.

23. Wu H, Tian GH, Li Y, et al. Spatial semantic hybrid map building and application of mobile service robot. *Robot Auton Syst* 2014; 62(6): 923–941.

24. Li X, Zhang X, Zhu B, et al. A visual navigation method of mobile robot using a sketched semantic map. *Int J Adv Robot Syst* 2012; 9: 1–15.

25. Kostavelis I and Gasteratos A. Semantic mapping for mobile robotics tasks: a survey. *Robot Auton Syst* 2015; 66: 86–103.

26. Landsiedel C, Rieser V, Walter M, et al. A review of spatial reasoning and interaction for real-world robotics. *Adv Robotics* 2017; 31(5): 222–242.

27. Hart PE, Nilsson NJ, and Raphael B. A formal basis for the heuristic determination of minimum cost paths. *IEEE Trans Syst Sci Cybern* 1968; 4(2): 100–107.

28. Fox D, Burgard W, and Thrun S. The dynamic window approach to collision avoidance. *IEEE Robot Autom Mag* 1997; 4(1): 23–33.

29. Ko DW, Yi C, and Suh IH. Semantic mapping and navigation: a Bayesian approach. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Tokyo, Japan, 3–8 November 2013, pp. 2630–2636.

30. Crespo J, Barber R, and Mozos OM. Relational model for robotic semantic navigation in indoor environments. *J Intell Robot Syst* 2017; 86(3–4): 617–639.

31. Lee K, Lee SJ, and Wan KC. Enhanced maximum likelihood grid map with reprocessing incorrect sonar measurements. *Auton Robot* 2013; 35(2–3): 123–141.

32. Dai J, Li Y, He K, et al. R-FCN: object detection via region based fully convolutional networks. In: *Advances in neural information processing systems*, Barcelona, Spain, 5–11 December 2016, pp. 379–387.

33. Mur-Artal R, Montiel JMM, and Tardos JD. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Trans Robot* 2015; 31(5): 1147–1163.

34. Engel J, Schöps T, and Cremers D. LSD-SLAM: large-scale direct monocular SLAM. In: *European conference on computer vision (ECCV)*, Zurich, Switzerland, 6–12 September 2014, pp. 834–849.

35. Rublee E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to SIFT or SURF. In: *2011 international conference on computer vision*, Barcelona, Spain, 6–13 November 2011, pp. 2564–2571.

36. Kriegel HP, Kröger P, Sander J, et al. Density-based clustering. *Wires Data Min Knowl* 2011; 1(3): 231–240.

37. Shi J and Malik J. Normalized cuts and image segmentation. *IEEE Trans Pattern Anal Mach Intell* 2000; 22(8): 888–905.

38. Hess W, Kohler D, Rapp H, et al. Real-time loop closure in 2D LIDAR SLAM. In: *IEEE international conference on robotics and automation (ICRA)*, Stockholm, Sweden, 16–21 May 2016, pp. 1271–1278.

39. Kummerle R, Grisetti G, Strasdat H, et al. g2o: a general framework for graph optimization. In: *IEEE international conference on robotics and automation (ICRA)*, Shanghai, China, 9–13 May 2011, pp. 3607–3613.