

SeGuE: Semantic Guided Exploration for Mobile Robots

Cody Simons, Aritra Samanta, Amit K. Roy-Chowdhury, Konstantinos Karydis

Abstract—The rise of embodied AI applications has enabled robots to perform complex tasks that require sophisticated understanding of their environment. To allow successful robot operation in such settings, maps must be constructed to include both semantic and geometric information. In this paper, we address the novel problem of semantic exploration, whereby a mobile robot must autonomously explore an environment to fully map its structure and features’ semantic appearance. We develop a method based on next-best-view exploration, where potential poses are scored based on the semantic features visible from that pose. We explore two alternative methods for sampling potential views and demonstrate the effectiveness of our framework in both simulation and physical experiments. The automatic creation of high-quality semantic maps can enable robots to better understand and interact with their environments and facilitate the deployment of future embodied AI applications.

I. INTRODUCTION

Advances in machine learning and edge computing have propelled robots toward acquiring reasoning and decision-making capabilities that are higher than ever before. These capabilities are critical for successfully deploying them in workplace environments such as automated kitchens [1] or unmanned farms [2], where a robot must solve a joint navigation and semantic scene understanding problem.

To solve this problem, classical metric-based maps have been fused with semantic-based maps to offer appropriate information-rich maps. For example, in visual object navigation [3], an agent must find an object that matches a text or picture. Similarly, in language-guided navigation [4], an agent must navigate based on natural language instructions that reference the appearance of the environment. These methods are often zero-shot, assuming no prior knowledge of the environment. Although this setting is challenging, it is inherently transient. In real-world cases where continuous operation is required, constructing a comprehensive map during or before operation would be highly beneficial. Simultaneously, several existing methods (e.g., [1], [3]–[5]) use maps that have been augmented with semantic classes. Although semantic class information is valuable, its limited reusability may hinder transfer to physical robot deployment across environments. To address this challenge, we propose an exploration method that maps the semantic *features* of the environment, thus enabling reusability across tasks.

The authors are with the Dept. of Electrical and Computer Eng. at the Univ. of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA. Email: {csimo005, asama004, amitrc, karydis}@ucr.edu. We gratefully acknowledge the support of NSF # IIS-1901379, # CNS-2312395 and ONR # N00014-18-1-2252, # N00014-19-1-2264. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

Modern AI-enhanced semantic maps can jointly encode the geometry and appearance features of a scene. They can hold features extracted by machine learning models, which can be reused across many different tasks. Semantic feature maps can be created online [3], typically to aid in visual object search, or offline [4], which can be used for language-guided navigation or other downstream tasks. However, offline mapping typically requires manual robot operation or hand-engineered routes to ensure the comprehensive coverage of appearance features. Although autonomous exploration methods for mapping environmental structures are well-established in robotics, there is a notable gap in *automatically guiding exploration to create high-quality semantic maps*.

Autonomous environment exploration and mapping have received significant attention over the years. Mapping the structure of a static environment in 2D has largely been solved by existing SLAM approaches [6]. Exploration has also been extensively studied, with conventional approaches divided broadly into frontier-based and next-best-view-based methods. In frontier-based methods [7], the robot moves to locations that are on the border between the mapped and unmapped portions of the environment, whereas next-best-view methods [8]–[10] sample and score potential views based on information-theoretic metrics. Given sufficient time, both methods can fully capture the geometry of the environment. However, they do not consider the appearance of the environment. Furthermore, the increased range and field of view of modern depth sensors can result in incomplete and low-quality semantic feature maps [11].

In this work, we introduce a novel autonomous exploration method based on the next-best-view approach, which maps the structure of the environment and simultaneously captures its semantic appearance. Our method scores potential views by computing the average entropy of the visible semantic features, excluding the occluded and converged features. We investigate two alternative view-sampling techniques: uniform sampling across reachable poses and an iterative importance-sampling method inspired by particle filter optimization. Off-the-shelf components are used for metric-based mapping and navigation. Our approach is validated through both simulation and physical hardware experiments using a wheeled mobile robot. In sum, our contributions include:

- 1) We propose a method for scoring individual semantic map features and show how these scores can be combined to score a particular view.
- 2) We explore two alternative viewpoint sampling methods.
- 3) We demonstrate the feasibility of our framework in simulated and physical experiments.

