

S-Graphs+: Real-time Localization and Mapping leveraging Hierarchical Representations

Hriday Bavle¹, Jose Luis Sanchez-Lopez¹, Muhammad Shaheer¹,
Javier Civera² and Holger Voos¹

Abstract—In this paper, we present an evolved version of the Situational Graphs, which jointly models in a single optimizable factor graph, a SLAM graph, as a set of robot keyframes, containing its associated measurements and robot poses, and a 3D scene graph, as a high-level representation of the environment that encodes its different geometric elements with semantic attributes and the relational information between those elements.

Our proposed *S-Graphs+* is a novel four-layered factor graph that includes: (1) a keyframes layer with robot pose estimates, (2) a walls layer representing wall surfaces, (3) a rooms layer encompassing sets of wall planes, and (4) a floors layer gathering the rooms within a given floor level. The above graph is optimized in real-time to obtain a robust and accurate estimate of the robot's pose and its map, simultaneously constructing and leveraging the high-level information of the environment. To extract such high-level information, we present novel room and floor segmentation algorithms utilizing the mapped wall planes and free-space clusters.

We tested *S-Graphs+* on multiple datasets including, simulations of distinct indoor environments, on real datasets captured over several construction sites and office environments, and on a real public dataset of indoor office environments. *S-Graphs+* outperforms relevant baselines in the majority of the datasets while extending the robot situational awareness by a four-layered scene model. Moreover, we make the algorithm available as a docker file.

Project web: https://snt-arg.github.io/s_graphs_docker/

I. INTRODUCTION

ROBOTS require a deep understanding of the situation for their autonomous and intelligent operations. Works like [1], [2], [3], [4] generate 3D scene graphs modeling the environment with high-level semantic abstractions (such as chairs, tables, or walls) and their relationships (such as a set of walls forming a room or a corridor). While providing a rich understanding of the scene, they rely on separate SLAM methods, such as [5], [6], [7], that previously estimate the robot's pose and its map using metric/semantic representations

*This work was partially funded by the Fonds National de la Recherche de Luxembourg (FNR), under the projects C19/IS/13713801/5G-Sky, by a partnership between the Interdisciplinary Center for Security Reliability and Trust (SnT) of the University of Luxembourg and Stugalux Construction S.A., by the Spanish Government under Grant PID2021-127685NB-I00 and by the Aragón Government under Grant DGA T45 17R/FSE. For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

¹Authors are with the Automation and Robotics Research Group, Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg. Holger Voos is also associated with the Faculty of Science, Technology and Medicine, University of Luxembourg, Luxembourg. {hriday.bavle, joseluis.sanchezlopez, muhammad.shaheer, holger.voos}@uni.lu

²Author is with I3A, Universidad de Zaragoza, Spain jcivera@unizar.es

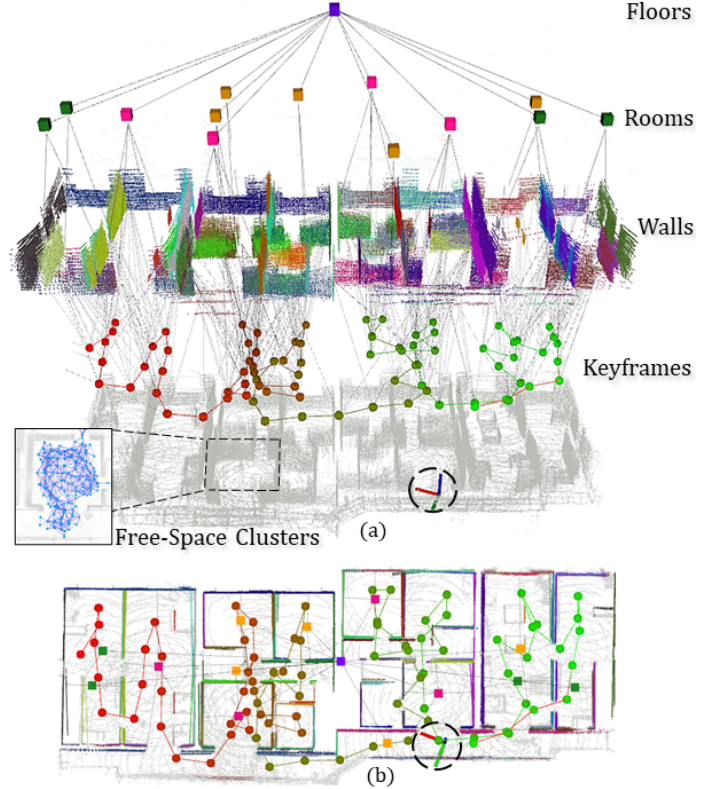


Fig. 1: *S-Graph+* built using a legged robot (circled in black) as it navigates a real construction site consisting of four adjacent houses. (a) 3D view of the four-layered hierarchical optimizable graph. The zoomed-in image shows a partial view of the free-space clusters utilized for room segmentation. (b) Top view of the graph.

without exploiting this hierarchical high-level information of the environment. Thus in general, there exists a loose coupling between 3D scene graphs and the underlying SLAM graphs.

Our previous work *S-Graphs* [8] bridges this gap proposing for the first time a tightly coupled geometric LiDAR SLAM with 3D scene graphs, demonstrating state-of-the-art metrics. However, the performance of *S-Graphs* is limited in complex environments, due to its loosely-coupled factors of walls and rooms, the influence of its multiple hand-tuned parameters for room extraction, and the lack of higher-level abstractions other than rooms.

To overcome these limitations, we present *S-Graphs+*, with updated front-end and back-ends relying on 3D LiDAR measurements. The new **front-end** includes (1) a novel *room*

segmentation algorithm using free-space clusters and wall planes. We define two- and four-walled rooms, corresponding to corridors and squarish rooms respectively. (2) a *floor segmentation* algorithm extracting the floor centers using all the currently extracted wall planes.

Its **back-end** consists of a real-time optimizable factor graph composed of four layers. A *keyframes layer* constraining a subset of robot poses at specific distance-time intervals. A *walls layer* constraining of the wall plane parameters and linked to the keyframes using pose-plane constraints. A *rooms layer* modeling detected rooms to its corresponding wall planes. The highest-level layer of the graph, the *floors layer*, denotes the current floor level in the graph and constrains the rooms at that level. See Fig. 1 for an illustrative example of an *S-Graph+* of a real building.

Our main contributions are, therefore, summarized as:

- A novel real-time factor graph organized in four hierarchical layers.
- A real-time extraction of high-level information using the novel room and floor segmentation algorithms.
- A thorough experimental evaluation in different simulated and real construction/office environments as well as software release for the research community.

II. RELATED WORKS

A. SLAM and Scene Graphs

The literature on LiDAR SLAM is huge, and there are several well known geometric approaches like LOAM [5] and its variants [6], [9], [10], and also semantic ones like LeGO-LOAM [7], SegMap [11], SUMA++ [12] that provide robust and accurate localization and 3D maps of the environments. However, geometric SLAM lacks meaning in the representation of the environments, which causes failures in aliased environments and limitations for high-level tasks or human-robot interaction. And its semantic SLAM counterpart lacks in most occasions geometric accuracy and robustness, due among other to wrong matches between the semantic elements and the absence of relational constraints between them.

