Probabilistic 3D Semantic Map Fusion Based on Bayesian Rule

Yufeng Yue, Ruilin Li, Chunyang Zhao, Chule Yang, Jun Zhang, Mingxing Wen
Guohao Peng, Zhenyu Wu and Danwei Wang

***Abstract*— Performing collaborative semantic mapping is a critical challenge for multi-robot systems to maintain a comprehensive contextual understanding of the surroundings. In this paper, a novel hierarchical semantic map fusion framework is proposed, where the problem is addressed in low-level single robot semantic mapping and high level global semantic map fusion. In the single robot semantic mapping process, Bayesian rule is used for label fusion and occupancy probability updating, where the semantic information is added to the geometric map grid. High level global semantic map fusion covers decentralized map sharing and global semantic map updating. Collaborative semantic mapping is conducted in two scenarios, that is, NTU dataset and the KITTI dataset. The results show the high quality of the global semantic map, which demonstrates the utility and versatility of 3D semantic map fusion algorithm.**

I. Introduction

With the development in the domain of robotics and deep learning, the acquisition of semantic information becomes possible. To enhance the perception capability of intelligent mobile robots, it is critical to build the semantic map with high-level semantic information. Semantic mapping consists of geometry mapping [1] and semantic information [2], which is crucial for many applications such as localization, path planning and navigation [3]. Robots not only need to obtain the 3D geometry information of the environment to avoid obstacles [4], [5] but also need to recognize objects and scenes for high-level tasks [6] [7]. With the development of deep learning-based semantic segmentation algorithms, the semantic map will have the characteristics of semantic labels, spatial scope and interval attributes, which can enable robots to perform more intelligent tasks.

However, most researchers focus more on single robot algorithms for semantic segmentation [8] and mapping [9]. For multi-robot systems, approaches have been proposed to perform collaborative geometry mapping [10] [11]. Until now, comprehensive analysis and modeling of the deployment of the semantic mapping for multi-robot are still not available. Therefore, there is a gap in performing collaborative semantic mapp fusion.

In this paper, a novel hierarchical semantic map fusion framework is proposed. To the best of our knowledge, this is the first attempt to address multi-robot semantic mapping problem. CNN model is deployed on the robot platform, performing the real-time semantic segmentation. Based on the geometry 3D reconstruction algorithms, semantic information obtained by robots is added to the map by applying the Bayes' rule, building the semantic map. Then, the 3D semantic map fusion is performed to integrate maps built by different robots into a complete map. In conclusion, Bayesian rule is applied to achieve both single-robot and multi-robot level semantic reconstruction. The main contributions are listed as follows:

- A novel hierarchical probabilistic semantic map fusion framework is proposed to address the problems in both low-level and high-level semantic mapping.
- Bayesian rule is applied to perform label fusion and occupancy probability updating, fusing semantic information to geometry mapping.
- A real robot system is developed that implements semantic understanding and global semantic reconstruction, which is tested on NTU dataset and KITTI dataset.

*The research was partially supported by the ST Engineering NTU Corporate Lab through the NRF corporate lab@university scheme.

Yufeng Yue, Ruilin Li, Chunyang Zhao, Chule Yang, Jun Zhang, Mingxing Wen, Guohao Peng, Zhenyu Wu and Danwei Wang are with School of Electrical and Electronic Engineering, Nanyang Technological University(NTU), Singapore. Email: yyue001@e.ntu.edu.sg.

II. Related Work

A. Semantic Segmentation

The concept of deep convolutional neural network (DCNN) is firstly proposed in 1998 and applied to handwritten file recognition. In recent years, deep convolutional neural network (DCNN) has gradually become mainstream in high-dimensional applications, achieving the best results in a series of computer vision tasks such as image classification, segmentation, and object detection [12].