II. RELATED WORK

A. Semantic Mapping

Semantic mapping has been explored in both the machine learning and robotics communities. Machine learning research has focused on dense semantic features [4]. These maps allow for fine-grained search within the environment [4] and can be used to predict the layout of unexplored portions of the environment [12]. Within robotics, there has been a focus on augmenting maps with semantic classes [13], [14]. Semantic categories have been used to increase loop closures [15] and create more efficient routes [16]. As the resources available on robots (sensors, computing units, etc.) continue to grow and AI models become more amenable to edge computing, the demand for maps of dense semantic features is expected to rise. Our method can help in the automatic construction of such maps.

B. Frontier Exploration

Frontier-based exploration has been extensively studied in the literature. Initially proposed in [7], frontier-based methods seek to fully explore an environment by constantly moving toward the border between the known and unknown spaces. More recent works have sought to increase the computational efficiency of computing frontiers [17] and create frontier ranking mechanisms [18]. These methods are solely concerned with mapping the geometry of an environment, which can be directly observed using LiDAR sensors that are commonly integrated into contemporary robots. In contrast, mapping semantic features can be significantly complicated by the variability in object appearance at different distances, the salience of various objects to feature extractors, and the limited field of view of most cameras.

C. Next-Best-View Determination

The next-best-view pipeline [8]–[10] consists of a pose sampler, a metric to rank poses, either in terms of information gain [8]–[10] or a task-specific metric [1], [5], and a mapper which updates an internal map while navigating to the determined next-best-view pose. Several contemporary implementations use sampling-based planners to select poses [8], [19]. However, this can lead to being trapped in local maxima. One way to address this is by integrating an additional global sampler [10]. Information-theoretic metrics [9], [20] encode the structure and confidence in the current map to rank the value of a particular view. Semantic categories have also been used [1] to focus on task-relevant structures. To the best of our knowledge, no method has incorporated a semantic feature map into a next-best-view algorithm to date.

III. PROBLEM DEFINITION

In this work, we address a novel semantic exploration problem. We aim to simultaneously explore and map an environment in real time autonomously while ensuring that the map captures the semantic information necessary for downstream embodied AI applications. Specifically, we ask how a mobile robot equipped with an RGB camera and depth sensor can explore an environment and construct a 2D map

$\mathcal{M} \in \mathbb{R}^{H \times W \times N+1}$, where H and W represent the height and width of the map, respectively, and N represents the dimension of the semantic feature vectors. The additional scalar value in the third dimension represents the occupancy information in the form of an occupancy grid. Although semantic map generation is a thoroughly studied field [4], in this work, we specifically address the exploration problem to ensure sufficient coverage of the environment and capture high-quality semantic features. To reduce implementation complexity, we assume that the robot operates in a physically bounded environment and that the environment is relatively flat and traversable. These assumptions are not restrictive because they relate to the most common indoor navigation paradigms (as tested herein) and can directly carry over to outdoor navigation in structured environments.

IV. PROPOSED APPROACH

We propose the Semantic Guided Exploration (SeGuE) method for mobile robots. Our method is based on next-best-view algorithms. During exploration, a semantic map, S , is constructed using data from two different sensing modalities, specifically an RGB camera and a 3D LiDAR. SeGuE samples collision-free and reachable poses from the occupancy grid generated during navigation based on the current semantic understanding of the environment. The average prediction entropy of each semantic feature visible from a pose is then computed to estimate the potential information gain from each pose. The average is used to favor higher entropy features rather than large collections of low-entropy features. To ensure that cells with higher aleatoric uncertainty are not overly exploited, the entropy of each semantic feature is tracked over time and features whose entropy has converged are excluded from score calculations. This process is repeated until a pose with a score above the threshold cannot be found. SeGuE is summarized in Algorithm 1. Next, we elaborate on each step of the approach.