Scene graphs, on the other hand, model scenes as structured representations, specifically in the form of a graph comprising objects, their attributes, and the inter-relationships among them. This high-level representation has the potential to boost several aspects in mapping, as for example the map compacity or the robot understanding. Focusing on 3D scene graphs for understanding, the pioneering work [1] creates an offline semi-autonomous framework using object detections from RGB images, generating a multi-layered hierarchical representation of the environment and its components, divided mainly into layers of camera, objects, rooms, and building. [13] presents a framework for generating a 3D scene graph using a sequence of images to verify its applicability to visual questioning and answering and to task planning. 3D SSG (Semantic Scene Graph) presents a learning method based on PointNet and Graph Convolutions Networks (GCN) to semi-automatically generate graphs for 3D scenes, while SceneGraphFusion [3] on the other hand, generate a real-time incremental 3D scene graph using RGB-D sequences, accurately handling partial and

missed semantic data. 3D DSG (Dynamic Scene Graph) [14] extend the 3D scene graph concept to environments with static parts and dynamic agents in an offline manner, while Hydra [4], presents research in the direction of real-time 3D scene graph generation as well as its optimization using loop closure constraints. Approaches have also been presented such as [15] exploiting scene graphs to efficiently generate a 3D scene. Though promising in terms of scene representation and higher-level understanding, a major drawback of these models is that they do not tightly couple the estimate of the scene graph with the SLAM state, in order to simultaneously optimize them. They thus generate a scene graph and a SLAM graph in an independent manner. Our previous work *S-Graphs* [8] bridged this gap showcasing the potential of tightly coupling SLAM graphs and scene graphs. However, for several reasons, it was limited to simple structured environments. Our current work *S-Graphs+* overcomes these limitations generating a four-layered hierarchical optimizable graph while simultaneously representing the environment as a 3D scene graph, able to provide an excellent performance even in complex environments.

B. Room Segmentation

For a robot to understand structured indoor environments, it is necessary to first understand their basic components, such as walls, and their composition into higher-level structures such as rooms. Hence, room identification and segmentation is one of the critical tasks in *S-Graphs+*. In the literature, different room segmentation techniques are presented over pre-generated maps using 2D LiDARs [16]–[18]. Their performance is, however, degraded in presence of clutter. While [19] presents a room segmentation approach based on pre-generated 2D occupancy maps in cluttered indoor environments, it still lacks real-time capabilities. Methods such as [20]–[22] perform segmentation of indoor spaces into meaningful rooms, although they require a pre-generated 3D map of the environment and cannot segment it in real-time. Given the current state-of-the-art for room segmentation, there was a need to develop a room segmentation algorithm capable of running in real-time as the robot explores its environment, to simultaneously incorporate this high-level information into the optimizable *S-Graphs+*.

III. OVERVIEW

The architecture of *S-Graphs+* is illustrated in Fig. 2. Its pipeline can be divided into six modules, and its estimates are referred to four frames: the LiDAR frame L_t , the robot frame R_t , the odometry frame O , and the map frame M . L_t and R_t are rigidly attached to the robot and then depend on the time instant t , while O and M are fixed. The first module receives the 3D LiDAR point cloud in frame L_t , which is pre-filtered and downsampled. The second module estimates the robot odometry in frame O either from LiDAR measurements or the robot encoders. Four additional front-end modules generate the four-layered topological graph modelling the understanding of the environment, namely: 1) The plane segmentation module, segmenting and initializing wall planes

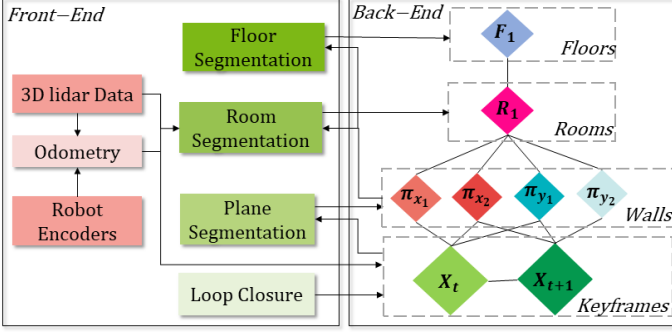


Fig. 2: *S-Graphs+* overview. Our inputs are the 3D LiDAR measurements and robot odometry, which are pre-filtered and processed in the front-end to extract wall planes, rooms, floor, and loop closures. Note the four-layered *S-Graph+*, whose parameters are jointly optimized in the back-end.

in the map frame M using the point clouds at each keyframe. 2) The room segmentation module, generating first free-space clusters from the robot poses and 3D LiDAR measurements, and then using such clusters along with the mapped planes to detect room centers in frame M . 3) The floor segmentation module, utilizes the information of all the currently mapped walls to extract the center of the current floor level in frame M . 4) Finally, the loop closure module, which utilizes a scan-matching algorithm to recognize revisited places and correct the drift. We define the global state as:

$$\mathbf{s} = [\begin{matrix} {}^M\mathbf{x}_{R_1}, \dots, {}^M\mathbf{x}_{R_T}, {}^M\boldsymbol{\pi}_1, \dots, {}^M\boldsymbol{\pi}_P, \\ {}^M\boldsymbol{\rho}_1, \dots, {}^M\boldsymbol{\rho}_S, {}^M\boldsymbol{\kappa}_1, \dots, {}^M\boldsymbol{\kappa}_K, \\ {}^M\boldsymbol{\xi}_1, \dots, {}^M\boldsymbol{\xi}_F, {}^M\mathbf{x}_O \end{matrix}]^T, \quad (1)$$

where ${}^M\mathbf{x}_{R_t} \in SE(3)$, $t \in \{1, \dots, T\}$ are the robot poses at T selected keyframes, ${}^M\boldsymbol{\pi}_i \in \mathbb{P}^3$, $i \in \{1, \dots, P\}$ are the plane parameters of the P wall planes in the scene, ${}^M\boldsymbol{\rho}_j \in \mathbb{R}^2$, $j \in \{1, \dots, S\}$ contains the parameters of the S four-wall rooms and ${}^M\boldsymbol{\kappa}_k \in \mathbb{R}^2$, $k \in \{1, \dots, K\}$ the parameters of the K two-wall rooms, ${}^M\boldsymbol{\xi}_f \in \mathbb{R}^2$, $f \in \{1, \dots, F\}$ are the F floors levels, and ${}^M\mathbf{x}_O$ models the drift between the odometry frame O and the map frame M .

IV. FRONT-END

A. Robot Odometry

S-Graphs+ is agnostic to the source of odometry, thus it can utilize odometry from different sources. It can utilize odometry estimated either from 3D LiDAR measurements or directly generated from encoders of the robotic platforms.

