



From entities to areas: A semantically driven clustering approach for area delimitation on 3D scene graphs^{☆,☆☆,★}

Mario A.V. Saucedo^{1,*}, Akash Patel¹, Taxiarchis-Foivos Blounas, Christoforos Kanellakis, George Nikolakopoulos

Robotics and AI Group, Luleå University of Technology, Luleå, SE 97187, Sweden

ARTICLE INFO

Keywords:

Robotics
Robot perception and sensing
Semantic scene understanding
Hierarchical graph representations
Formal algorithms and methods

ABSTRACT

3D scene graph (3DSG) generation is a rapidly evolving field that plays a significant role in robotic autonomy. Traditionally, the focus has been on indoor environments, where robots understand and navigate spaces by abstracting objects and geometric information in a structured graph format. Expanding upon this idea, this paper introduces a 3DSG construction architecture, which enables scene-agnostic abstraction of the environment, with the goal of facilitating the adoption of 3DSG for autonomous agents in both indoor and outdoor environments. We propose a novel approach for area delimitation in 3DSGs that leverages label propagation to cluster entities (i.e. objects of interest) into areas that are both semantically and topologically distinguishable within a scene. Towards this end, we establish label propagation for 3DSGs, by formulating a dynamic set of propagation factors that accommodate to the relevance of semantic information and their natural decay through the topological structure of the 3DSG. Additionally, to achieve scene-agnostic area delimitation, we introduce a single-step optimization process for the calculation of clutter-aware propagation factors based on the approximation of an optimal set of factors that maximize inter-area eccentricity while minimizing intra-area eccentricity. Finally, the proposed framework is extensively validated through simulations and real-world deployments using a Boston Dynamics Spot legged robot and a Clearpath Husky mobile robot. The experimental results showcase the scalability of the proposed framework to indoor and outdoor environments for real-time 3DSG construction.

1. Introduction

In recent years, 3D Scene Graphs (3DSGs) have been broadly adopted in a wide range of robotic applications such as object search [1], semantic-driven navigation [2,3], and semantic scene understanding [4–7]. Since their introduction in [8], 3DSGs have been used as an abstraction of indoor environments into semantically structured graph representations based on layers, where each layer has a set of nodes, each node has a set of attributes, and the edges between nodes represent their relationships. Nevertheless, the integration of 3DSGs in robotic missions oriented at unstructured scenes has been hindered by the restrictive nature of the first proposed, and still widely used, set of layers. Despite numerous robotic applications having started

to move towards the use of graph representations, as in the case of exploration [9], inspection [10] and multi-agent collaboration [11], the current state-of-the-art still seldom uses 3DSGs for field-oriented applications.

In recent works [12–16], 3DSGs share three core layers: (i) Building, (ii) Rooms, and (iii) Entities (i.e. objects and agents), where the building typically defines the robot's workspace, which is subdivided into rooms that contain the entities. However, room delimitation is usually based solely on geometric information, defining rooms as a geometric property rather than a semantically meaningful division of the scene. While in most indoor environments rooms are both geometrically and semantically meaningful, in unstructured scenes, where areas

[☆] This article is part of a Special issue entitled: 'TC 4.3 Robotics (IFAC WC 2026)' published in Mechatronics.

^{☆☆} This work has been partially funded by the European Union's Horizon Europe Research and Innovation Programme, under the Grant Agreement No. 101138330 CIRCULess.

[★] The code used in this study can be made available upon request. For additional multimedia material of the simulations and field evaluations of this work, please refer to the following link: <https://sites.google.com/view/from-entities-to-areas/home>.

* Corresponding author.

E-mail address: marval@ltu.se (M.A.V. Saucedo).

¹ These authors contributed equally to this work.

are not physically constrained or clearly separated, purely geometric approaches become insufficient.

In this work, we depart from the state-of-the-art paradigm of dividing scenes into “rooms” and instead propose a more meaningful spatial delimitation based on the entities that populate the environment. Starting from the observation that, in human-centric environments, entities (i.e. objects) form the base unit of reasoning for higher-level abstractions, we delimit the environment into semantically definable areas by clustering entities. This enables scene segmentation even in open spaces. To address this challenge, we propose a novel method for area delimitation in 3DSGs that leverages label propagation to cluster entities into semantically and geometrically distinguishable areas. The proposed method is integrated into a state-of-the-art 3DSG construction framework and experimentally validated in real-world environments.

— Contributions

The focus of this work is to allow for a more versatile abstraction of structured and unstructured environments throughout 3DSGs. Towards this end, we introduce a formulation of the label propagation algorithm for 3DSGs that replaces the fixed similarity matrix with a dynamic set of propagation factors reflecting the decay on the relevance of the labels after each propagation step (Section 3.2). This new formulation allow to account for the natural relevance decay of the semantic information contained in the *Entities Layer* of a given 3DSG when propagated through the topological sub-graph of the *Places Layer*.

Furthermore, we leverage the proposed formulation for the area delimitation task by optimizing the initial set of propagation factors with a single-step optimization based on the dispersion of the *Entities*. The proposed method, namely clutter-aware propagation factors, allows us to approximate the optimal set of factors that maximize the inter-Area eccentricity while minimizing the intra-Area eccentricity on a 3DSG (Section 3.3).

The proposed method has been integrated into a state-of-the-art 3DSG construction framework and experimentally tested in indoor and outdoor environments. The extensive experimental results prove the adaptability of the framework to different environments and showcase the scalability and real-time performance of the proposed method.

2. Related work

3D Scene graphs have emerged as a popular alternative for the encoding of scenes into intuitive and manageable data representations [17]. Among the first works to introduce this concept we can find [8], where a semi-automatic framework for the construction of 3D scene graphs is presented. In this pioneering work the authors leverage the semantic parsing method presented in [18] for the delimitation of room elements in indoor environments. Subsequent works as [12], make use of a 3D Euclidean Signed Distance Function (ESDF) to partition the environment into different rooms. The presented method works by taking a horizontal 2D section of the 3D ESDF, cut below the level of the detected ceiling, the result of this partitioning operation is a set of disconnected 2D ESDFs corresponding to each room.

Furthermore, Bavle et al. [15,16], introduced a formulation for the room and corridor nodes, where both can be represented as a set of planes, such that, a room node is defined as composed of four planar walls, and a corridor node as composed of two parallel wall planes. Their topological layer assesses if the mapped planes at a given time instance belong to a particular object or structural component and further constrain their geometry. On the other hand, the authors of Hughes et al. [13,14], proposed the use of dilation operations on the voxel-based map to help expose rooms in the environment. The idea behind this concept is based in the fact that these operations can be directly mapped into topological changes in the sub-graph of “places”, which serves as a topological map, where each place is an obstacle-free location and an edge between places denotes straight-line

traversability. Nevertheless, the previous methods for room detection only reasons over the topology of the environment, as such, even in indoor environments, they will fail to segment semantically distinct areas in wider spaces (e.g. an open floor-plan).