Compared with traditional visual algorithms such as N-cut [13] and Grab cut [14], DCNN achieves good results in its end-to-end approach with fully connected networks (FCN) [15]. At present, Deeplab [16] is the best model that has excellent performance in the field of semantic segmentation. Therefore, in our robot platform, Deeplab is employed to process the image data collected by mobile robots. In the model of Deeplab V3 [17], batch normalization is included within ASPP which uses multiple parallel atrous convolutions with different rates to process the feature map so that multiple-scale information can be effectively captured. Deeplab v3+ [18] combines the advantages of ASPP and the Encoder-Decoder structure that captures clear target boundaries by gradually restoring spatial information.

There are several different CNN models to perform the semantic segmentation. In 2017, a novel FCN architecture for semantic segmentation, Segnet [19], was presented. The decoder of segnet up-samples its lower resolution input feature map(s). Refinenet [2] is proposed which explicitly exploits all the information to enable prediction with high resolution, where the information is along the down-sampling process. Pyramid Scene Parsing Network [8] is developed in which exploits global environment information, aggregating different-place information through pyramid pooling model.

978-1-7281-3458-1/19/$31.00 ©2019 IEEE

B. Semantic Mapping

The SLAM problem was originally proposed to address the two-dimensional place localization and mapping optimization problem under the probabilistic framework [20] [21]. There are several approaches to perform the dense surface reconstruction. An efficient system for precise and real-time reconstruction of complicated indoor scenes is presented in [22]. To perform dense map geenration, [23] fuse RGB and depth images into the global map. In geometry mapping, the problem of multiple scales also need to be considered. A scalable, real-time approach for powerful reconstruction is proposed to perform surface reconstruction in multiple scales [24]. Regarding the reconstruction of moving objects, [25] incrementally fuses sensor observations into a consistent semantic map.

Many algorithms have been proposed for semantic mapping. Based on the Octomap framework, an approach for constructing multi-label semantic 3D octree mapping is proposed in [9]. The image classification is projected into the 3D lidar point cloud. The resulting point cloud feeds into the octree map and calculates the corresponding probability (occupancy and label) for each 3D voxel. A semantic SLAM system [26] is presented that uses object-level entities to construct semantic maps and integrate them into the RGB-D SLAM framework. For performing the large-scale mapping of dynamic urban environments, a stereo-based dense mapping algorithm is proposed [27]. However, the aforementioned approaches only address the problem on the level of single robot semantic mapping, and multi-robot semantic mapping is still an open problem.

III. Geometry Mapping

Building a 3D geometry reconstruction from stereo image pairs contains three steps: depth estimation, sensor pose estimation, and 3D reconstruction. Based on the RGB images and depth information of each image, point cloud can be obtained where the three-dimensional structure of the environment is shown. The point cloud, however, has several obvious drawbacks: (1) The scale of point cloud maps is usually very large. (2) Point cloud maps cannot handle moving objects. Therefore, a flexible and compressed Octomap [28] is employed, which can also be updated dynamically.

Based on the structure of octree, the probability that expresses whether a node is occupied is adopted. Generally speaking, the relevant probability is updated according to the occupancy condition of one node, that is to say, when it is observed that the node is occupied, the score of this node is increased. Conversely, as a node is observed as blank, the occupancy score shall gradually decrease. In this way, obstacle information in the map can be dynamically modeled. The probability of occupancy for each node is described by the probability logarithm (Log-odds), where $l(n) = log\frac{P(n)}{1-P(n)}$. $P(n)$ denotes the initial occupancy probability and is set as 0.5.

Given a node n and the observed data $z_{1:t}$, the probability $P(n|z_{1:t})$ of a node n to be occupied given the sensor measurements $z_{1:t}$ is estimated according to:

$$P(n|z_{1:t}) = [1 + \frac{1 - P(n|z_t)}{P(n|z_t)} \frac{1 - P(n|z_{1:t-1})}{P(n|z_{1:t-1})} \frac{P(n)}{1 - P(n)}]^{-1} \quad (1)$$

The occupancy probability logarithm of this node from the beginning to the time t is $l(n|z_{1:t})$:

$$l(n|z_{1:t}) = l(n_i|z_{1:t-1}) + l(n|z_t) \quad (2)$$

The updating of (1) depends on the current observation z_t, the previous estimation $p(n|z_{1:t-1})$ and the prior probability $p(n)$. The term $p(n|z_t)$ is the occupancy probability based on the current observation.