B. Wall Extraction

We use sequential RANSAC to detect and initialize wall planes. Compared to our previous work [8], in which wall detection was done on a different thread leading to missed planar detections, in *S-Graphs+* we extract the wall planes every time a new keyframe is registered. This results in efficient detection and mapping of all the wall planes at a

given time instant. Each wall plane extracted at time t , ${}^{L_t}\boldsymbol{\pi}$, is referred to the LiDAR frame L_t , we need to convert it to its Closest Point (CP) representation [8], and then to the map frame ${}^M\boldsymbol{\pi}$ using the estimated robot pose at time t . The wall plane normals with their Mn_x or Mn_y components greater than the Mn_z component are classified as vertical planes. Furthermore, normals where Mn_x is greater than Mn_y are classified as x -plane normals, and otherwise they are classified as y -plane normals. Finally, planes whose normals' bigger component is Mn_z are classified as horizontal planes or ground surfaces. After initializing each plane in the global map, correspondences are searched for every subsequent plane observation. Data association is performed using the Mahalanobis distance between each mapped plane and the newly extracted ones.

C. Room Segmentation

In this work, we present a novel room segmentation strategy capable of segmenting different room configurations in a structured indoor environment. It consists on two steps, **Free-Space Clustering** and **Room Extraction**, and the output are the parameters of **four-wall** and **two-wall** rooms. All this is explained in the next paragraphs.

Free-Space Clustering. Our free-space clustering algorithm divides the Free-Space Graph of a scene into several clusters that should correspond to the rooms of that scene. We first generate a complete sparse connected graph of free spaces using [23], given the robot poses and a Euclidean Signed Distance Field (ESDF) representation [24]. We then cluster this free-space graph into different free-space regions as follows. Given the graph \mathcal{G} , we create a filtered graph \mathcal{G}_f removing the vertices v_d whose distance to obstacles is greater than a given threshold t_λ . We also remove from \mathcal{G}_f all the edges e_d that are connected to the node set v_d . We then run the connected components method [25] on \mathcal{G}_f to divide it into several connected sub-graphs \mathcal{G}_{f_i} , $i \in \{1, \dots, N\}$. In order to re-connect the deleted vertices v_d and their edges e_d to the filtered sub-graphs \mathcal{G}_{f_i} , we check within the entire graph \mathcal{G} , each edge e_{d_i} that connects vertex v_{f_i} of a filtered sub-graph \mathcal{G}_{f_i} to the deleted vertex v_{d_i} , thus inserting vertex v_{d_i} within \mathcal{G}_{f_i} . Using this technique we can obtain disconnected free-space clusters belonging to different rooms as vertices close to room openings have distances closer to walls (obstacles) and thus vote for disconnecting the graph. Algorithm. 1 and Fig. 3 give further details on this free-space clustering.

Room Extraction. Room extraction uses the free-space clusters \mathcal{G}_{f_i} and the wall planes from a keyframe at time t to detect different room configurations. Wall planes are represented in the map frame as ${}^M\boldsymbol{\Pi} = [{}^M\boldsymbol{\pi}_1, \dots, {}^M\boldsymbol{\pi}_j]$, where each plane ${}^M\boldsymbol{\pi}_i = [{}^M\mathbf{n}, {}^Md]$ is defined by its normal ${}^M\mathbf{n} = [{}^Mn_x, {}^Mn_y, {}^Mn_z]$ and distance Md to the origin. All extracted wall planes are first categorized as x -direction planes ${}^M\boldsymbol{\Pi}_x$, for which their highest normal component is n_x , and y -direction planes ${}^M\boldsymbol{\Pi}_y$ for which the highest normal dimension is n_y . ${}^M\boldsymbol{\Pi}_x$ plane are further classified as ${}^M\boldsymbol{\Pi}_{x_a}$, with $n_x > 0$, and ${}^M\boldsymbol{\Pi}_{x_b}$ with $n_x < 0$. Analogously ${}^M\boldsymbol{\Pi}_{y_a}$ and ${}^M\boldsymbol{\Pi}_{y_b}$ represent y -planes with positive and negative n_y respectively.

Algorithm 1: Free-Space Clustering

Input: Free-space graph \mathcal{G} , generated using [23]
Output: Clustered sub-graphs $\mathcal{G}_{f_i}, i \in \{1, \dots, N\}$

1: Filter nodes far from obstacles: $\mathcal{G}_f \leftarrow \mathcal{G}$

```

for  $v_i \in \mathcal{V}$  &  $v \in \mathcal{G}$  do
  if  $v_i.distance > t_\lambda$  then
     $v_d \leftarrow v_i$  &  $e_d \leftarrow v_i.edges$ 
  else
     $v_f \leftarrow v_i$  &  $e_f \leftarrow v_i.edges$ 
  end
end
 $\mathcal{G}_f \leftarrow (v_f, e_f)$ 

```

2: Graph clustering by connectivity: $\mathcal{G}_{f_1} \dots \mathcal{G}_{f_n} \leftarrow \mathcal{G}_f$

```

 $c \leftarrow 0$ 
for  $v_{f_i} \in \mathcal{V}_f$  do
   $v_{f_i}.visited \leftarrow False$  &  $v_{f_i}.cluster \leftarrow 0$ 
end
Function  $visit(v_f)$ :
  if  $v_{f_i}.visited = False$  then
     $v_{f_i}.visited \leftarrow True$ 
    if  $v_{f_i}.cluster = 0$  then
       $c \leftarrow c + 1$ 
       $v_{f_i}.cluster \leftarrow c$  &  $\mathcal{G}_c \leftarrow v_{f_i}$ 
    end
    for  $v_{f_n} \in v_{f_i}.neighbours$  do
       $v_{f_n}.cluster \leftarrow c$  &  $visit(v_{f_n})$ 
    end
  end
return
for  $v_{f_i} \in \mathcal{V}_f$  do
   $visit(v_{f_i})$ 
end

```

3: Inclusion of deleted vertices and edges in the cluster: $\mathcal{G}_{f_i} \leftarrow (v_d, e_d)$

```

for  $v_{d_i} \in \mathcal{V}_d$  do
  for  $v_n \in v_{d_i}.neighbours$  do
    if  $v_n \in \mathcal{G}_{f_i}$  then
       $c \leftarrow v_n.cluster$  &  $\mathcal{G}_{f_i} \leftarrow (v_{d_i}, e_{d_i})$ 
    end
  end
end

```

Given each sub-category of the wall planes, our room extraction method first checks the $L2$ norm between the 3D points of each plane and the vertices of each cluster \mathcal{G}_{f_i} , to find the set of walls lying closer to each specific cluster.