On the other hand, works such as Greve et al. [19] and Deng et al. [20] have addressed the challenge of 3DSG construction for outdoor environments by reasoning about intersections as structural elements of the graph. These intersection entities serve to aggregate multiple incoming and outgoing segments (i.e., roads), enabling the graph to capture higher-order connectivity and navigational affordances of the environment. Nevertheless, despite their effectiveness in structured outdoor settings, these approaches remain limited by their strong reliance on well-defined topological patterns. As a result, their intersection detection mechanisms are not suited for unstructured open spaces or areas with ambiguous or evolving connectivity, where clear segment convergence or trajectory branching is absent.

Among the works trying to address these challenges we can find [21], where an approach to build 3D scene graphs in arbitrary indoor and outdoor environments is presented. The authors formulate region identification as a node classification problem on the “places” layer, where given a spatial ontology defining concepts and relations for indoor and outdoor environments, they employ a Graph Neural Network (GNN) for generating node embeddings and a Multi Layer Perceptron (MLP) for predicting node labels. Likewise, Maggio et al. [22] identifies regions in indoor environments by constructing a set of place primitives and then incrementally clustering them using an agglomerative Information Bottleneck (IB) algorithm, which groups place primitives into higher-level region nodes based on their semantic similarity and relevance to a set of given natural-language tasks.

However, both previous methods rely on prior semantic knowledge encoded in predefined ontologies and label sets to perform classification, constraining their ability to discover novel or previously unseen spatial concepts. While this semantic grounding improves interpretability and consistency, it also limits adaptability, as the resulting room or region definitions are bounded by the expressiveness and completeness of the prior semantic models rather than emerging purely from the observed structure of the environment. In contrast to previous works, our proposed method requires no prior knowledge about the environment or the entities. Instead, it leverages the proposed label propagation method for area delimitation by using the encoded semantic and topological information on the *Entities* and *Places* layers.

3. Methodology

3.1. Preliminaries

3D Scene Graph: Let us define a 3DSG as a hierarchical representation composed of multiple abstraction layers, where the abstractions are defined based on the resolution of the information they encode [8]. To that extent, we denote $\mathbb{L} = [\textit{Scene}, \textit{Areas}, \textit{Entities}, \textit{Places}]$, as the set of abstraction layers that constitute a 3DSG. Where each layer is defined as follows:

- *Scene*: The top layer represents the environment of the robotic platform.
- *Areas*: Distinguishable sub-classifications of the scene that can be semantically labeled.
- *Entities*: Any semantically labeled class that can be classified as an object.
- *Places*: A sparse topological representation of the unoccupied space within the scene.

Let $\mathcal{G}^l = (\mathcal{V}^l, \mathcal{E}^l, \mathcal{A}^l)$ be an undirected sub-graph representing the layer $l \in \mathbb{L}$ with an ordered set of nodes $\mathcal{V}^l$ and edges $\mathcal{E}^l$. Let $a_i^l \in \mathcal{A}^l$

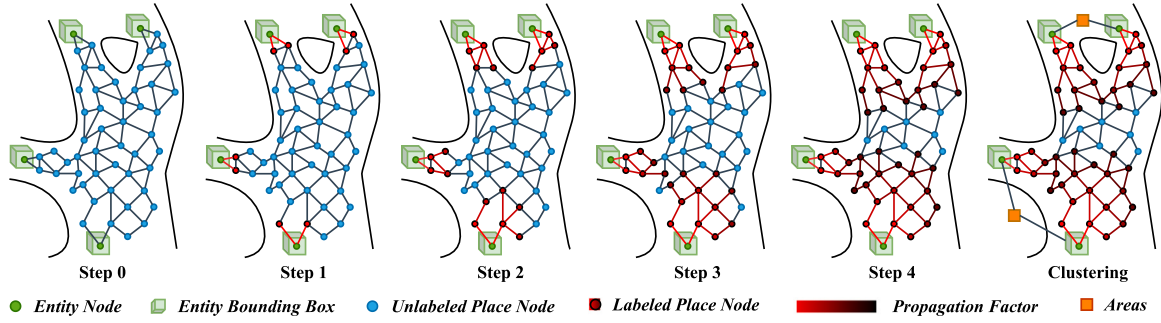


Fig. 1. Area delimitation on 3DSGs. Overview of the proposed area delimitation method based on the introduced label propagation algorithm. The process starts by propagating the semantic information of the *Entities* into the *Places*, while progressively updating the set of labeled and unlabeled nodes. Finally, the *Entities* are clustered into *Areas* such that those within the same cluster are connected by edges belonging to only labeled nodes.

be the attribute of a node $v_i^l \in \mathcal{V}^l$ encoded with necessary metric and semantic information.

Eccentricity: In graph theory, the eccentricity of a vertex is defined as the greatest distance between that vertex and any other vertex in the graph

$$\epsilon(v) = \max_{u \in \mathcal{V}} d(v, u). \quad (1)$$

For the purposes of this article we define the eccentricity of a 3DSG in two ways: (i) intra-layer eccentricity and (ii) inter-layer eccentricity.

The first one is defined as:

$$\epsilon(v_i^l) = \max_{\substack{u_j, u_k \in \mathcal{V}^{l+1} \\ u_j \neq u_k}} d(u_j, u_k), \quad (2)$$

$$\text{s.t. } e_{ij}, e_{ik} \in \mathcal{E}^{l+1}, \quad e_{ij}, e_{ik} \neq 0,$$

where $\epsilon(v_i^l)$ represents the intra-layer eccentricity of $v_i^l \in \mathcal{V}^l$ and u_j, u_k are child nodes of v_i^l belonging to the layer $\mathcal{G}^{l+1}$ (i.e., the next layer from top to bottom).

While the second one is defined as:

$$\epsilon(\mathcal{V}^l) = \max_{\substack{v_i^l, v_j^l \in \mathcal{V}^l \\ v_i^l \neq v_j^l}} d(v_i^l, v_j^l), \quad (3)$$

$$\text{s.t. } e_{ij} \in \mathcal{E}^l, \quad e_{ij} \neq 0,$$

representing the inter-layer eccentricity of the layer $\mathcal{G}^l$.