IV. Collaborative Semantic 3D Mapping

A. The Framework of Hierarchical Semantic 3D Mapping

The hierarchical semantic map fusion framework performs the global semantic 3D reconstruction from single-robot maps construction to multi-robot maps integration. For single-robot level, RGB information and depth information of the surroundings, as well as the coordinate transformation, can be obtained in real-time by the stereo camera. According to these information, semantic information is obtained by the CNN model deployed in the robot platform and an octomap [9] can be built. By applying the Bayes' rule, the label probabilities are added to the grid occupancy probabilities, obtaining the single-robot semantic maps. For multi-robot level, map fusion is performed to generate a global enhanced map. In particular, it is to integrate the grid data of different robots into a new grid so that the global octomap can be formed. The framework of hierarchical semantic 3D mapping is shown in Fig. 1.

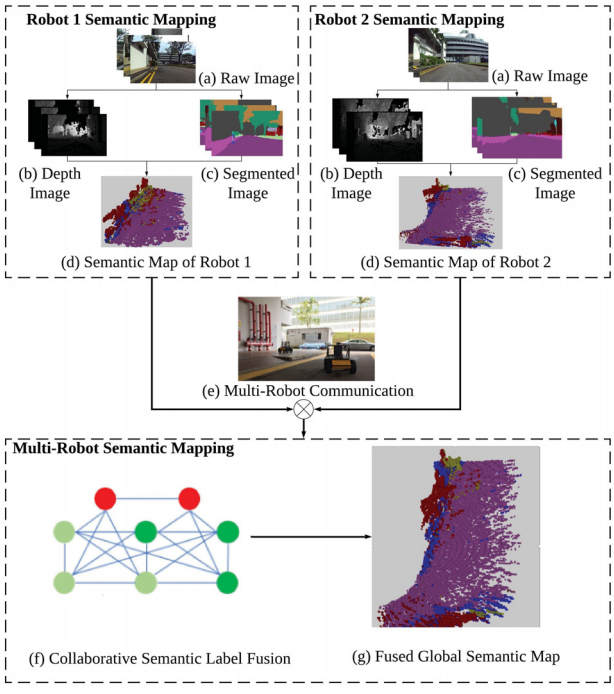

Fig. 1: The framework of hierarchical semantic 3D mapping

B. Semantic Segmentation of Single Robot

Deployment of semantic segmentation model on mobile robots contains three steps. At the first, a ROS node is built to subscribe the real-time information of the mobile robot. Secondly, the images are subscribed by the ROS node and processed by Deeplab model [19], obtaining the class and the probabilities of different labels of each pixel. Finally, the processed semantic images are published and label probabilities are used to perform the semantic mapping. The whole process is shown in Fig. 2.

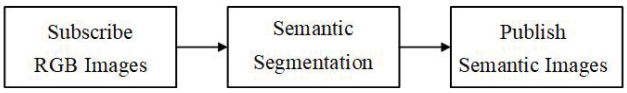

Fig. 2: Flowchart of deployment of semantic segmentation model on robot system

C. Label fusion

The Octomap generates a 3D map (cubic volume unit) for a set of point clouds. The sensor information integration is done by using the occupancy grid mapping. The probability of the voxel to be occupied is $P(n|z_{1:t})$.