A. Pose Scoring

The poses are scored according to the amount of information that can be gained from each pose. We use entropy as a proxy for this, encouraging the robot to explore the areas of the map that it is most uncertain about. During exploration, the semantic score map is continually updated by extracting dense semantic features from the current image observation. These features are then associated with different points in the point cloud *ptCloud*. Because the 3D points now represent semantic features, they are projected onto the 2D semantic map to associate each feature with a cell. The features associated with the same cell are averaged. (Details of this procedure can be found in [4].) Semantic features are extracted from the image *img* using a machine learning model $h = f \circ g$, which can be decomposed into a feature extractor f and classifier g . Consider g can classify M classes. The features stored in our semantic map are extracted from f . Throughout our experiments, we use DinoV2 [21], trained in a self-supervised manner to encourage scale invariance, to extract features and a linear classifier fine-tuned on the ADE20K

Algorithm 1: SeGuE Algorithm

Require: Termination Threshold τ

- 1: Initialize an empty semantic map S and empty occupancy map O
 - 2: Get image img , point cloud $ptCloud$, and current pose p_{curr} from robot
 - 3: $S.update(img, ptCloud, p_{curr})$
 - 4: $O.update(ptCloud, p_{curr})$
 - 5: **while** True **do**
 - 6: Sample a set of pose $P = \{p_i\}_{i=0}^N$ using PoseSampler(S, O)
 - 7: $p^* = \max_{p \in P} \text{PoseScore}(p, S, O)$
 - 8: **if** $\text{PoseScore}(p^*, S, O) < \tau$ **then**
 - 9: End exploration
 - 10: **end if**
 - 11: Begin navigating to pose p^*
 - 12: **while** p^* not reached **do**
 - 13: Get image img , point cloud $ptCloud$, and current pose p_{curr} from robot
 - 14: $S.update(img, ptCloud, p_{curr})$
 - 15: $O.update(ptCloud, p_{curr})$
 - 16: **end while**
 - 17: **end while**
-

dataset [22]. The remainder of this section describes the scoring mechanism in detail.

1) *Feature Scoring*: Each cell in the semantic map contains a semantic feature vector that encodes the appearance of the location. The quality of each semantic feature associated with a particular cell changes over time as the corresponding viewpoint varies. The prediction entropy of each cell is computed to assess the quality of the semantic feature based on the assumption that lower entropy predictions correspond to higher quality features and semantic maps. Using the classification head g , which classifies M classes, the score for a semantic feature x is

$$s(x) = \frac{H(g(x))}{H(\mathcal{U}(M))}, \quad (1)$$

where H is the Shannon entropy, and \mathcal{U} is the uniform distribution. The score is normalized by the maximum entropy to be in the range $[0, 1]$, since the maximum entropy can grow arbitrarily large as M increases. Any cell with no semantic feature is assigned a score of 1, indicating maximal uncertainty about cells with no observations. Once these scores are computed for every cell in the semantic map, potential views can be ranked by averaging the scores of the individual cells that are visible from that viewpoint.

2) *Visibility Mask*: We require that only locations visible from a given pose should be considered when providing a score for that pose. To enforce this, a ray-tracing algorithm similar to that in [12] is used to construct a view mask of cells visible from a particular pose. An example of the mask is shown in Fig. 1. In the figure, only the cells highlighted in gray are included in the aforementioned averaging operation;

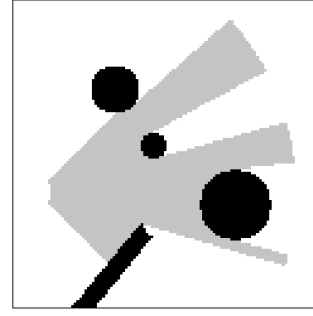


Fig. 1: An example of the view mask generated by our raytracing algorithm. Known obstacles are shown in black and the view mask is shown in gray.

all other cells are ignored. This ensures that the overall score of a pose is not affected by individual subscores that may be gathered from cells that provide no new information.

3) *Feature Convergence Tracking*: Even after collecting several observations, some cells may still have high entropy owing to the aleatoric uncertainty present in some classes. Thus, in cases where additional observations would not improve a semantic feature score, any new observation should not contribute to the pose score computation. This can help prevent the exploitation of inherently uncertain cells. To address this, the feature score of each cell is tracked over time and monitored for convergence.