3.2. Label propagation

For any given 3DSG we define label propagation as a function in terms of two connected layers $(\mathcal{G}^l, \mathcal{G}^{l+1}) \mapsto \mathcal{W}$ where the objective is to propagate semantic information from the upper layer $\mathcal{G}^l$ to the subsequent layer $\mathcal{G}^{l+1}$. Let the initial set of labeled nodes, with known labels $\mathcal{C} = \{c_0, c_1, \dots, c_n\} \in \mathbb{Z}^+$, be denoted as $\mathcal{V}^L = \mathcal{V}^l$, and the set of unlabeled nodes as $\mathcal{V}^U = \mathcal{V}^{l+1}$. Let $\mathcal{W}_l$ be the initial set of propagation factors represented by the biadjacency matrix

$$\mathcal{W}_l = \begin{bmatrix} 0 & \Omega \\ \Omega^\top & 0 \end{bmatrix}, \quad (4)$$

where Ω is a matrix of size $|\mathcal{V}^L| \times |\mathcal{V}^U|$, which contains the propagation factors between $\mathcal{V}_l^L$ and $\mathcal{V}_l^U$.

Thereafter, all unlabeled nodes $j \in \mathcal{V}_l^U$ are updated based on the neighboring labeled nodes as follows:

$$\hat{c}_j = \sum_{i \in \mathcal{N}_l^L(j)} w_{ij} c_i, \quad (5)$$

where $\mathcal{N}_l^L(j) = \{i \in \mathcal{V}_l^L | w_{ij} \in \mathcal{W}_l, w_{ij} \neq 0\}$ is the set of neighboring labeled nodes. Likewise, the new factors matrix $\mathcal{W}_{l+1}$ is computed using the update rule

$$w_{jk} = \max\{w_{ij} | i \in \mathcal{N}_l^L(j)\} - \lambda, \quad (6)$$

for all propagation factors $w_{jk} | k \in \mathcal{N}_l^U(j)$ where $\mathcal{N}_l^U(j) = \{k \in \mathcal{V}_l^{l+1} | e_{jk} \in \mathcal{E}^{l+1}, e_{jk} \neq 0\}$ and $\lambda \in \mathbb{R}^+$ is a decay factor.

After all nodes in $\mathcal{V}_l^U$ have been parsed the sets $\mathcal{V}_{l+1}^L$ and $\mathcal{V}_{l+1}^U$ are recomputed and the updating process is repeated using the new factors matrix $\mathcal{W}_{l+1}$. The process is terminated once all propagation factors for any remaining unlabeled nodes fall under a defined threshold,

$$\max_{w \in \mathcal{W}_{l+m}} w \leq \tau. \quad (7)$$

3.3. Area delimitation

In the case of area delimitation, we seek to propagate the semantic information encoded on the *Entities* into the geometric information of the *Places* Layer. The resulting propagation factors are then used to cluster the *Entities* such that those belonging to the same cluster are connected without traversing edges belonging to unlabeled nodes. A depiction of this process is presented in Fig. 1.

The overall task can then be formulated as an optimization problem, where the objective is to determine the optimal initial set of propagation factors $\mathcal{W}^*$ that maximize the eccentricity among *Areas* (i.e. inter-Area) while minimizing the eccentricity of *Entities* belonging to the same *Area* (i.e. intra-Area), expressed as:

$$\mathcal{W}^* = \arg \max_{\mathcal{W}} \left(\underbrace{\epsilon(\mathcal{V}^{Areas})}_{\text{inter-Area}} - \gamma \max_{a \in \mathcal{V}^{Areas}} \underbrace{\epsilon(a)}_{\text{intra-Area}} \right), \quad (8)$$

where the parameter γ balances the trade-off between maximizing semantic separation across *Areas* and minimizing spatial dispersion within each *Area*. Higher γ favors compact, semantically tight *Areas*, while lower values promote broader region aggregation. However, the search space of possible propagation configurations grows combinatorially with the number of edges and nodes in the 3DSG, making it an NP-hard problem.

To overcome this limitation, we introduce a single-step optimization that approximates the optimal set of propagation factors $\mathcal{W}^*$ based on the dispersion of the *Entities*. The intuition is that regions with higher clutter (i.e., denser local neighborhoods) should yield smaller propagation factors to prevent over-segmentation, while sparser regions should favor larger factors to preserve connectivity across open spaces. These insights motivate our approach for clutter-aware propagation factors, where for each node $i \in \mathcal{V}^{Entities}$, we approximate its optimal propagation factor as:

$$\hat{w}_i = \left(\sum_{j \in \mathcal{N}_p(i)} d(i, j) \right) - \min\{d(i, j) | j \in \mathcal{N}_p(i)\}, \quad (9)$$

where $\mathcal{N}_p(i)$ denotes the set of p nearest neighbors of the *Entity* node i . This formulation reduces the complexity from an intractable combinatorial search to a locally computed update, making it suitable for online execution while retaining sensitivity to the spatial distribution of the environment.

Table 1

Sensitivity of the proposed method to key parameters. The values on each cell correspond to consistent, ambiguous and inconsistent.

	$p = 2$			$p = 3$			$p = 4$		
	$\tau = 0.0$	$\tau = 0.5$	$\tau = 1.0$	$\tau = 0.0$	$\tau = 0.5$	$\tau = 1.0$	$\tau = 0.0$	$\tau = 0.5$	$\tau = 1.0$
$\lambda = 0.5$	11.84%	14.36%	17.29%	8.57%	9.34%	10.97%	8.10%	8.21%	8.21%
	2.33%	2.59%	3.57%	2.17%	2.29%	2.75%	2.17%	2.17%	2.17%
	85.83%	83.05%	79.14%	89.26%	88.37%	86.28%	89.73%	89.62%	89.62%
$\lambda = 1.0$	34.62%	34.62%	51.76%	22.69%	22.69%	32.79%	17.17%	17.17%	23.19%
	7.74%	7.74%	16.08%	4.71%	4.71%	9.98%	3.61%	3.61%	5.54%
	57.63%	57.63%	32.17%	72.60%	72.60%	57.23%	79.22%	79.22%	71.27%
$\lambda = 1.5$	50.80%	60.06%	64.31%	41.58%	46.95%	52.90%	33.54%	37.81%	42.25%
	12.49%	15.71%	15.06%	10.01%	12.51%	14.12%	6.27%	8.86%	12.12%
	36.71%	24.24%	20.63%	48.41%	40.54%	32.98%	60.19%	53.33%	45.64%

4. Experimental evaluation

In this section, we first present a performance comparison between the proposed method, ground truth data, and a room segmentation baseline, demonstrating the robustness and consistency of our approach in indoor environments. Next, we conduct a scalability analysis of the proposed label propagation algorithm, highlighting the impact of the clutter-aware propagation factors, followed by an analysis on the effects of sparsity. Finally, we describe the modules involved in constructing a 3DSG using the proposed framework and present extensive experiments conducted in real-world outdoor environments.