The interpretation of a multi-label octree depends on the premise that each 3D point can be classified into different classes. In this work, the probability of updating each label of one node is performed as follows:

(1) z denotes the observation (2) x is the current label of the observation z. (3) $P(x|z_t)$ denotes the probability of the 3D point to belong to the class x (4) x_j denotes the label class of one node (where j=1:11). (5) $P(x)$ denotes the initial labeling probability and is set as 0.5.

In order to calculate the posterior distribution $P(x|z_{1:t})$ from the corresponding posterior on time step earlier $P(x|z_{1:t-1})$, we follow the Bayes' update rule [9]:

$$P(x|z_{1:t}) = \frac{P(x|z_t)P(z_t)P(x|z_{1:t-1})}{P(x)P(z_t|z_{1:t-1})} \tag{3}$$

The opposite event $\neg x$ of posterior distribution can be obtained by:

$$P(\neg x|z_{1:t-1}) = \frac{(1-P(x|z_t))P(z_t)(1-P(x|z_{1:t-1}))}{(1-P(x))P(z_t|z_{1:t})} \tag{4}$$

Calculate the proportion of (3) and (4):

$$\frac{P(x|z_{1:t})}{P(\neg x|z_{1:t})} = \frac{P(x|z_t)P(x|z_{1:t-1})(1-P(x))}{(1-P(x|z_t))(1-P(x|z_{1:t-1}))P(x)} \tag{5}$$

We utilize log-odds to represent the label probabilities:

$$l_t(x) = l(x|z_t) + l(x|z_{1:t-1}) \tag{6}$$

From equation (6), it is clear that the label of one voxel depends on enough observations of this voxel from time 1 to the current time. The probability $P(n, x_{max})$ of the most probable class x_{max} for each node n is computed as follows:

$$P(n, x_{max}) = \arg\max_x [P(n, x_1), P(n, x_2), \ldots, P(n, x_{11})] \tag{7}$$

D. Collaborative Semantic Label Fusion

Based on the single-robot semantic maps, collaborative map merging can be performed. Denote M_r as the semantic map generated by robot r, where M_{r_n} represents the semantic map generated by its neighboring robot r_n. For robot r, the semantic map fusion problem is to estimate the occupancy probability and label probability in each grid, and generate the global map M conditioned on available partial maps M_r and M_{r_n}. Here, we assume the relative transformaion matrix T between M_r and M_{r_n} is known.

Here, the maps generated by neighboring robot r_n will be transformed in the coordinate frame of robot r given T. Then the collaborative semantic mapping problem is a process of combining the semantic information of common objects from partial maps to form a global enhanced map. The map merging is formulated bellow.

$$P(M|M_r, M_{r_n}) \tag{8}$$

Map merging should retain all of the useful information in the input partial maps while reducing the uncertainty of the final fused map. Since the same object can be observed in different perspectives by different robots, the voxels representing the same object can have different occupancy probabilities and label probabilities in separate maps. Therefore, it is significant to consider the dissimilarities when integrating them into a global map. In the process of semantic map fusion, the formula of occupancy probabilities updating of a node is shown as:

$$P(n|z_M) = [1 + \frac{1-P(n|z_{M_r})}{P(n|z_{M_r})}\frac{1-P(n|z_{M_{r_n}})}{P(n|z_{M_{r_n}})}\frac{P(n)}{1-P(n)}]^{-1} \tag{9}$$

The occupancy probability logarithm of this node is:

$$l(n|z_M) = l(n_i|z_{M_r}) + l(n|z_{M_{r_n}}) \tag{10}$$

Regarding label probabilities updating, equation 5 can be rewritten as:

$$\frac{P(x|z_M)}{P(\neg x|z_M)} = \frac{P(x|z_{M_r})P(x|z_{M_{r_n}})(1-P(x))}{(1-P(x|z_{M_r}))(1-P(x|z_{M_{r_n}}))P(x)} \tag{11}$$

The label probabilities can be represented as:

$$l_M(x) = l(x|z_{M_r}) + l(x|z_{M_{r_n}}) \tag{12}$$

When the system perform the collaborative mapping, label fusion need compute all 11 label probabilities of each voxel in different single-robot maps. The probability $P(n, x_{max})$ of global map for each node is computed as equation (7).