A ratio convergence test is employed to determine when a feature score converged. During exploration, when the feature score value of a non-converged cell is updated, the ratio $\frac{s_0}{s_1}$ is computed, where s_0 is the previous score and s_1 is the new score. If the ratio is less than a user-defined threshold, the score for that cell is considered to have converged. (A threshold of 1.1 was found to work well throughout our experiments.) Converged features can still be updated if they are incidentally observed during exploration; however, once the feature score converges, the respective cell will no longer be used in pose scoring.

B. Pose Sampling

Diverse poses across the map must be considered to find those that maximize metric (1), and hence ensure that entropy is minimized and the semantic information is mapped fully. To this end, two pose-sampling strategies, Uniform Sampling (US) and Importance Sampling (IS), are proposed. Both strategies employ the same scoring mechanism. Regardless of the sampling method, exploration is terminated if no pose is found within a certain percentage of the maximum score. This is also a user-defined hyperparameter. The threshold was set to 5% throughout our experiments; however, future work can systematically investigate its effect. The number of samples and iterations for each method were selected as per Section V-B and were kept fixed in all the experiments.

1) *Uniform Sampling*: In the first method, the poses are sampled uniformly across the entire map, and unreachable poses are discarded. Sampling continues until the desired number of reachable poses is obtained. These poses are then scored according to (1). The pose with the highest score is selected as the next pose to navigate to.

2) *Importance Sampling*: Although the US method can acquire poses with high entropy, a large number of samples must be considered for assessment, which may be inefficient in practice. Inspired by [23], we also implement a sampling approach based on particle filter optimization. This method samples an initial set of poses uniformly across reachable poses, which are used to fit a mixture of Gaussian models, where the likelihood of each pose is directly proportional to its score (1). This distribution is then used to sample a new set of poses, subject to the same reachability constraint, and the process is repeated for a fixed number of iterations. The mean of the mixture component with the highest score is taken as the next pose to navigate to.

V. EXPERIMENTS

We evaluated SeGuE in both simulated and real-world environments using a wheeled mobile robot. The entire software suite is written in the Robot Operating System (ROS) framework. Odometry and mapping were conducted using 3D LiDAR information, as described in [24]. Point cloud data were projected onto the ground plane to obtain a 2D occupancy grid for motion planning. Waypoint navigation was conducted using the built-in ROS package *move_base*. Scene features were extracted using an RGB camera.

To evaluate the quality of the generated semantic maps, we assessed the map coverage and average entropy. Map coverage provides an insight into the proportion of the appearance features that have been captured during exploration. It is computed as the percentage of occupancy map cells that are either free or occupied and have a corresponding semantic feature. Higher map coverage values are better. The average entropy indicates the quality of the observations on the map and is computed over the entire semantic feature map. Any cell observed in the occupancy map with no corresponding semantic feature was assumed to have maximum entropy. Lower values of average entropy are better.

A. Simulation Results

We simulated the Clearpath Jackal in Gazebo [25] and the relevant sensors using the calibration data available in [26]. All simulations were performed on a desktop computer equipped with an Intel i7 processor, NVIDIA RTX 3090, and 32 GB of RAM. We present the results in two different environments. The Small House environment replicates a residential home, and the Bookstore environment is a retail space. For each environment, the initial conditions were fixed across SeGuE and each baseline method.

We compared SeGuE using both the US and IS variants against a frontier-based exploration method [27] and a next-best-view method (RRG-NBV) [19]. (Sample sizes chosen as per Section V-B.) We also compared SeGuE to a *No Score* baseline, where we replaced our map scoring metric with a simple metric in which every unseen cell has a score of one and every seen cell has a score of zero. View scoring and importance sampling were kept fixed.

The map quality metrics for each method are listed in Table I. It can be observed that, across the board, the frontier-

TABLE I: Simulation Results.