4.1. Evaluation metric

The evaluation metric was the consistency rate measured as the percentage of the segmented rooms to be classified in one of the following three categories:

- *Consistent*: The room contains one or more areas that encompass all the entities in the room.
- *Ambiguous*: The room contains one or more areas that encompass most of the entities in the room.
- *Inconsistent*: The room does not contain any areas, or the areas fail to encompass most of the entities in the room.

4.2. Scalability and stability analysis

Sensitivity to Key Parameters: We evaluate the sensitivity of the proposed method to the decay factor λ , stopping threshold τ , and neighborhood size p using the Habitat-Matterport 3D dataset [23]. The results are summarize on Table 1, where we can observe that low λ values lead to increasing ambiguity or inconsistency, while larger values improve segmentation coherence, likewise τ regulates propagation depth and improves the overall area formation. Finally, the neighborhood size p has a stabilizing effect, where lower values reduce sensitivity to local spatial variations.

Graph Size: Fig. 2 shows the computation time (ms) for the proposed label propagation algorithm using clutter-aware propagation factors under different graph sizes. As expected, the computational cost increases more significantly with the number of *Entities* in comparison to the number of *Places*, by approximately a 10X factor. This quality allows the algorithm to better scale to outdoors and open spaces where the relation between *Entities* and *Places* is lower than in structured indoor environments.

Variable Sparsity: The test environment consisted of the Satsop nuclear power plant virtual environment, from the DARPA subterranean challenge [24]. The environment is further divided in 3 zones, each one presenting a different degree of sparsity on the *Places Layer*, such that Zone [i] is the densest and Zone [iii] the sparsest, as depicted in Fig. 3. Furthermore, all rooms that compose each of the zones, and contain at least one entity, have been classified into one of three categories

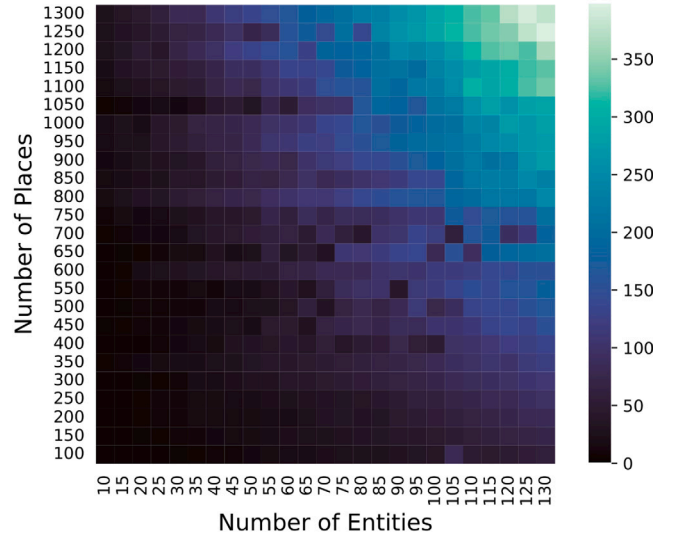


Fig. 2. Computation time (ms). Computation time needed to estimate the areas using the proposed method under different graph configurations.

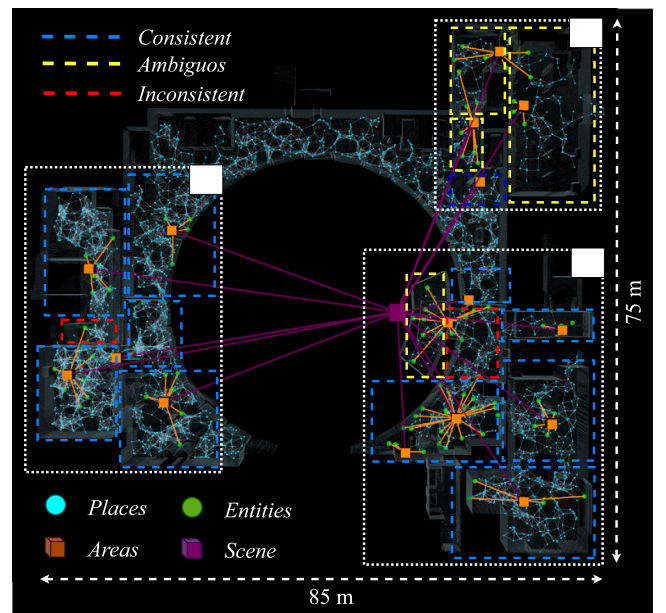


Fig. 3. Sparsity effect. Visualization of the effect of sparsity when generating a 3DSG using the proposed method, where the three main zones in the environment present each a different degree of sparsity on the *Places Layer*, from dense [i] to sparse [iii].

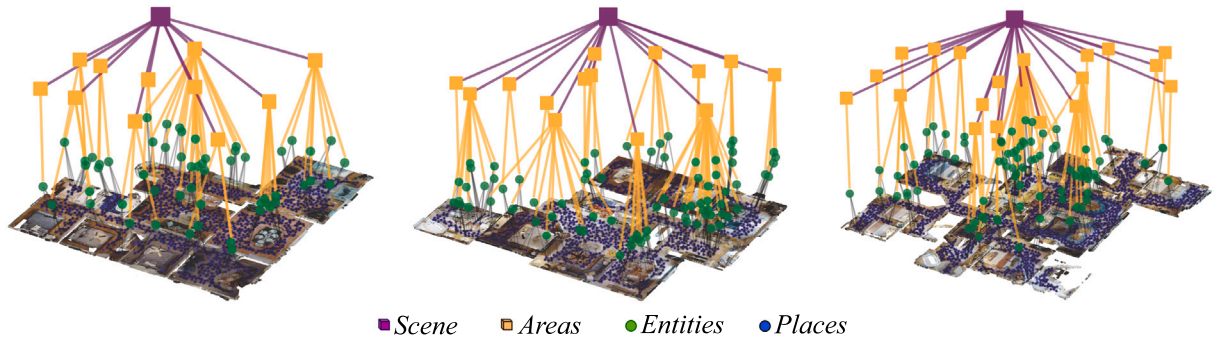


Fig. 4. Indoor experiments. Depiction of the *Area* delimitation performance of the proposed clutter-aware propagation factors on different environments of the Matterport 3D dataset, showcasing the advantages of the proposed method to avoid overlapping while improving the clustering overall performance.

Table 2

Performance under variable sparsity.

Category	Zone [i]	Zone [ii]	Zone [iii]
Consistent	83.4%	71.4%	25.0%
Ambiguous	0.0%	14.3%	75.0%
Inconsistent	16.6%	14.3%	0.0%

(Section 4.3) based on the computed areas. Table 2 summarizes the obtained results for each of the zones, and illustrates how an increase in the sparsity of the *Places Layer* directly affects the performance of the proposed method, resulting in a greater ambiguity of the delimited areas.

