Robust Structure Identification and Room Segmentation of Cluttered Indoor Environments from Occupancy Grid Maps

Matteo Luperto[*1], Tomasz Piotr Kucner[2,3], Andrea Tassi[4], Martin Magnusson[2], and Francesco Amigoni[4]

Abstract— Identifying the environment's structure, i.e., to detect core components as rooms and walls, can facilitate several tasks fundamental for the successful operation of indoor autonomous mobile robots, including semantic environment understanding. These robots often rely on 2D occupancy maps for core tasks such as localisation, motion and task planning. However, reliable identification of structure and room segmentation from 2D occupancy maps is still an open problem due to clutter (e.g., furniture and movable object), occlusions, and partial coverage. We propose a method for the RObust StructurE identification and ROom SEgmentation (ROSE^2) of 2D occupancy maps, which may be cluttered and incomplete. ROSE^2 identifies the main directions of walls and is resilient to clutter and partial observations, allowing to extract a clean, abstract geometrical floor-plan-like description of the environment, which is used to segment, i.e., to identify rooms in, the original occupancy grid map. ROSE^2 is tested in several real-world publicly-available cluttered maps obtained in different conditions. The results show how it can robustly identify the environment structure in 2D occupancy maps suffering from clutter and partial observations, while significantly improving room segmentation accuracy. Thanks to the combination of clutter removal and robust room segmentation ROSE^2 consistently achieves higher performance than the state-of-the-art methods, against which it is compared.
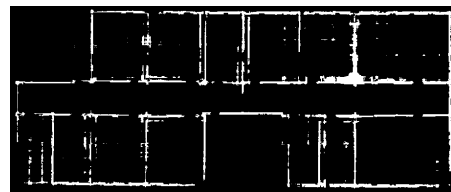
I. INTRODUCTION

In recent years, ground mobile robots have been deployed in numerous indoor applications including industrial, public, office, and domestic environments. These robots often rely on 2D occupancy maps for core robotic tasks such as localization, motion and task planning.

Occupancy maps represent the shape of the environment through a grid, where each cell is associated to a probability of being occupied by an obstacle [1]. However, such maps do not provide higher-level semantic information about the environment (i.e., the type of objects, the structure of the environment, the room type [2], [3]).
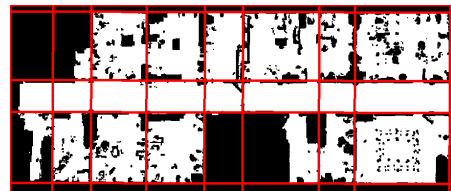
Considering how semantic information about structural layout plays an important role in numerous robotic applications such as localisation [4], [5], exploration [6]–[8], mapping performance benchmarking [9], and room segmentation [10], it is not surprising that the structure extraction problem has received substantial attention recently. Even though indoor environments are predominantly well *structured* it is still challenging to extract such information

1: Università degli studi di Milano, Italy.
2: Örebro Universitet, Sweden.
3: Aalto University, Finland.
4: Politecnico di Milano, Italy.
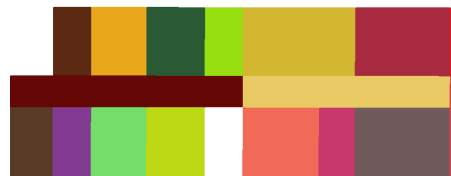* Corresponding Author: `matteo.luperto@unimi.it`
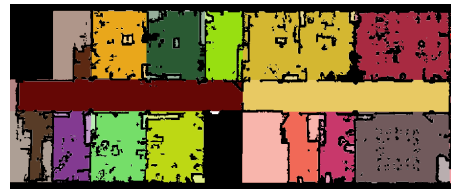
(a) Occupancy grid map.

(b) Removal of non-structural components.

(c) Wall lines identification.

(d) Geometric floor-plan-like representation.

(e) Room segmentation and room-shape prediction.

Fig. 1: Our method (ROSE^2) starts from a (cluttered or partial) 2D occupancy grid map of an indoor environment and identifies the structure by retrieving its rooms and walls. To do so, it filters clutter and non-structural features (b), it retrieves the main wall directions (c), and obtains a geometrical representation of the environment (d), that is used to segment the map in rooms while predicting their shape (e).

from 2D occupancy maps. This is usually caused by the presence of clutter (e.g., from furniture, movable objects) that occludes the background (e.g., walls) [11]. Several methods that try to identify structure from 2D occupancy maps (e.g., by segmenting the map into rooms) perform poorly in cluttered environments [12], as they are based on local map features such as corners and narrow passages. However, 2D occupancy maps provide little insights to distinguish between structural components of the environment (e.g., a wall) or furniture (e.g., a chair or a table), that can both give rise to such local features. Some works tackled this problem by enriching the 2D range data with additional modalities like camera images or by using 3D point clouds [11], [13]. However, additional sensing modalities are not always available.

In this work, we propose a method for RObust StructurE identification and ROom SEgmentation (ROSE^2) from (cluttered) 2D occupancy maps. This method stems from the assumption that the structure of the environment can be defined through a (limited) set of main directions that are followed by walls.

The proposed method for structure extraction and room segmentation consists of two principal steps, as shown in Fig. 1. First, we identify all occupied cells of a 2D occupancy map that belong to the main structure (e.g., walls), by using a robust frequency-based structure extraction method [14]. At the same time, we filter out map components that are due to clutter, noise, and non-structural components.

Second, we group locally-perceived portions of walls along the identified main directions of walls in the environment, following the assumption that a wall could be shared by multiple rooms. This allows us to align walls detected from individual rooms at the building level. We use these features to build a clean, abstract geometrical floor-plan-like-representation of the environment.