	Small House			Bookstore		
	Coverage \uparrow	Average Entropy \downarrow	Time	Coverage \uparrow	Average Entropy \downarrow	Time
Frontiers Based [27]	0.303	3.619	0 : 05 : 13	0.477	2.862	0 : 06 : 20
RRG-NBV [19]	0.730	1.643	0 : 05 : 26	0.255	3.917	0 : 01 : 03
No Score, US	0.905	0.806	0 : 23 : 12	0.925	0.801	1 : 08 : 48
No Score, IS	0.897	0.842	0 : 15 : 42	0.929	0.787	1 : 05 : 23
SeGuE, US	0.938	0.673	0 : 47 : 32	0.940	0.723	1 : 22 : 51
SeGuE, IS	0.932	0.702	0 : 40 : 11	0.939	0.730	1 : 19 : 31

based and RRG-NBV approaches do not perform well. This can be explained by the fact that no appearance features are considered, but instead, the scene geometry is mapped. Both SeGuE and the *No Score* baselines are significantly better, with lower average entropy and higher coverage of the semantic features, demonstrating the necessity of exploration methods that are aware of scene semantics. Our semantic feature scoring method lowers the average entropy when combined with either sampling method.

We visualize the occupancy grid and semantic maps produced in each environment using the frontier-based method and SeGuE in Fig. 2 (for Small House) and Fig. 3 (for Bookstore). Because semantic features exist in a high-dimensional abstract feature space, we pass the features through a prediction head and visualize the class predictions, where each color represents a different class. Both methods produce a complete occupancy grid; however, in frontier-based cases, the semantic maps are incomplete. In these cases, frontier-based exploration ends because it is only aware of the structure of the environment; however, our semantic-aware method fully maps the semantic maps.

B. Sample Number Study

We evaluate the sensitivity of both sampling methods to the number of samples chosen in the Small House environment. In the case of the uniform sampling method, this is a direct sweep over the number of samples selected. The results of this ablation study, with and without our semantic scoring method, are shown in Table II. We note that, overall, our semantic scoring method increases coverage and decreases the average entropy. While no direct relation between the number of samples and the performance appears to exist, we observe that, with and without our semantic scoring method, the optimal performance seems to be between 100 and 200 samples. This is likely because too few samples do not approach an optimal score, whereas too many samples tend to over-exploit, leading to less exploration overall.

We performed the same study using the IS method, varying the number of samples chosen in every iteration and the total number of iterations. The quantitative results are presented in Table III. When comparing the *No Score* performance between the sampling methods (cf. Tables II and III), the performance does not change significantly. We observe a marked improvement when the *Semantic Score* is combined with IS, indicating that IS can better exploit the information within our metric. Increasing the number of samples resulted in marginally worse performance, likely caused by excessive exploitation and less overall exploration.

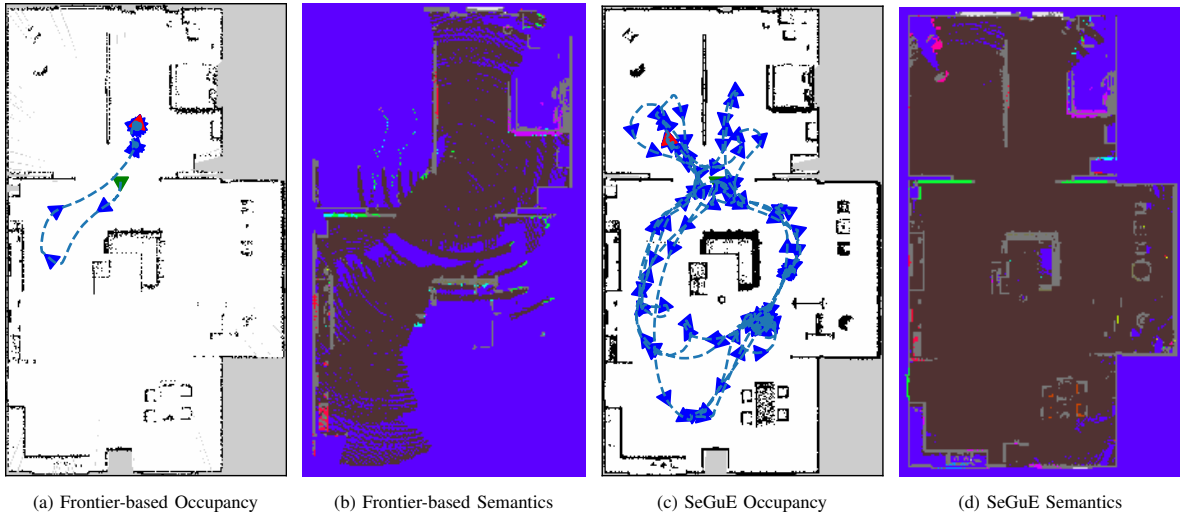


Fig. 2: Visualization of the mapping results of (a-b) the frontier-based baseline and (c-d) SeGuE in the Small House environment. We show both the occupancy grid and semantic prediction, computed from the map of semantic features. On the occupancy grid, we plot the trajectory of the robot.