The label fusion rules are employed to execute the sequential map merging. In this work, based on the posterior probabilities in different mobile robots, the posterior probabilities of occupancy and label can be computed. The fusion process effectively extracts the valuable information of each grid, ensuring the accuracy of the global map.

V. Experimental Results

We evaluate our system on the KITTI dataset [29] and NTU dataset. The semantic segmentation model is trained by using the cityscapes dataset [30]. NTU dataset is obtained by our mobile robot HUSKY Clearpath which is shown in Fig. 3.

The perception sensor is the ZED stereo camera (resolution: 672376, FOV: 8756). All algorithms are running on the Ubuntu 16.04 and ROS kinectic.

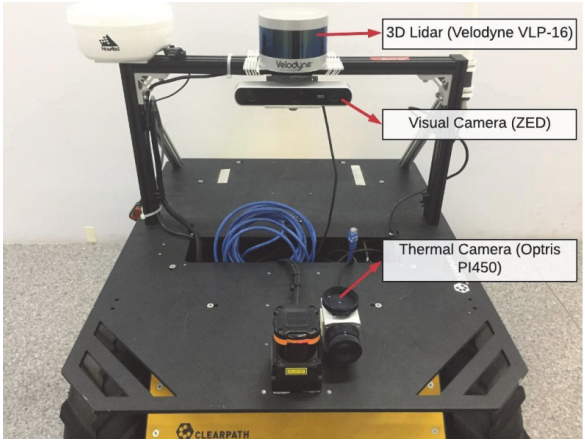

Fig. 3: HUSKY Clearpath with ZED stereo camera

A. Semantic 3D Reconstruction

We first test our system on the NTU dataset. Fig. 4 displays the input of the system, including RGB image, depth image, and corresponding semantic image. After processing the input images by our system, semantic 3D reconstruction results of the NTU dataset can be obtained, which is shown in Fig. 6.

(a) RGB image (b) Depth image (c) Semantic image

Fig. 4: An example of input of NTU datasts

Fig. 5 shows the geometric mapping results which applies Octomap [31] by using the point cloud published by stereo camera. Comparing the Octomap results of Fig. 5 and Fig. 6, it is shown that our approach can reconstruct the 3D map and recognize the classes of objects on the road with high accuracy successively.

Due to the low precision of our stereo camera, the results of NTU datasets cannot demonstrate the correctness of our system. We apply our system on the KITTI dataset which has more precise depth information of the environment. Fig. 7 shows the smooth and more precise Octomap on KITTI dataset with high precision of stereo camera. This shows that our system can perform semantic reconstruction with high accuracy.

B. Semantic Map Fusion

The algorithm of the high level collaborative map merging is tested on the NTU dataset (Left side of car park B) which is shown in Fig. 8. In this environment, we divide the dataset into two sets to simulate two robots. Firstly, the semantic maps are generated separately. Then, the two semantic maps have been fused into a global semantic map. In the right part of