4.3. Performance comparison

Ground Truth Comparison: We employed the Habitat-Matterport 3D dataset² [23], one of the largest collections of real-world building-scale 3D scenes encompassing residential, commercial, and civic environments, as the ground truth reference for this comparison. Fig. 4 illustrates representative examples from the dataset alongside the corresponding segmented areas using the proposed clutter-aware propagation factors. Overall, the proposed method effectively mitigates overlap in densely cluttered regions while enhancing the coherence and separation of clusters in sparser or more open spaces.

Furthermore, the comparison in Fig. 5 evaluates the proposed clutter-aware propagation factors (i.e., adaptive) against three fixed configurations: the global mean, lower quartile (Q1), and upper quartile (Q3) of the clutter-aware factors. The fixed values were statistically estimated from the distribution of the clutter-aware propagation factors computed over all samples in the dataset, thereby representing central and boundary behaviors of the adaptive scheme. The results show that the adaptive approach achieves the highest proportion of consistent room segmentations while maintaining low variability across scenes. In contrast, the fixed configurations exhibit reduced performance, with a notable increase in ambiguous and inconsistent cases. Specifically, the mean factor fails to capture local scene variability, leading to under- or over-propagation in different contexts, whereas the Q1 and Q3 settings further exaggerate these effects by biasing the propagation towards overly conservative or overly diffusive behavior, respectively. These findings underscore the importance of dynamically adapting propagation strength to the structural and clutter characteristics of each environment, thereby enhancing robustness and generalization.

On the other hand, the adaptive clutter-aware propagation introduces a higher computational cost compared to fixed propagation factors, due to the per-sample estimation of local clutter statistics and

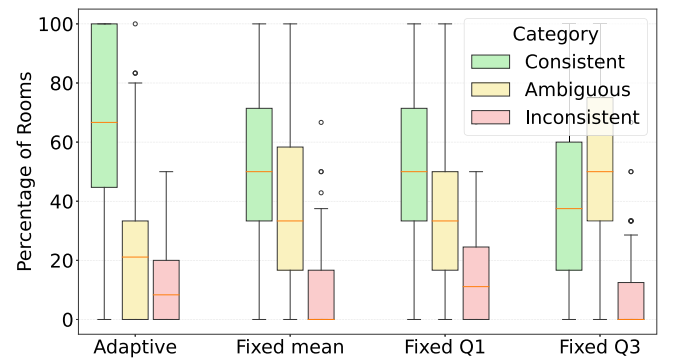


Fig. 5. Clutter-aware propagation factors. Comparison for the *Area* delimitation performance of the proposed clutter-aware propagation factors (i.e., adaptive) against fixed propagation factors.

the dynamic adjustment of propagation strength. As a result, the adaptive method exhibits a higher average processing time per sample of 0.154 s, compared to the fixed Q1, mean, and Q3 settings, ranging from 0.008 s to 0.010 s. However, this overhead remains modest in absolute terms and scales linearly with the number of samples. Importantly, the fixed configurations benefit from precomputed global statistics that eliminate runtime adaptation but do so at the expense of robustness to scene-specific structural variability.

Room-Segmentation Comparison: For the purposes of this comparison, the proposed method is evaluated against room-based strategies for the task of room segmentation in indoor environments. We select the latest implementation of the Hydra³ framework presented in [13,14] as the baseline, and conduct experiments on the uHumans2 dataset [12]. For this comparison, we employ the places layer provided by the Hydra framework. This choice is consistent with our problem formulation, as our method does not impose any specific requirements on the place generation process, beyond capturing a generic description of the environment through a sparse topological representation. The comparison results showcase an overall consistency of 71.4% for the proposed method, with an ambiguity rate of 28.6% and no inconsistencies. Among the zones with ambiguity, we could observe a trend where objects placed in close proximity to the point of union between two rooms might be clustered together (e.g. a plant and a trash can located on opposite sides of a door).

On the other hand, in the zones where the proposed method was consistent with the baseline, it also presented the desired behavior of delimiting large rooms (e.g. long halls or open floor plans) into smaller areas (e.g. different working zones within the same room). Fig. 6

² <https://github.com/matterport/habitat-matterport-3dresearch>.

³ <https://github.com/MIT-SPARK/Hydra>.

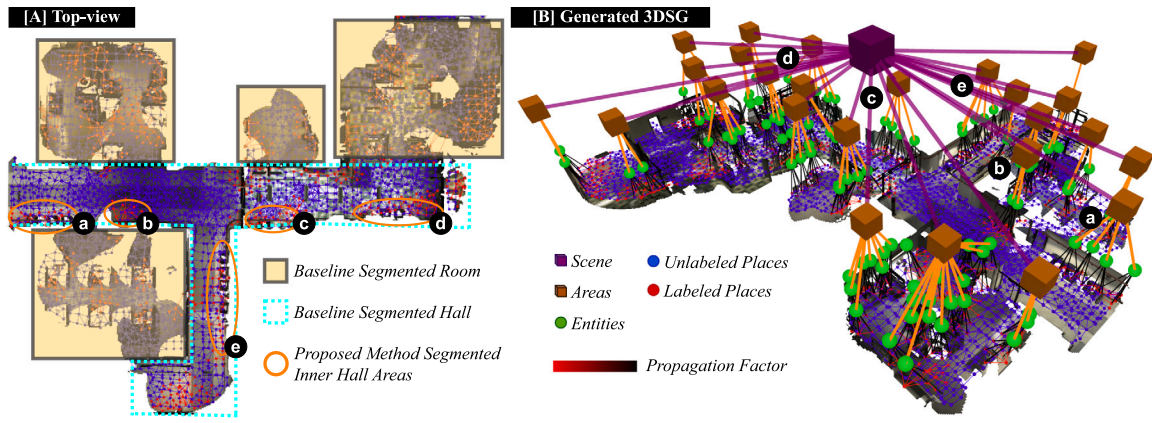


Fig. 6. Geometric baseline comparison. Depiction of the segmented areas using the propose method versus the geometric-based baseline on a long hall with multiple waiting zones along the way.

depicts one of this scenarios where a long hall is segmented as a single room by the baseline, while the proposed method identifies multiple areas representing different waiting zones along the main hall. This behavior is not limited to elongated spaces, but is consistently observed across other rooms where the proposed approach further refines the spatial abstraction by decomposing large or multifunctional rooms into semantically meaningful sub-areas. This results in a higher degree of granularity, where distinct functional zones are explicitly represented despite sharing the same enclosing geometry.

Consequently, the generated areas are not only consistent with traditional room-level segmentation but also provide richer semantic grounding, enabling the representation to better capture the internal organization and usage patterns of the space. This is particularly advantageous for robotic applications as it enables finer-grained spatial reasoning and more task-aware behaviors. Allowing robots to adapt to local context, reduce perceptual ambiguity in elongated spaces, and support more robust planning and interaction in human-centric environments.