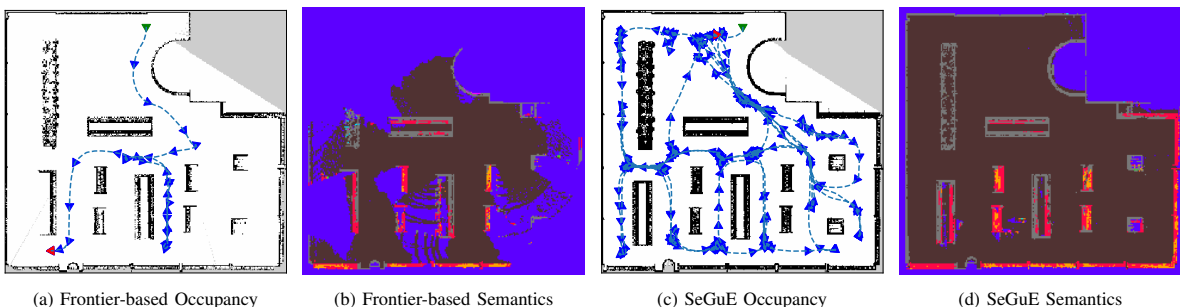


Fig. 3: Visualization of the mapping results of (a-b) the frontier-based baseline and (c-d) SeGuE in the Bookstore environment. We show both the occupancy grid and semantic prediction, computed from the map of semantic features. On the occupancy grid, we plot the trajectory of the robot.

TABLE II: Ablation Study: Sample Number in US.

Samples	Semantic Score			No Score		
	Coverage \uparrow	Average Entropy \downarrow	Time	Coverage \uparrow	Average Entropy \downarrow	Time
10	0.897	0.842	0 : 18 : 49	0.895	0.850	0 : 17 : 33
50	0.928	0.711	0 : 26 : 46	0.899	0.837	0 : 12 : 16
100	0.908	0.801	0 : 39 : 15	0.911	0.775	0 : 14 : 23
200	0.938	0.673	0 : 47 : 32	0.905	0.806	0 : 23 : 12
500	0.904	0.812	0 : 34 : 47	0.887	0.903	0 : 15 : 49
1000	0.912	0.772	0 : 36 : 26	0.895	0.851	0 : 16 : 18

TABLE III: Ablation Study: Number of Samples and Iterations in IS.

Iterations	Samples	Semantic Score			No Score		
		Coverage \uparrow	Average Entropy \downarrow	Time	Coverage \uparrow	Average Entropy \downarrow	Time
2	50	0.932	0.702	0 : 40 : 11	0.897	0.842	0 : 15 : 42
5	20	0.926	0.704	0 : 33 : 38	0.900	0.844	0 : 13 : 44
10	10	0.905	0.807	0 : 28 : 58	0.881	0.921	0 : 14 : 18
10	20	0.918	0.750	0 : 50 : 46	0.888	0.876	0 : 17 : 37
10	50	0.919	0.771	0 : 59 : 51	0.916	0.761	0 : 21 : 17
10	100	0.898	0.836	0 : 27 : 02	0.898	0.850	0 : 20 : 07

C. Real-world Experiments

In addition to the simulation studies, we evaluated SeGuE in a real-world environment (campus cafeteria). We used the Clearpath Jackal wheeled mobile robot equipped with the Velodyne VLP-16 LiDAR and Zed2i stereo camera; for a full description of the sensor calibration, see [28]. Additionally, we connected the Clearpath Jackal robot with a laptop equipped with an Intel i7 processor, an NVIDIA RTX 3060, 32 GB of RAM, and 512 GB of storage.