Four-Wall Rooms. For a given cluster \mathcal{G}_{f_i} , if the room extraction module finds a set of four wall planes ${}^M\Pi_s = [{}^M\pi_{x_{a_1}}, {}^M\pi_{x_{b_1}}, {}^M\pi_{y_{a_1}}, {}^M\pi_{y_{b_1}}]$ close to the cluster vertices, it is considered as a four-wall room candidate and further tests are carried out. First, the widths w_x and w_y of ${}^M\Pi_x =$

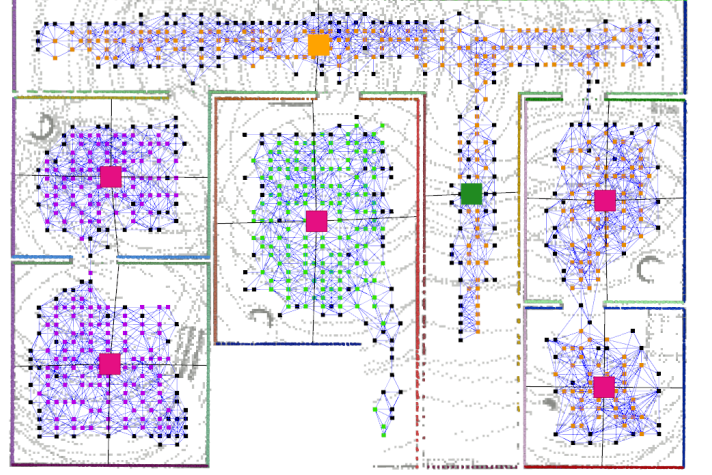


Fig. 3: Free space clustering and rooms segmentation, obtained from the estimated wall planes surrounding each cluster. Pink colored squares represent a four-wall room, while yellow and green colored squares represent two-wall rooms in x and y directions respectively. Nodes colored in black are those that are closest to walls and vote for splitting the graph.

$\{{}^M\pi_{x_{a_1}}, {}^M\pi_{x_{b_1}}\}$ and ${}^M\Pi_y = \{{}^M\pi_{y_{a_1}}, {}^M\pi_{y_{b_1}}\}$ should be greater than a given threshold t_w , where

$$w_x = [{}^M d_{x_{a_1}} \cdot {}^M \mathbf{n}_{x_{a_1}} - |{}^M d_{x_{b_1}}| \cdot {}^M \mathbf{n}_{x_{b_1}}]$$

$$w_y = [{}^M d_{y_{a_1}} \cdot {}^M \mathbf{n}_{y_{a_1}} - |{}^M d_{y_{b_1}}| \cdot {}^M \mathbf{n}_{y_{b_1}}] \quad (2)$$

${}^M d_{x_{a_1}}$ and ${}^M d_{x_{b_1}}$ are the plane distances to the origin and ${}^M \mathbf{n}_{x_{a_1}}$ and ${}^M \mathbf{n}_{x_{b_1}}$ are the normals of x -planes. Similarly, ${}^M d_{y_{a_1}}$, ${}^M d_{y_{b_1}}$, ${}^M \mathbf{n}_{y_{a_1}}$ and ${}^M \mathbf{n}_{y_{b_1}}$ are the distances and normals for in y -planes. For Eq. 2 to hold true, $|{}^M d_{x_{a_1}}| > |{}^M d_{x_{b_1}}|$ and $|{}^M d_{y_{a_1}}| > |{}^M d_{y_{b_1}}|$. All plane normals are converted to point away from the map M frame as:

$${}^M \mathbf{n} = \begin{cases} -1 \cdot {}^M \mathbf{n} & \text{if } {}^M d > 0 \\ {}^M \mathbf{n} & \text{otherwise} \end{cases} \quad (3)$$

If the above test is successful, the 3D points in each wall are checked to be enclosed within the two apposed walls. For example, in-plane points belonging to ${}^M\pi_{x_{a_1}}$ are checked to lie within the points of ${}^M\pi_{y_{a_1}}$ and ${}^M\pi_{y_{b_1}}$. Given a room candidate with a planar set ${}^M\Pi_s$ consisting of four walls, we first calculate the room center as follows:

$${}^M \mathbf{r}_{x_i} = \frac{1}{2} [{}^M d_{x_{a_1}} \cdot {}^M \mathbf{n}_{x_{a_1}} - |{}^M d_{x_{b_1}}| \cdot {}^M \mathbf{n}_{x_{b_1}}] + |{}^M d_{x_{b_1}}| \cdot {}^M \mathbf{n}_{x_{b_1}}$$

$${}^M \mathbf{r}_{y_i} = \frac{1}{2} [{}^M d_{y_{a_1}} \cdot {}^M \mathbf{n}_{y_{a_1}} - |{}^M d_{y_{b_1}}| \cdot {}^M \mathbf{n}_{y_{b_1}}] + |{}^M d_{y_{b_1}}| \cdot {}^M \mathbf{n}_{y_{b_1}}$$

$${}^M \rho_i = {}^M \mathbf{r}_{x_i} + {}^M \mathbf{r}_{y_i} \quad (4)$$

Eq. 4 holds true when $|d_{x_1}| > |d_{x_2}|$. Again, all planes are converted to point away from the origin M using Eq. 3.

Data association for the room node follows two steps. First, the $L2$ norm between the positions of the mapped rooms with the newly detected ones is calculated. Second, the shortlisted rooms using the first step undergo further checks at each wall

plane. The *id* of each wall plane comprised by the newly detected room is checked with the wall planes of the mapped room. In case of *id* mismatch, Mahalanobis distance between wall planes constituting the wall planes is checked. If all wall planes are matched a detected room is matched. This process allows for the identification and merging of duplicate wall planes (*id* mismatch) for a given room, arising from inaccuracies in the wall plane matching step (Section. IV-B) as the room and its respective wall plane matching thresholds can be safely tuned to be larger than the single wall plane matching threshold.

Two-Wall Rooms. The room extraction method is sometimes able to find only two walls that surround a free-space cluster \mathcal{G}_{f_i} . These two-wall rooms can be rooms with some undetected walls or corridor-like structures. If a wall plane set ${}^M\Pi_s = [{}^M\pi_{x_{a_1}}, {}^M\pi_{x_{b_1}}]$ contains two x -planes then it is a two-wall room in x direction. Analogously, two-wall rooms in the y direction are composed of opposed y -planes. Walls forming two-wall rooms undergo the same checks as four-wall rooms, shown in equations 3 and 2. Given the missing information, in addition to the wall planes, the corresponding center ${}^M\mathbf{c}_i$ of the cluster \mathcal{G}_{f_i} is utilized to compute the two-wall room center as follows:

$$\begin{aligned} {}^M\mathbf{r}_{x_i} &= \frac{1}{2} [|{}^Md_{x_{a_1}}| \cdot {}^M\mathbf{n}_{x_{a_1}} - |{}^Md_{x_{b_1}}| \cdot {}^M\mathbf{n}_{x_{b_1}}] + |{}^Md_{x_{b_1}}| \cdot {}^M\mathbf{n}_{x_{b_1}} \\ {}^M\mathbf{r}_{x_i} &= {}^M\mathbf{r}_{x_i} + [{}^M\mathbf{c}_i - [{}^M\mathbf{c}_i \cdot {}^M\hat{\mathbf{r}}_{x_i}] \cdot {}^M\hat{\mathbf{r}}_{x_i}] \end{aligned} \quad (5)$$