4.4. Field experiment results

Graph Construction Pipeline: The proposed method can be implemented in any given 3DSG as long as it contains an *Entities Layer* and a *Places Layer*, thus it can be adapted to already existing frameworks such as Rosinol et al. [12] and Hughes et al. [13,14]. Nevertheless, most of the frameworks are not designed for outdoors, reason why we have developed a 3DSG construction pipeline, based on state-of-the-art modules, structured for both indoors and outdoors.

- (1) *Places Layer*: In this work, we generate the *Places Layer* using the exploration architecture implemented in the STAGE planner framework [9]. As the robot navigates the environment, a collision-free dense graph is incrementally built from sparsely sampled poses along its exploration path.
- (2) *Entities Layer*: The construction of the *Entities Layer* is an extension of the object detection and localization framework presented in [25]. The resulting bounding boxes are then used to populate $\mathcal{V}^{Entities}$. The set of edges $\mathcal{E}^{Entities}$ is then obtained using a k-d tree nearest neighbor search combined with collision checks, ensuring that only valid connections are established.
- (3) *Areas Layer*: The construction of the *Areas Layer* is based on the presented label propagation algorithm (Section 3.2), where the initial propagation factors are selected based on the proposed single-step optimization (Section 3.3).
- (4) *Scene Layer*: The final layer consists of a scene node connecting all delimited *Areas*.

Outdoor Experiment I: The experiment was conducted in an urban outdoor environment characterized by a wide open space surrounded by buildings using a Boston Dynamics Spot legged robot. Fig. 7 presents the generated 3DSG for this environment, along with representative images of the *Entities* detected within the delineated *Areas*. These results illustrate the capability of the proposed method to generate semantically coherent and distinguishable *Areas* even in unstructured outdoor settings, for instance, identifying a bicycle parking spot in Fig. 7.E or a group of people in Fig. 7.D.

Outdoor Experiment II: This experiment was conducted in an urban environment composed of multiple parking lots using a Husky mobile robot, where the detected entities were filtered to only include vehicles. Fig. 8 shows the generated 3DSG for this environment, along with images of the *Entities* present in the delimited *Areas*. This experiment highlights the capability of the proposed method to intrinsically understand the underlying nature of human-designed outdoor environments. In particular, it overcomes two main challenges. The first one is the presence of open-concept spaces, such as the one shown in Fig. 8.C and .D, which exhibits a partially defined geometry due to nearby buildings yet cannot be categorized as a single semantically distinguishable zone. The second challenge is the segmentation of open space with human-defined abstract boundaries, as is the case in Fig. 8.E and .F. Despite the lack of any geometries delimiting both parking lots, the framework is still able to delimit them as two separate areas.

4.5. Discussions

This section will elaborate on some of the properties of the proposed method to have in consideration during its implementation. First, while it is independent of the number and semantic labeling of the *Entities*, the distribution of the *Entities* actively affects the delimitation of *Areas*, in other words, different semantic classes may present different distributions and result in different *Areas* within the same environment. Moreover, the sparsity of the *Places Layer* will also actively change the delimited *Areas*, however, based on the experimental evaluation, if the sparsity of the *Places Layer* remains consistent across different topologies, then the same set of parameters will be functional regardless of the environment. Finally, walls surrounded geometries (e.g. rooms) will not be delimited as *Areas* unless they contain *Entities*. While the topology of the environment, encoded as the *Places Layer*, is considered for the delimitation of *Areas*, it is the *Entities* that define the *Areas*, both topologically and semantically.

Nevertheless, by grounding area delimitation on the distribution and semantics of entities rather than on structural boundaries alone, the resulting areas exhibit a higher degree of granularity and semantic

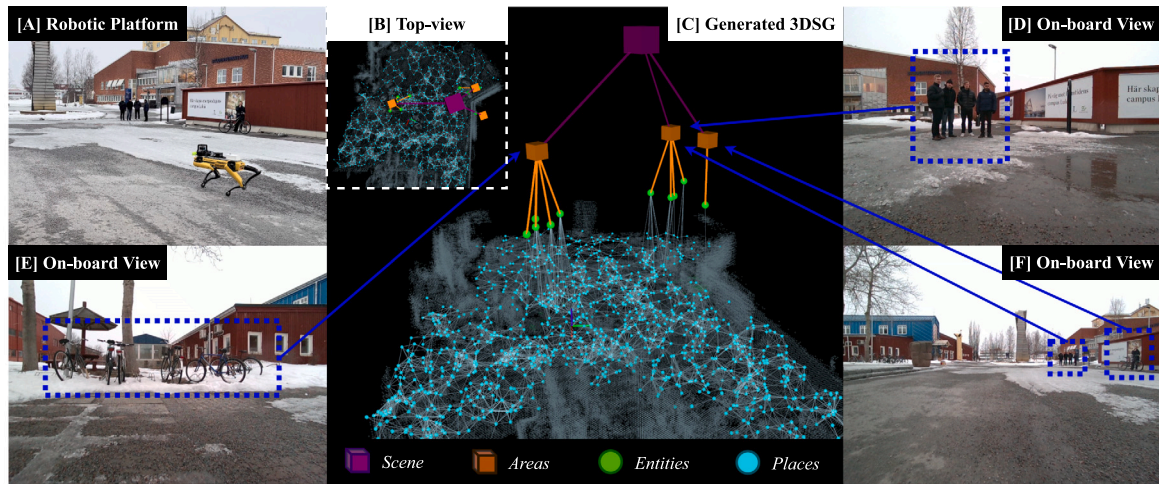


Fig. 7. Outdoor Experiment I. Depiction of the generated 3DSG using the proposed framework [C], alongside a top-view visualization of the environment [B] and the robotic platform used in the experiment [A]. Furthermore, [E]–[F] present visualizations of the detected *Entities* and their corresponding *Areas*.