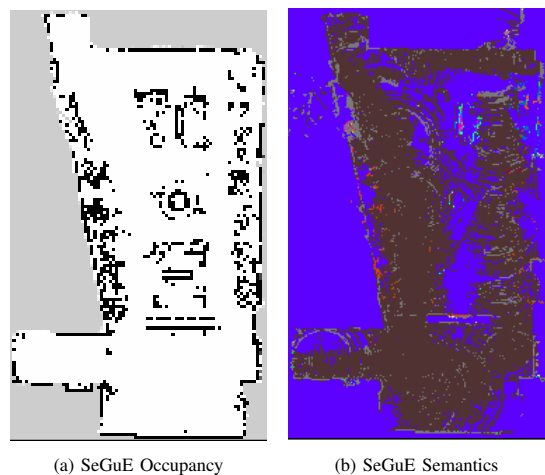


Fig. 4: Mapping results of SeGuE in real-world experiments.

We tested SeGuE with uniform sampling because it was found to perform consistently well across simulation studies. Experimental results are shown in Fig. 4. The computed coverage is 0.735 and the average entropy is 1.753. Although the overall coverage is lower than the values found in simulation studies, our method can still map most of the environment. This performance degradation can be attributed to two factors. First, the underlying motion planning library did not transfer well to a more cluttered location. Second,

battery constraints limit the total runtime. Adjusting the underlying motion planner to be more energy-aware (e.g., by building on top of [29] and [30]) could enable optimal coverage subject to motion budget and task (i.e. viewpoint) constraints, so that the environment can be explored timely.

VI. CONCLUSION

We developed SeGuE, a method to autonomously explore an environment that ensures that appearance features are sufficiently observed so that a high-quality and complete map of semantic features can be constructed. We proposed a method to score the individual semantic features contained within the map and showed how these scores may be aggregated to rank potential viewpoints. Two viewpoint sampling methods were considered, and our tests showed how they interacted with the system as a whole. Results of both the simulation and physical hardware experiments showed that our method can fuse semantic feature information well, improving overall exploration metrics, such as map coverage and average entropy, compared with the baseline methods.

Because SeGuE builds on top of a next-best-view algorithm, it remains bound to the inherent limitations of the latter, particularly the performance dependence on the sampling method and scoring. Furthermore, we focused only on global sampling methods to avoid getting trapped in local minima. In future work, we aim to integrate local sampling.