where ${}^M\mathbf{r}_{x_i}$ is the two-wall room center in x direction, ${}^M\hat{\mathbf{r}}_{x_i} = {}^M\mathbf{r}_{x_i} / \|{}^M\mathbf{r}_{x_i}\|$, and ${}^M\mathbf{c}_i$ is the cluster center obtained from the endpoints of the cluster \mathcal{G}_{f_i} as:

$$\begin{aligned} {}^M\mathbf{c}_{x_i} &= \frac{1}{2} [{}^Mp_{x_1} - {}^Mp_{x_2}] + {}^Mp_{x_2} \\ {}^M\mathbf{c}_{y_i} &= \frac{1}{2} [{}^Mp_{y_1} - {}^Mp_{y_2}] + {}^Mp_{y_2} \\ {}^M\mathbf{c}_i &= [{}^M\mathbf{c}_{x_i}, {}^M\mathbf{c}_{y_i}] \end{aligned} \quad (6)$$

where ${}^Mp_{x_1}$ to ${}^Mp_{y_2}$ are the cluster endpoints. Two-wall room center in y direction can be calculated analogously.

Data association of two-wall rooms follows a similar concept as four-wall rooms. In the case of a two-wall room in the x direction, we first compute the $L2$ norm along the x -axis of the two-wall room center followed by the *id* check of individual wall planes comprising the two-wall room. In case of *id* mismatch, $L2$ norm of the planar points between the detected and the mapped wall planes is computed. A new two-wall room is created in case either of the matching distances is greater than the threshold and as in the case of four-wall rooms, this procedure allows for the identification of duplication wall planes. Both the detected four and two-wall rooms are optimized along with their corresponding wall planes in the back-end explained in Section. V.

D. Floor Segmentation

The floor segmentation module extracts the widest wall planes within the current explored floor level by the robot which can then be used to calculate the center of the current

floor level. Our floor segmentation utilizes the information from all mapped walls to create a sub-category of wall planes as described in the room segmentation (Section. IV-C) as, ${}^M\Pi_{s_t}$ where $t = \{1, \dots, T\}$. After receiving a complete plane set it computes the widths w_x between all x -direction planes and similarly w_y for y -direction planes using Eq. 2. The wall plane set with the largest w_x and w_y is the chosen candidate for the current floor level. These planar pairs in both x and y direction undergo an additional dot product check between their corresponding normal orientations, $|\mathbf{n}_{x_{a_1}} \cdot \mathbf{n}_{x_{b_1}}| < t_n$ and $|\mathbf{n}_{y_{a_1}} \cdot \mathbf{n}_{y_{b_1}}| < t_n$, to remove wall planes originating outside the building structure. The floor segmentation computes the floor center node using the obtained wall plane candidates following Eq. 4. Whenever the robot ascends or descends to a different floor level, the newly mapped wall planes are incorporated with the new floor, and the current floor level is computed only using the wall planes with the same floor.

E. Loop Closure

As in [8], room detections provide “soft” loop closure constraints when the robot revisits previously mapped rooms, while a scan matching-based “hard” loop closure method add constraints on the relative pose of neighboring keyframes.

V. BACK-END

The back-end is responsible for creating and optimizing the four layered *S-Graph+*, which is explained in detail as follows.

Keyframes. This layer creates a factor node ${}^M\mathbf{x}_{R_t} \in SE(3)$ with the robot keyframe pose at time t in the map frame M . The pose nodes are constrained by pairwise odometry readings between consecutive poses as in [8].

Walls. This layer creates the planar factor nodes for the wall planes extracted by the wall segmentation (Section. IV-B). The planar nodes are factored as ${}^M\pi = [{}^M\phi, {}^M\theta, {}^Md]$, where ${}^M\phi$ and ${}^M\theta$ stand for the azimuth and elevation of the plane in frame M . The planar nodes are constrained with their corresponding keyframes using pose-plane constraints as in [8]. The room segmentation module utilizes mapped walls at current keyframe k_t (Section. IV-C) to identify different room candidates, whereas mapped walls from all the mapped keyframes $\mathbf{k} = \{k_1, \dots, k_T\}$ are utilized by the floor segmentation module (Section. IV-D) to identify the center of the floor level.

Rooms. The rooms layer receives the extracted room candidates and their corresponding wall planes from the room segmentation module (Section IV-C) to create appropriate constraints between them.

Four-Wall Rooms: We propose a novel edge formulation between the detected room node (generated from its center) and its four mapped wall planes, where the total cost function to minimize the room node and its plane set can be given as:

$$\begin{aligned} &c_{\rho}({}^M\rho, [{}^M\pi_{x_{a_i}}, {}^M\pi_{x_{b_i}}, {}^M\pi_{y_{a_i}}, {}^M\pi_{y_{b_i}}]) \\ &= \sum_{t=1, i=1}^{T, S} \|{}^M\hat{\rho}_i - f({}^M\tilde{\pi}_{x_{a_i}}, {}^M\tilde{\pi}_{x_{b_i}}, {}^M\tilde{\pi}_{y_{a_i}}, {}^M\tilde{\pi}_{y_{b_i}})\|_{\tilde{\Lambda}_{\hat{\rho}_i, t}}^2 \end{aligned} \quad (7)$$

Where ${}^M\hat{\rho}_i$ is the estimated four-wall room center obtained from Section. IV-C and $f({}^M\tilde{\pi}_{x_{a_i}}, {}^M\tilde{\pi}_{x_{b_i}}, {}^M\tilde{\pi}_{y_{a_i}}, {}^M\tilde{\pi}_{y_{b_i}})$ is the function mapping the four wall planes estimated to a four-wall room center using Eq. 4. The goal of this cost function is to maintain the structural consistency between the four planes forming the room.

Two-Wall Rooms: We propose a similar cost function to minimize room nodes and their two corresponding wall planes as follows:

$$c_{\kappa}({}^M\kappa_i, [{}^M\pi_{x_{a_1}}, {}^M\pi_{x_{b_1}}, {}^M\mathbf{c}_i]) = \sum_{t=1, i=1}^{T, K} \| {}^M\hat{\kappa}_i - f({}^M\tilde{\pi}_{x_{a_1}}, {}^M\tilde{\pi}_{x_{b_1}}, {}^M\mathbf{c}_i) \|_{\Lambda_{\tilde{\kappa}_i, t}}^2 \quad (8)$$

${}^M\mathbf{c}_i$ is the cluster center, which is kept constant during the optimization, and ${}^M\hat{\kappa}_i$ is the estimated two-wall room center in x direction obtained from Section. IV-C. $f({}^M\tilde{\pi}_{x_{a_1}}, {}^M\tilde{\pi}_{x_{b_1}}, {}^M\mathbf{c}_i)$ maps the two wall planes and its cluster center to a room center using Eq. 5. The cost function to minimize two-wall rooms in y direction follows Eq. 8 for wall planes $({}^M\pi_{y_{a_j}}, {}^M\pi_{y_{b_j}})$ and cluster center ${}^M\mathbf{c}_j$. Duplicate wall plane nodes identified during the four-wall or two-wall room segmentation are constrained by a factor minimizing the difference between their respective parameters.