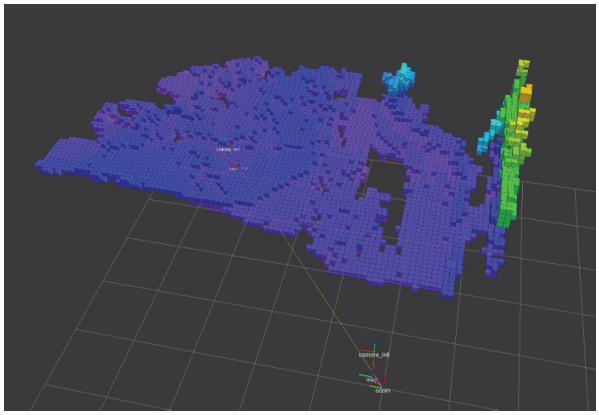

Fig. 5: Octomap of NTU datasets without labeling

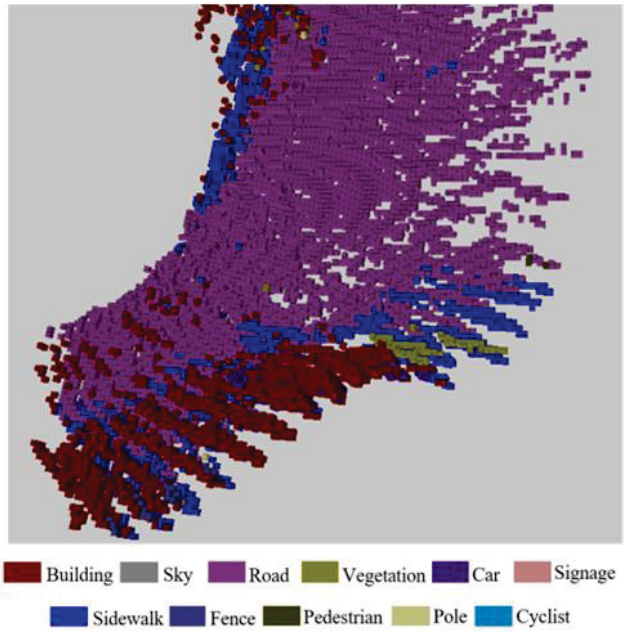

Fig. 6: Octomap of NTU datasets with labeling

fig. 9, it is the low-level single-robot mapping part of our system, utilizing the environment views to build semantic map for each robot. The integrated global map which is presented in Fig. 9 is the high-level multi-robot map merging part. From the whole process of collaborative mapping, it is shown that our system performs well in both single-robot level and multi-robot level. For single-robot level, the semantic octomap can be completely built by utilizing the environment view of mobile robots. For multi-robot level, collaborative map merging achieves the fusion of maps of two robots, generating an improved global map.

C. Quantitative analysis

Fig. 10 shows the continuous probability updating process of a single voxel. Update 0 corresponds to the initial state that the probabilities of all labels are all initialized with the same value 9.09%.