REFERENCES

- [1] N. Blodow, L. C. Goron, Z.-C. Marton, D. Pangercic, T. Rühr, M. Tenorth, and M. Beetz, "Autonomous semantic mapping for robots performing everyday manipulation tasks in kitchen environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4263–4270, 2011.
- [2] A. Dechemi, D. Chatziparaschis, J. Chen, M. Campbell, A. Shamshirgaran, C. Mucchiani, A. Roy-Chowdhury, S. Carpin, and K. Karydis, "Robotic assessment of a crop's need for watering: Automating a time-consuming task to support sustainable agriculture," *IEEE Robotics & Automation Magazine*, vol. 30, no. 4, pp. 52–67, 2023.
- [3] S. Y. Gadre, M. Wortsman, G. Ilharco, L. Schmidt, and S. Song, "Cows on pasture: Baselines and benchmarks for language-driven zero-shot object navigation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23171–23181, 2023.
- [4] C. Huang, O. Mees, A. Zeng, and W. Burgard, "Visual language maps for robot navigation," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 10608–10615, 2023.
- [5] S. Achat, Q. Serdel, J. Marzat, and J. Moras, "A case study of semantic mapping and planning for autonomous robot navigation," *SN Computer Science*, vol. 5, 2023.
- [6] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on Robotics (TRO)*, vol. 32, no. 6, pp. 1309–1332, 2016.
- [7] B. Yamauchi, "A frontier-based approach for autonomous exploration," in *IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA)*, pp. 146–151, 1997.
- [8] A. Bircher, M. Kamel, K. Alexis, H. Oleynikova, and R. Siegwart, "Receding horizon "next-best-view" planner for 3d exploration," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1462–1468, 2016.
- [9] E. Palazzolo and C. Stachniss, "Effective exploration for mavs based on the expected information gain," *Drones*, vol. 2, no. 1, 2018.
- [10] M. Selin, M. Tiger, D. Duberg, F. Heintz, and P. Jensfelt, "Efficient autonomous exploration planning of large-scale 3-d environments," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1699–1706, 2019.
- [11] M. Brossard, S. Bonnabel, and A. Barrau, "A new approach to 3d icp covariance estimation," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 744–751, 2020.
- [12] N. Yokoyama, S. Ha, D. Batra, J. Wang, and B. Bucher, "Vlfm: Vision-language frontier maps for zero-shot semantic navigation," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 42–48, 2024.
- [13] J. M. Correia Marques, A. J. Zhai, S. Wang, and K. Hauser, "On the overconfidence problem in semantic 3d mapping," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11095–11102, 2024.
- [14] D. Morilla-Cabello, J. Westheider, M. Popović, and E. Montijano, "Perceptual factors for environmental modeling in robotic active perception," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4605–4611, 2024.
- [15] X. Chen, A. Milioto, E. Palazzolo, P. Giguère, J. Behley, and C. Stachniss, "Suma+: Efficient lidar-based semantic slam," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4530–4537, 2019.
- [16] J. Liu, Y. Lv, Y. Yuan, W. Chi, G. Chen, and L. Sun, "An efficient robot exploration method based on heuristics biased sampling," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 7, pp. 7102–7112, 2023.
- [17] M. Keirdar and G. A. Kaminka, "Efficient frontier detection for robot exploration," *The International Journal of Robotics Research*, vol. 33, no. 2, pp. 215–236, 2014.
- [18] D. L. da Silva Lubanco, M. Pichler-Scheder, and T. Schlechter, "A novel frontier-based exploration algorithm for mobile robots," in *IEEE International Conference on Mechatronics and Robotics Engineering (ICMRE)*, pp. 1–5, 2020.
- [19] M. Steinbrink, P. Koch, B. Jung, and S. May, "Rapidly-exploring random graph next-best view exploration for ground vehicles," in *European Conference on Mobile Robots (ECMR)*, pp. 1–7, 2021.
- [20] R. Monica and J. Aleotti, "Contour-based next-best view planning from point cloud segmentation of unknown objects," *Autonomous Robots*, pp. 443–458, 2017.
- [21] M. Oquab, T. Darcet, T. Moutakanni, H. V. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, R. Howes, P.-Y. Huang, H. Xu, V. Sharma, S.-W. Li, W. Galuba, M. Rabbat, M. Assran, N. Ballas, G. Synnaeve, I. Misra, H. Jegou, J. Mairal, P. Labatut, A. Joulin, and P. Bojanowski, "Dinov2: Learning robust visual features without supervision," 2023.
- [22] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene parsing through ade20k dataset," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5122–5130, 2017.
- [23] B. Liu, S. Cheng, and Y. Shi, "Particle filter optimization: A brief introduction," in *Advances in Swarm Intelligence*, pp. 95–104, Springer, 2016.
- [24] H. Teng, Y. Wang, D. Chatziparaschis, and K. Karydis, "Adaptive lidar odometry and mapping for autonomous agricultural mobile robots in unmanned farms," *Computers and Electronics in Agriculture*, vol. 232, p. 110023, 2025.
- [25] N. Koenig and A. Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. 3, pp. 2149–2154, 2004.
- [26] H. Teng, D. Chatziparaschis, X. Kan, A. K. Roy-Chowdhury, and K. Karydis, "Centroid distance keypoint detector for colored point clouds," in *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1196–1205, 2023.
- [27] J. Hörner, "Map-merging for multi-robot system," Master's thesis, Charles University, Faculty of Mathematics and Physics, 2016.
- [28] H. Teng, Y. Wang, X. Song, and K. Karydis, "Multimodal dataset for localization, mapping and crop monitoring in citrus tree farms," in *International Symposium on Visual Computing*, pp. 571–582, Springer, 2023.
- [29] X. Kan, H. Teng, and K. Karydis, "Online exploration and coverage planning in unknown obstacle-cluttered environments," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5969–5976, 2020.
- [30] X. Kan, T. C. Thayer, S. Carpin, and K. Karydis, "Task planning on stochastic aisle graphs for precision agriculture," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3287–3294, 2021.