Floors. The floor node consists of the center of the current floor level calculated from the floor segmentation (Section. IV-D). We add a cost function between the floor node and all the mapped four-wall rooms at that floor level as follows:

$$c_{\xi}({}^M\xi_i, {}^M\rho_j) = \sum_{t=1, i=1, j=1}^{T, F, S} \| {}^M\hat{\delta}_{\xi_i, \rho_j} - f({}^M\xi_i, {}^M\rho_j) \|_{\Lambda_{\xi_i, t}}^2 \quad (9)$$

where ${}^M\hat{\delta}_{\xi_i, \rho_j}$ stands for the relative distance between the floor i with center ξ_i and the four-wall room j with center ρ_j , and $f({}^M\xi_i, {}^M\rho_j)$ maps the relative distance between the centers of floor node and four-wall room node. Two-wall room nodes are constrained with the floor node using the same Eq. 9. While the robot navigates in the surroundings and discovers new wall planes, the estimate of the floor node might change due to the insertion of such planes into the map. If the current floor center calculated from the new wall planes gets updated beyond a threshold t_f , the estimate of the floor node is updated in the graph accordingly along with the relative distances between the floors and all the rooms.

VI. EXPERIMENTAL RESULTS

A. Methodology

We validate *S-Graphs+* on simulated and real-world scenarios, comparing it against several state-of-the-art LiDAR SLAM frameworks and the baseline *S-Graphs*. The experiments cover a wide array of scenes, from construction sites to office spaces, and use data recorded in-house and from the public TIERS dataset. We report standard error metrics for the map and trajectory estimations (RMSE and ATE) as well as qualitative results. For comparing the room detection of *S-Graphs+*

TABLE I: Absolute Trajectory Error (ATE) [m], of *S-Graphs+* and relevant baselines on simulated data. Best results are boldfaced, second best are underlined.

Method		Dataset				
Mapping	Odometry	<i>C1F0</i>	<i>C1F2</i>	<i>SE1</i>	<i>SE2</i>	<i>SE3</i>
HDL-SLAM [10]	VGICP [26]	0.15	<u>0.02</u>	0.04	0.15	0.14
ALOAM [5]	ALOAM	0.10	0.09	0.16	0.32	0.20
MLOAM [9]	MLOAM	1.14	0.52	0.65	2.82	0.17
FLOAM [6]	FLOAM	0.11	0.15	0.15	0.24	0.76
LeGO-LOAM [7]	LeGO-LOAM	-	-	-	-	0.73
<i>S-Graphs</i> [8]	VGICP	<u>0.07</u>	<u>0.02</u>	<u>0.03</u>	0.17	0.33
<i>S-Graphs+</i> (ours)	VGICP	0.08	0.01	0.02	0.12	0.05
<i>S-Graphs+</i> (ours)	FLOAM	0.05	0.03	0.15	<u>0.14</u>	<u>0.12</u>

against the heuristics in *S-Graphs*, we report the precision and recall of the four-walled and two-walled room detections in the simulated environments, for which we have the ground truth number of rooms defined in the architectural plans.

Simulated Data. We conduct a total of five simulated experiments. Two of them, *C1F1* and *C1F2*, are generated from the 3D meshes of two floors of actual architectural plans. The other three, *SE1*, *SE2*, and *SE3*, are performed in additional simulated environments resembling typical indoor environments with different room configurations. Due to absence of odometry from robot encoders, in all simulated experiments the odometry is estimated only from LiDAR measurements. For a fair validation, *S-Graphs+* is run using two different odometry inputs, specifically VGICP [26] and FLOAM [6].

In-House Dataset. In all our in-house data we utilize the robot encoders for estimating the odometry. The first two experiments, denoted as *C1F1* and *C1F2*, are performed on two floors of a construction site consisting of a single house. Additionally, experiments are performed over larger construction sites combining several houses. *C2F0*, *C2F1*, and *C2F2* consist of three floors of an ongoing construction site combining four individual houses. *C3F1*, and *C3F2* are two combined houses, while *C4F0* is a basement area with different storage rooms. *LC1* consists of an office environment in which the robot traverses back and forth a long corridor. To validate the accuracy of each method in all the real experi-

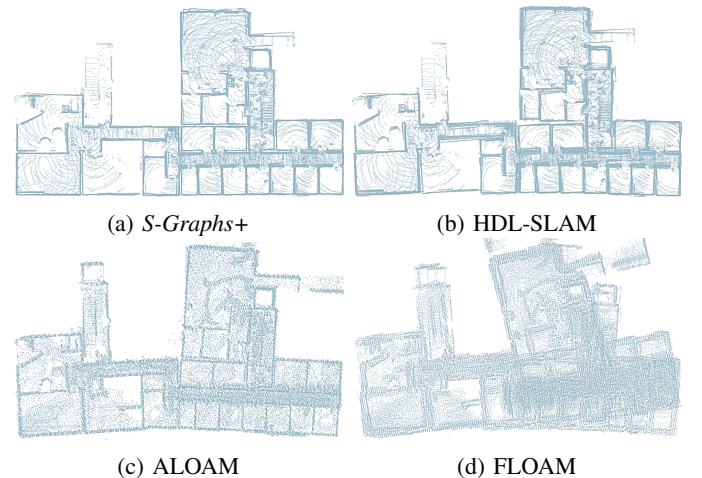


Fig. 4: Maps by *S-Graphs+* and baselines, in-house seq. *C4F0*.

TABLE II: Point cloud RMSE [m] for our in-house real sequences. All methods use odometry from robot encoders. Best results are boldfaced, second best underlined.

	Dataset							
Method	Point Cloud RMSE							
Mapping	<i>C1F1</i>	<i>C1F2</i>	<i>C2F0</i>	<i>C2F1</i>	<i>C2F2</i>	<i>C3F1</i>	<i>C3F2</i>	<i>LC1</i>
HDL-SLAM [10]	1.34	0.37	<u>0.26</u>	0.46	<u>0.26</u>	1.41	1.08	1.45
ALOAM [5]	8.03	1.20	0.31	0.43	0.57	0.57	0.58	3.14
MLOAM [9]	3.73	1.93	0.34	0.61	0.51	-	0.37	1.68
FLOAM [6]	7.63	1.15	0.46	0.74	0.35	1.66	0.41	2.90
LeGO-LOAM [7]	4.08	0.70	0.46	0.48	0.53	0.32	0.52	3.40
<i>S-Graphs</i> [8]	<u>0.33</u>	<u>0.37</u>	0.30	0.48	0.50	0.41	0.29	1.24
<i>S-Graphs+</i> (ours)	0.31	0.33	0.17	<u>0.46</u>	0.13	<u>0.39</u>	<u>0.32</u>	<u>1.27</u>

ments we report the RMSE of the estimated 3D maps against the actual 3D map generated from the architectural plan except for experiment *C4F0*, for which we provide qualitative results due to the absence of a ground truth plan.