Table I shows the updating process of one voxel in which

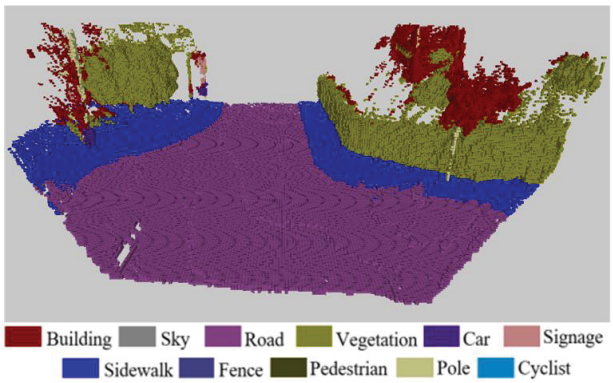

Fig. 7: Octomap of KITTI datasets with labeling

Fig. 8: Overview of car park B

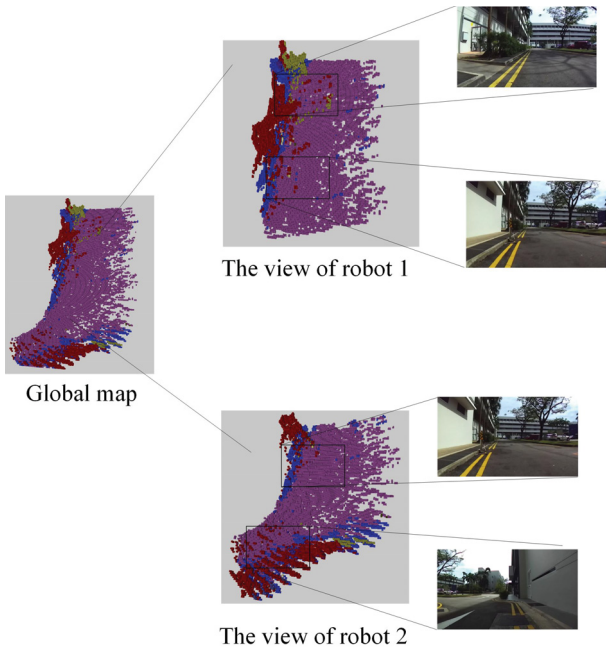

Fig. 9: Collaborative mapping result

points are entered in order, specifying the label and its corresponding probability. The first point used for updating the voxel belongs to the label Road with a probability of 0.978569. From Fig. 10 we can observe the increase in the probability of the label ”Road” (gray) and the decrease in that of the remaining labels. As the second point is entered, an increasing trend in the confidence that this voxel belongs to the ”Road” can be seen, which is the same as the behavior of the first point.

An efficiency test is performed in our real robot system, which can process the information of the environment in real-time. In the test, when an NVIDIA 1060 is used to handle the CNN model, the frame rate of the mobile robot is 2.2 fps, which is sufficient for mobile robots in practical scenario.

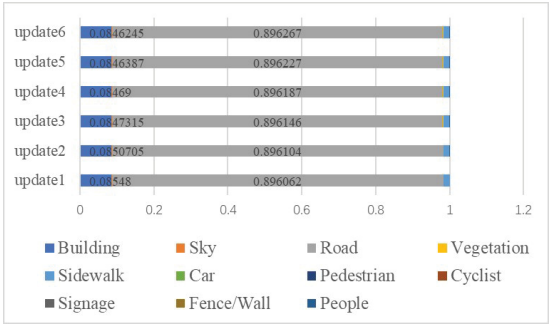

Fig. 10: Update process of a single voxel (road)

TABLE I: Probability of the grid label before each label fusion (6 update steps)

Update	Label	Probability
1	Road	0.978569
2	Road	0.978861
3	Road	0.9791
4	Road	0.98038
5	Road	0.981926
6	Road	0.983317

VI. Conclusion

This paper established a semantic 3D mapping framework for multi-robot systems. On the basis of geometric occupancy mapping, we firstly applied the semantic segmentation model to mobile robot system. Then, the label information is fused with the Octomap to achieve semantic 3D mapping on single-robot level. To achieve multi-robot semantic mapping, this paper provided a theoretical basis and algorithm for the global reconstruction of 3D semantic maps between mobile robots. In the future, semantic map fusion with real multi-robot setting will be investigated. Multi-robot localization and place recognition can be performed based on the semantic maps, which shall lead to a significant improvement in performances of mobile robots.

References

[1] P. Sun, J. Chen, and H. Y. K. Lau, “Programming human-like point-to-point approaching movement by demonstrations with large-scale direct monocular slam,” in *2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec 2016, pp. 1498–1503.