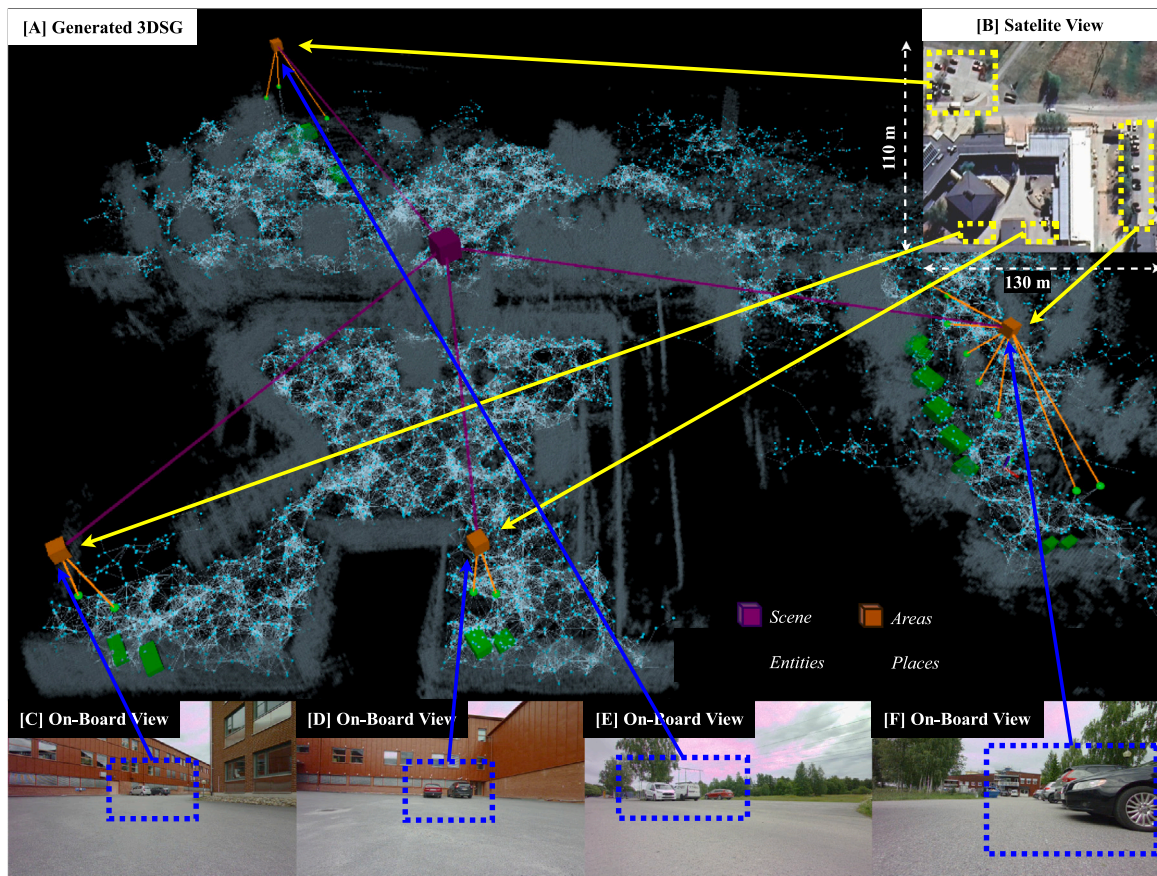


Fig. 8. Outdoor Experiment II. Depiction of the generated 3DSG using the proposed framework [A], alongside a satellite image of the environment [B], highlighting the parking lots corresponding to each *Area*. Furthermore, [C]–[F] present visualizations of the detected *Entities* and their corresponding *Areas*.

coherence. This allows the framework to capture functional sub-regions within large or open spaces that would otherwise be merged into a single region under geometry-only segmentation. As a result, areas are not merely spatial partitions but semantically grounded abstractions that reflect how environments are actually structured and used.

This semantic grounding further enables the proposed method to adapt naturally across different environments, including unstructured

and outdoor scenes where clear geometric delimiters are absent. Since area formation emerges from entity-level interactions and their propagation through the topological structure, the framework remains robust to variations in layout, scale, and scene typology without relying on predefined room concepts or environment-specific priors.

This adaptability is particularly advantageous in robotics, where higher-level reasoning, task planning, and human-aware interaction

benefit from spatial abstractions that encode meaning rather than geometry alone. By producing areas that align more closely with functional and semantic context, the proposed representation facilitates reasoning over space, supports more informed decision-making, and lays the groundwork for integrating symbolic reasoning, task constraints, and commonsense knowledge into autonomous robotic systems.

5. Conclusions

This article introduced a 3DSG construction architecture, which enables scene-agnostic abstraction of the environment, with the goal of facilitating the adoption of 3DSG for autonomous agents in both indoor and outdoor environments. We presented a novel approach for area delimitation in 3DSGs that leverages label propagation to cluster *Entities* into *Areas* that are both semantically and topologically distinguishable within a scene. Furthermore, we establish a label propagation formulation for 3DSGs, by introducing a dynamic set of propagation factors that accommodate to the relevance of semantic information and their natural decay through the topological structure of the 3DSG. In addition, we also develop a single-step optimization process for the calculation of clutter-aware propagation factors to achieve scene-agnostic area delimitation, by approximating the optimal set of factors that maximize inter-area eccentricity while minimizing intra-area eccentricity. Finally, we provide extensive simulation and experimental results from different environments, showcasing the scalability of the framework to indoor and outdoor environments, for real-time 3DSG construction.