TIERS LiDARs dataset. We also validate *S-Graphs+* on the public TIERS dataset [27], recorded by a moving platform in a variety of scenarios. We show results in the five indoor sequences of the dataset. Experiments *T6* to *T8* are done in a single small room in which the platform does several passes at increasing speeds. Experiments *T10* and *T11* are performed in a larger indoor hallway with longer trajectories of the moving platform. We use the VLP-16 LiDAR data and report the ATE against the provided ground truth. Due to the absence of encoder readings in this dataset, each baseline method uses its own LiDAR-based odometry. As in the simulated datasets, we validate *S-Graphs+* with VGICP and FLOAM odometry.

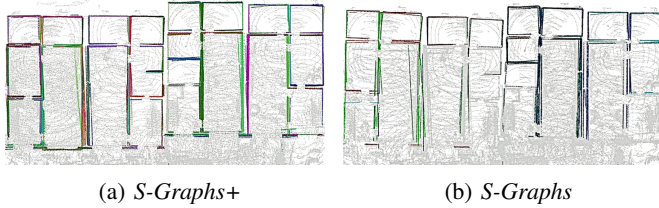


Fig. 5: *S-Graphs+* and *S-Graphs* maps, in-house seq. *C2F0*.

B. Results and Discussion

Simulated Data. Table. I showcases the ATE for the simulated experiments. We outperform *S-Graphs* [8] thanks to

TABLE III: Absolute Trajectory Error (ATE) [m], of *S-Graph+* and relevant baselines on the TIERS dataset [27]. Best results boldfaced, second best underlined.

Method	Dataset					
Mapping	Odometry	<i>T6</i>	<i>T7</i>	<i>T8</i>	<i>T10</i>	<i>T11</i>
HDL-SLAM [10]	VGICP [26]	<u>0.26</u>	0.29	0.34	1.42	2.98
ALOAM	ALOAM [5]	<u>0.26</u>	<u>0.27</u>	0.34	0.76	2.29
MLOAM	MLOAM [9]	<u>0.26</u>	0.26	0.35	3.27	0.92
FLOAM	FLOAM [6]	0.25	0.26	<u>0.33</u>	<u>0.65</u>	2.05
LeGO-LOAM	LeGO-LOAM [7]	0.27	0.33	0.38	1.49	<u>0.80</u>
<i>S-Graphs</i> [8]	VGICP	<u>0.26</u>	0.28	0.37	1.41	1.41
<i>S-Graphs+</i> (ours)	VGICP	0.25	0.27	0.34	1.27	0.95
<i>S-Graphs+</i> (ours)	FLOAM	0.25	0.26	0.32	0.54	0.55

the new plane segmentation module, new rooms factors and the new room segmentation algorithm. Experiments are sorted by scene size, from left to right the scene size being larger. Note how the baseline errors tend to grow for larger scenes, and how our *S-Graphs+* achieves bigger error reductions for larger scenes due to its better representation. It can also be seen in Table. I how *S-Graphs+* has been run using two odometry methods, VGICP and FLOAM, and how it is able to improve the results of the two odometries.

In-House Dataset. Table II presents the point cloud RMSE obtained by comparing the generated 3D maps against the 3D maps from the building plans. As it can be observed in the table, *S-Graphs+* is more accurate than the baseline in most of the cases. For experiment *C4F0*, Fig. 4 shows a top view of the final maps estimated by *S-Graphs+* and three other baselines. Observe the higher degree of accuracy and cleaner map elements in the *S-Graphs+* case, the latest indicating a better alignment for different robot passes. Similarly, observe the precise map generated by *S-Graphs+* in Fig. 5 for experiment *C2F0* when comparing with *S-Graphs*, mainly due to the improved wall plane segmentation and improved rooms-wall planes constraints. Fig. 1, shows the entire four-layered *S-Graphs+* for *C2F2* along with its map accuracy. In all the above experiments, no fine-tuning of parameters of *S-Graphs+* was required for room detection and the same prior tuned parameters sufficed for all.

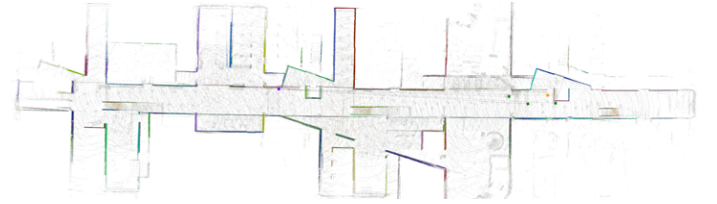


Fig. 6: Map estimated by *S-Graphs+* on TIERS sequence *T11*.

Fig. 7 presents the precision/recall of the room detection in *S-Graphs+* (the method described in Section IV-C) and the one in *S-Graphs*. Note how the precision is slightly higher for *S-Graphs+*. More importantly, the recall is clearly higher for *S-Graphs*, in particular for scenarios with complex layouts such as *C2F1*. The latest is one of the main strengths of *S-Graphs*, extracting a higher number of rooms adds a higher number of constraints leading to more accurate estimates and a better representation.

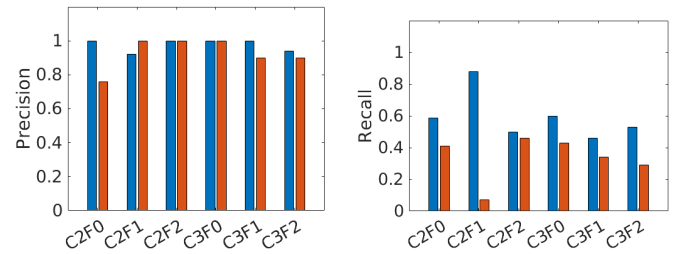


Fig. 7: Precision and recall for *S-Graphs+* (blue) and *S-Graphs* (red) on six different scenes of our in-house dataset.

TIERS LiDARs dataset. Table. III presents the ATE for all baseline methods and our *S-Graphs+* in the indoor sequences of the public TIERS dataset [27]. *S-Graphs+* with FLOAM odometry gives the best results in all the experiments. Again, the sequences are sorted from left to right by increasing size. Note that all methods perform similarly for small scenes, but differ as scenes become larger, *S-Graphs+* presenting significant error reductions for large environments. The strength of our hierarchical representation is particularly evident in scenarios like *T11*, in which *S-Graphs+* keeps the errors small even if the FLOAM odometry error grows substantially. The accuracy of our method in such sequence can be observed qualitatively in the top view of the estimated map in Fig. 6.