[2] G. Lin, A. Milan, C. Shen, and I. Reid, "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1925–1934.
[3] Y. Wang, M. Shan, Y. Yue, and D. Wang, "Vision-based flexible leader-follower formation tracking of multiple nonholonomic mobile robots in unknown obstacle environments," *IEEE Transactions on Control Systems Technology*, pp. 1–9, 2019.
[4] Y. Yue, D. Wang, P. G. C. N. Senarathne, and D. Moratuwage, "A hybrid probabilistic and point set registration approach for fusion of 3d occupancy grid maps," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct 2016, pp. 001 975–001 980.
[5] Y. Yue, P. G. C. N. Senarathne, C. Yang, J. Zhang, M. Wen, and D. Wang, "Probabilistic fusion framework for collaborative robots 3d mapping," in *2018 21st International Conference on Information Fusion (FUSION)*, July 2018, pp. 488–491.
[6] C. Yang, Y. Yue, J. Zhang, M. Wen, and D. Wang, "Probabilistic reasoning for unique role recognition based on the fusion of semantic-interaction and spatio-temporal features," *IEEE Transactions on Multimedia*, 2018.
[7] C. Yang, D. Wang, Y. Zeng, Y. Yue, and P. Siritanawan, "Knowledge-based multimodal information fusion for role recognition and situation assessment by using mobile robot," *Information Fusion*, vol. 50, pp. 126 – 138, 2019.
[8] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.
[9] J. S. Berrio, J. Ward, S. Worrall, W. Zhou, and E. Nebot, "Fusing lidar and semantic image information in octree maps," in *ACRA Australasian Conference on Robotics and Automation 2017*, 2017.
[10] Y. Yue, P. G. C. N. Senarathne, C. Yang, J. Zhang, M. Wen, and D. Wang, "Hierarchical probabilistic fusion framework for matching and merging of 3-d occupancy maps," *IEEE Sensors Journal*, vol. 18, no. 21, pp. 8933–8949, Nov 2018.
[11] Y. Yue, C. Yang, Y.Wang, P. G. C. N. Senarathne, J. Zhang, M. Wen, and D. Wang, "A multi-level fusion system for multi-robot 3d mapping using heterogeneous sensors," *IEEE Systems Journal*, 2019.
[12] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov 1998.
[13] M. T. Yu and M. M. Sein, "Automatic image captioning system using integration of n-cut and color-based segmentation method," in *SICE Annual Conference 2011*, Sep. 2011, pp. 28–31.
[14] Y. Li, J. Zhang, P. Gao, L. Jiang, and M. Chen, "Grab cut image segmentation based on image region," in *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, June 2018, pp. 311–315.
[15] A. G. Schwing and R. Urtasun, "Fully connected deep structured networks," *arXiv preprint arXiv:1503.02351*, 2015.
[16] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.
[17] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.
[18] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 801–818.
[19] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
[20] A. Nüchter, K. Lingemann, J. Hertzberg, and H. Surmann, "6d slam3d mapping outdoor environments," *Journal of Field Robotics*, vol. 24, no. 8-9, pp. 699–722, 2007.
[21] M. Milford, E. Vig, W. Scheirer, and D. Cox, "Vision-based simultaneous localization and mapping in changing outdoor environments," *Journal of Field Robotics*, vol. 31, no. 5, pp. 780–802, 2014.
[22] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. W. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking." in *ISMAR*, vol. 11, no. 2011, 2011, pp. 127–136.
[23] S. Lee, H. Lim, H. Kim, and S. C. Ahn, "Rgb-d fusion: Real-time robust tracking and dense mapping with rgb-d data fusion," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sep. 2014, pp. 2749–2754.
[24] J. Zienkiewicz, A. Tsiotsios, A. Davison, and S. Leutenegger, "Monocular, real-time surface reconstruction using dynamic level of detail," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 37–46.
[25] D. Kochanov, A. Ošep, J. Stückler, and B. Leibe, "Scene flow propagation for semantic mapping and object discovery in dynamic street scenes," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 1785–1792.
[26] L. Zhang, L. Wei, P. Shen, W. Wei, G. Zhu, and J. Song, "Semantic slam based on object detection and improved octomap," *IEEE Access*, vol. 6, pp. 75 545–75 559, 2018.
[27] I. A. Bârsan, P. Liu, M. Pollefeys, and A. Geiger, "Robust dense mapping for large-scale dynamic environments," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 7510–7517.
[28] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: An Efficient Probabilistic 3D Mapping Framework Based on Octrees," *Autonomous Robots*, 2013.
[29] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 3354–3361.
[30] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 3213–3223.
[31] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "Octomap: An efficient probabilistic 3d mapping framework based on octrees," *Autonomous robots*, vol. 34, no. 3, pp. 189–206, 2013.