CRedit authorship contribution statement

Mario A.V. Saucedo: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Akash Patel:** Writing – review & editing, Software, Methodology. **Taxiarchis-Foivos Blounas:** Visualization, Software, Data curation. **Christoforos Kanellakis:** Writing – review & editing, Supervision. **George Nikolakopoulos:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used the ChatGPT model developed by OpenAI, in order to correct grammar and refine the wording of the manuscript. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Mario Alberto Valdes Saucedo reports financial support was provided by European Union's Horizon Europe Research and Innovation Programme, under the Grant Agreement No. 101138330 CIRCULess. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- [1] Giuliari F, Skenderi G, Cristani M, Bue AD, Wang Y. Leveraging commonsense for object localisation in partial scenes. *IEEE Trans Pattern Anal Mach Intell* 2023;45(10):12038–49. <http://dx.doi.org/10.1109/TPAMI.2023.3272523>.
- [2] Ginting MF, Kim S-K, Fan DD, Palieri M, Kochenderfer MJ, Akbar Aghamohammadi A. SEEK: Semantic reasoning for object goal navigation in real world inspection tasks. In: *Proc. of robotics: science and systems*. 2024.
- [3] Ginting MF, Kim S-K, Peltzer O, Ott J, Jung S, Kochenderfer MJ, Aghamohammadi A-a. Safe and efficient navigation in extreme environments using semantic belief graphs. In: *2023 IEEE international conference on robotics and automation. ICRA*, 2023, p. 5653–8. <http://dx.doi.org/10.1109/ICRA48891.2023.10161056>.
- [4] Saucedo MA, Viswanathan VK, Kanellakis C, Nikolakopoulos G. Estimating commonsense scene composition on belief scene graphs. In: *2025 IEEE international conference on robotics and automation. ICRA*, 2025, p. 2861–7. <http://dx.doi.org/10.1109/ICRA55743.2025.11127920>.
- [5] Saucedo MA, Patel A, Saradagi A, Kanellakis C, Nikolakopoulos G. Belief scene graphs: Expanding partial scenes with objects through computation of expectation. In: *2024 IEEE international conference on robotics and automation. ICRA*, 2024, p. 9441–7. <http://dx.doi.org/10.1109/ICRA57147.2024.10611352>.
- [6] Saucedo MA, Stathouloupoulos N, Patel A, Kanellakis C, Nikolakopoulos G. Leveraging computation of expectation models for commonsense affordance estimation on 3D scene graphs. In: *2024 IEEE/RSJ international conference on intelligent robots and systems. IROS*, 2024, p. 9797–802. <http://dx.doi.org/10.1109/IROS58592.2024.10802560>.
- [7] Amiri S, Chandan K, Zhang S. Reasoning with scene graphs for robot planning under partial observability. *IEEE Robot Autom Lett* 2022;7(2):5560–7. <http://dx.doi.org/10.1109/LRA.2022.3157567>.
- [8] Armeni I, He Z-Y, Zamir A, Gwak J, Malik J, Fischer M, Savarese S. 3D scene graph: A structure for unified semantics, 3D space, and camera. In: *2019 IEEE/CVF international conference on computer vision. ICCV*, 2019, p. 5663–72. <http://dx.doi.org/10.1109/ICCV.2019.00576>.
- [9] Patel A, Saucedo MA, Kanellakis C, Nikolakopoulos G. STAGE: Scalable and traversability-aware graph based exploration planner for dynamically varying environments. In: *2024 IEEE international conference on robotics and automation. ICRA*, 2024, p. 5949–55. <http://dx.doi.org/10.1109/ICRA57147.2024.10610939>.
- [10] Viswanathan VK, Saucedo MA, Satpute SG, Kanellakis C, Nikolakopoulos G. xFLIE: Leveraging actionable hierarchical scene representations for autonomous semantic-aware inspection missions. 2024, arXiv preprint [arXiv:2412.19571](https://arxiv.org/abs/2412.19571).
- [11] Stamatopoulos M-N, Banerjee A, Nikolakopoulos G. Collaborative aerial 3D printing: Leveraging UAV flexibility and mesh decomposition for aerial swarm-based construction. In: *2024 international conference on unmanned aircraft systems. ICUAS*, 2024, p. 45–52. <http://dx.doi.org/10.1109/ICUAS60882.2024.10557090>.
- [12] Rosinol A, Violette A, Abate M, Hughes N, Chang Y, Shi J, Gupta A, Carlone L. Kimera: From SLAM to spatial perception with 3D dynamic scene graphs. *Int J Robot Res* 2021;40(12–14):1510–46. <http://dx.doi.org/10.1177/02783649211056674>.
- [13] Hughes N, Chang Y, Carlone L. Hydra: A real-time spatial perception system for 3D scene graph construction and optimization. *Robot Sci Syst XVIII* 2022. <http://dx.doi.org/10.15607/RSS.2022.XVIII.050>.
- [14] Hughes N, Chang Y, Hu S, Talak R, Abdulhai R, Strader J, Carlone L. Foundations of spatial perception for robotics: Hierarchical representations and real-time systems. *Int J Robot Res* 2024;43(10):1457–505. <http://dx.doi.org/10.1177/02783649241229725>.
- [15] Bayle H, Sanchez-Lopez JL, Shaheer M, Civera J, Voos H. Situational graphs for robot navigation in structured indoor environments. *IEEE Robot Autom Lett* 2022;7(4):9107–14. <http://dx.doi.org/10.1109/LRA.2022.3189785>.
- [16] Bayle H, Sanchez-Lopez JL, Shaheer M, Civera J, Voos H. S-graphs+: Real-time localization and mapping leveraging hierarchical representations. *IEEE Robot Autom Lett* 2023;8(8):4927–34. <http://dx.doi.org/10.1109/LRA.2023.3290512>.
- [17] Bae J, Shin D, Ko K, Lee J, Kim U-H. A survey on 3D scene graphs: Definition, generation and application. In: Jo J, Choi H-L, Helbig M, Oh H, Hwangbo J, Lee C-H, Stantic B, editors. *Lecture notes in networks and systems*. vol. 642, 2023, p. 136–47. http://dx.doi.org/10.1007/978-3-031-26889-2_13.
- [18] Armeni I, Sener O, Zamir AR, Jiang H, Brilakis I, Fischer M, Savarese S. 3D semantic parsing of large-scale indoor spaces. In: *2016 IEEE conference on computer vision and pattern recognition. CVPR*, 2016, p. 1534–43. <http://dx.doi.org/10.1109/CVPR.2016.170>.
- [19] Greve E, Büchner M, Vödisch N, Burgard W, Valada A. Collaborative dynamic 3D scene graphs for automated driving. In: *2024 IEEE international conference on robotics and automation. ICRA*, 2024, p. 11118–24. <http://dx.doi.org/10.1109/ICRA57147.2024.10610112>.
- [20] Deng Y, Wang J, Zhao J, Tian X, Chen G, Yang Y, Yue Y. Opengraph: Open-vocabulary hierarchical 3D graph representation in large-scale outdoor environments. *IEEE Robot Autom Lett* 2024;9(10):8402–9. <http://dx.doi.org/10.1109/LRA.2024.3445607>.
- [21] Strader J, Hughes N, Chen W, Speranzon A, Carlone L. Indoor and outdoor 3D scene graph generation via language-enabled spatial ontologies. *IEEE Robot Autom Lett* 2024;9(6):4886–93. <http://dx.doi.org/10.1109/LRA.2024.3384084>.

- [22] Maggio D, Chang Y, Hughes N, Trang M, Griffith D, Dougherty C, Cristofalo E, Schmid L, Carlone L. Clio: Real-time task-driven open-set 3D scene graphs. *IEEE Robot Autom Lett* 2024;9(10):8921–8. <http://dx.doi.org/10.1109/LRA.2024.3451395>.
- [23] Ramakrishnan SK, Gokaslan A, Wijmans E, Maksymets O, Clegg A, Turner JM, Undersander E, Galuba W, Westbury A, Chang AX, Savva M, Zhao Y, Batra D. Habitat-matterport 3D dataset (HM3D): 1000 large-scale 3D environments for embodied AI. In: *Thirty-fifth conference on neural information processing systems datasets and benchmarks track*. 2021.
- [24] Orekhov VL, Chung TH. The DARPA subterranean challenge: A synopsis of the circuits stage. *Field Robot* 2022;2:735–47. <http://dx.doi.org/10.55417/fr.2022024>.
- [25] Saucedo MA, Stathouloupoulos N, Sumathy V, Kanellakis C, Nikolakopoulos G. BOX3D: Lightweight camera-lidar fusion for 3D object detection and localization. In: *2024 32nd mediterranean conference on control and automation. MED, 2024*, p. 101–6. <http://dx.doi.org/10.1109/MED61351.2024.10566236>.