VII. CONCLUSION

In this work we present *S-Graphs+*, a novel four-layered hierarchical factor graph composed of A *keyframes layer* constraining a sub-set of robot poses at specific distance-time intervals. A *walls layer* constraining the wall plane parameters and linking it to the keyframes. A *rooms layer* modeling detected rooms to their corresponding wall planes and a *floors layer*, denoting the current floor level in the graph and constraining the rooms at that level. To extract this high-level information we also propose a novel room segmentation algorithm using free-space clusters and wall planes and a floor segmentation algorithm extracting the floor centers using all the currently extracted wall planes. We demonstrate state-of-the-art results against several baselines on simulated and real experiments covering different office and construction indoor environments. In future work, we plan to exploit the hierarchical structure of the graph for efficient and faster optimization as well as enhance the reasoning over the graph for improving the detection of different relationship constraints between its semantic elements.

REFERENCES

- [1] I. Armeni, Z.-Y. He, J. Gwak, A. R. Zamir, M. Fischer, J. Malik, and S. Savarese, “3D Scene Graph: A structure for unified semantics, 3D space, and camera,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 5664–5673.
- [2] A. Rosinol, A. Gupta, M. Abate, J. Shi, and L. Carlone, “3D dynamic scene graphs: Actionable spatial perception with places, objects, and humans,” *arXiv preprint arXiv:2002.06289*, 2020.
- [3] S.-C. Wu, J. Wald, K. Tateno, N. Navab, and F. Tombari, “SceneGraphFusion: Incremental 3D Scene Graph Prediction from RGB-D Sequences,” 2021. [Online]. Available: <https://arxiv.org/abs/2103.14898>
- [4] N. Hughes, Y. Chang, and L. Carlone, “Hydra: A Real-time Spatial Perception Engine for 3D Scene Graph Construction and Optimization,” *arXiv preprint arXiv:2201.13360*, 2022.
- [5] J. Zhang and S. Singh, “LOAM: Lidar Odometry and Mapping in Real-time,” in *Robotics: Science and Systems*, 2014.
- [6] H. Wang, C. Wang, C. Chen, and L. Xie, “F-LOAM: Fast LiDAR Odometry and Mapping,” in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.
- [7] T. Shan and B. Englot, “LeGO-LOAM: Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 4758–4765.
- [8] H. Bavlle, J. L. Sanchez-Lopez, M. Shaheer, J. Civera, and H. Voos, “Situational graphs for robot navigation in structured indoor environments,” *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9107–9114, 2022.
- [9] J. Jiao, H. Ye, Y. Zhu, and M. Liu, “Robust Odometry and Mapping for Multi-LiDAR Systems With Online Extrinsic Calibration,” *IEEE Transactions on Robotics*, pp. 1–10, 2021.
- [10] K. Koide, J. Miura, and E. Menegatti, “A portable three-dimensional LIDAR-based system for long-term and wide-area people behavior measurement,” *International Journal of Advanced Robotic Systems*, vol. 16, no. 2, Mar. 2019.
- [11] R. Dubé, A. Cramariuc, D. Dugas, H. Sommer, M. Dymczyk, J. Nieto, R. Siegwart, and C. Cadena, “SegMap: Segment-based mapping and localization using data-driven descriptors,” *The International Journal of Robotics Research*, vol. 39, no. 2-3, pp. 339–355, jul 2019. [Online]. Available: <https://doi.org/10.1177%2F0278364919863090>
- [12] X. Chen, A. Milioto, E. Palazzolo, P. Giguère, J. Behley, and C. Stachniss, “SuMa++: Efficient LiDAR-based Semantic SLAM,” in *Proceedings of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2019.
- [13] U.-H. Kim, J.-M. Park, T. jin Song, and J.-H. Kim, “3-D Scene Graph: A Sparse and Semantic Representation of Physical Environments for Intelligent Agents,” *IEEE Transactions on Cybernetics*, vol. 50, no. 12, pp. 4921–4933, dec 2020. [Online]. Available: <https://doi.org/10.1109%2Ftcyb.2019.2931042>
- [14] A. Rosinol, A. Violette, M. Abate, N. Hughes, Y. Chang, J. Shi, A. Gupta, and L. Carlone, “Kimera: From SLAM to spatial perception with 3D dynamic scene graphs,” *The International Journal of Robotics Research*, vol. 40, no. 12-14, pp. 1510–1546, 2021.
- [15] H. Dharmo, F. Manhardt, N. Navab, and F. Tombari, “Graph-to-3D: End-to-End Generation and Manipulation of 3D Scenes Using Scene Graphs,” 2021. [Online]. Available: <https://arxiv.org/abs/2108.08841>
- [16] R. Bormann, F. Jordan, W. Li, J. Hampp, and M. Hägele, “Room segmentation: Survey, implementation, and analysis,” *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1019–1026, 2016.
- [17] M. Mielle, M. Magnusson, and A. J. Lilienthal, “A method to segment maps from different modalities using free space layout maoris: Map of ripples segmentation,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 4993–4999.
- [18] F. Foroughi, J. Wang, A. Nemati, Z. Chen, and H. Pei, “MapSegNet: A Fully Automated Model Based on the Encoder-Decoder Architecture for Indoor Map Segmentation,” *IEEE Access*, vol. 9, pp. 101 530–101 542, 2021.
- [19] M. Luperto, T. P. Kucner, A. Tassi, M. Magnusson, and F. Amigoni, “Robust structure identification and room segmentation of cluttered indoor environments from occupancy grid maps,” 2022. [Online]. Available: <https://arxiv.org/abs/2203.03519>
- [20] I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese, “3D Semantic Parsing of Large-Scale Indoor Spaces,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1534–1543.
- [21] R. Ambruş, S. Claiici, and A. Wendt, “Automatic Room Segmentation From Unstructured 3-D Data of Indoor Environments,” *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 749–756, 2017.
- [22] S. Ochmann, R. Vock, and R. Klein, “Automatic reconstruction of fully volumetric 3D building models from oriented point clouds,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 151, pp. 251–262, 2019.
- [23] H. Oleynikova, Z. Taylor, R. Siegwart, and J. Nieto, “Sparse 3D Topological Graphs for Micro-Aerial Vehicle Planning,” 2018. [Online]. Available: <https://arxiv.org/abs/1803.04345>
- [24] H. Oleynikova, Z. Taylor, M. Fehr, R. Siegwart, and J. Nieto, “Voxblox: Incremental 3D Euclidean Signed Distance Fields for on-board MAV planning,” in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, sep 2017. [Online]. Available: <https://doi.org/10.1109%2FIROS.2017.8202315>
- [25] J. Clark and D. A. Holton, *A first look at graph theory*. World Scientific, 1991.
- [26] K. Koide, M. Yokozuka, S. Oishi, and A. Banno, “Voxelized GICP for Fast and Accurate 3D Point Cloud Registration,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 11 054–11 059.
- [27] Q. Li, X. Yu, J. P. Queralta, and T. Westerlund, “Multi-Modal LiDAR Dataset for Benchmarking General-Purpose Localization and Mapping Algorithms,” 2022. [Online]. Available: <https://arxiv.org/abs/2203.03454>