The structure identified by our method may be used for multiple tasks such as (i) room segmentation, (ii) floor plan reconstruction, (iii) prediction of the missing portion of the environment, (iv) regularization, (v) map inpainting. In this work, we apply structural knowledge to improve performance of room segmentation, while we show qualitative results in tasks (ii)-(v).

ROSE^2 is applicable to different types of maps and environments as it does not rely on any assumption about the shape of the environment (as (pseudo-)Manhattan worlds), map completion (complete or partial maps), nor assume any parametrization specific for a given map. ROSE^2 does not require any training before being applied to new settings and can be used online during robot activities.

The experimental evaluation is performed by showing how our method effectively identifies the structure and performs room segmentation in several environments, both using real-world cluttered and partial occupancy maps and on the benchmark dataset for room segmentation of [10]. Thanks to the combination of clutter removal and robust room segmentation ROSE^2 consistently achieves a higher performance when compared with the three methods from

the state-of-the-art discussed in [10].ß

This paper builds on previously published work about different yet related topics, on structure extraction [14], [15] and shape prediction of unobserved rooms [16]. In this paper, we integrate an improved version of our work of [16] with the robust feature extraction method of [14] to achieve a framework that could be robustly used on all types of 2D occupancy maps for structure extraction and room segmentation. This improves the overall performance of the framework in two ways. First, it simplifies the steps of the method and removes the requirement of a manual tuning of a set of parameters for each type of maps (as partial or complete maps, simulated or real ones). Second, we remove the limitation of [16] to work only with clean simulated maps while also significantly improving performance in detecting the environment's structure.

II. RELATED WORK

A key feature of indoor environments is that they are built with an underlying *structure* that is due to the fact that they are composed of doors, walls, rooms, and hallways [17]. Identifying the environment' structure is a key factor for addressing tobot applications, among others, as localisation [4], [5] and exploration [6]–[8]. In this section, we provide an overview of methods proposed to detect structure from occupancy robot maps. In particular, we focus on the task of *room segmentation*, i.e., the detection and segmentation of a map into a set of rooms.

The survey of Bormann et al. [10] compares the state-of-the-art techniques used to perform room segmentation on 2D maps. In particular, [10] identifies three main groups of approaches. The first and most popular one is that of Voronoi-based approaches [18], that segment the environment using a Voronoi graph, a spatial partition of the map whose nodes and edges have a maximal distance from at least two points of a finite set of obstacles. The second group is that of morphological operators, which segment the environment by using a transform on the map. An example is the morphological segmentation method [19], that grows iteratively obstacles in the map until two connected areas become separated. A second example is the distance transform-based segmentation [20], that uses the distance of each empty pixel to the closest border/obstacle for segmentation. The third group, that includes [2], relies on learning grid-cell labels from local appearances (e.g., by classifying 2D raw sensorial inputs) and harmonizing neighbouring labels afterward. In this way, the process of room segmentation is fused with that of semantic mapping (e.g., determining that a room is also a corridor or an office). The performance of all these methods, as we show in Section IV, is unstable when applied to cluttered maps.

Mielle et al. [12] presents a method for room segmentation of 2D maps using the layout of the free space, by detecting ripple-like patterns and by merging neighboring regions with similar values. However, such a method cannot be applied to cluttered and non-empty environments.

The work of [21] addresses the task of room segmentation of a 2D occupancy map as a computer vision problem using an encoder-decoder architecture to identify rooms and corridors. Differently from us, this method requires training on the type of maps to be segmented. Clutter maps, which can present significantly different features than those considered for training in [21], were not investigated.

The method of [22], starting from a 2D occupancy map, reconstructs the geometrical shape of rooms by using Markov Logic Networks and data-driven Markov Chain Monte Carlo (MCMC) to sample over several possible room shape candidates, and selecting the fittest candidate according to the sensor data and a probabilistic room model.

Several methods extract the environment structure from 3D point clouds [13], [23]–[25]. While being partly inspired by those methods, our work focuses on a different type of input as 2D occupancy maps.

The task of using knowledge obtained from camera images and 3D point clouds to perform structure detection, room segmentation, and semantic mapping on 2D maps has also been investigated in several works. The work of [26] shows how the integration of heterogeneous multi-modal information from 2D laser range scanners and vision can be fused into a probabilistic framework to obtain a semantic segmentation of the map into rooms.

A similar approach is presented in [27] where a deep learning vision-based place categorization method provides a segmentation and semantic map of the 2D occupancy maps. Another recent work following a similar method is the one of [28], which integrates a vision-based scene classifier and an object detector to segment 2D grid maps into rooms and, at the same time, distinguish the semantic labels of that environment.

The work of [29] addresses a different, but related, problem. Starting from a 3D point-cloud map of the environment the method extracts a 2D occupancy map where different rooms are detected. To do so, they identify in 3D several structural features of the environment as regions, volumes, and passages, that are used to extract a topological map.

Our works of [15], [16] present a method to predict the shape of partially-observed rooms in 2D occupancy maps after computing a geometrical floor-plan-like representation of the environment called the building layout. While showing promising results in clean maps without furniture, the method required an ad hoc parametrization for each map in order to work, which jeopardizes its applicability to cluttered maps. Nevertheless, we show in [6] how such predicted structural knowledge could be used online to improve performance in exploration for map building. Another application is in [30], that shows how the identification of the structure of the environment can help to predict the shape of multiple closed rooms that are behind closed doors, then performing inpainting on the map of the predicted shape of unobserved closed rooms.

III. Our Method

ROSE^2 identifies the structure present in indoor environments from 2D occupancy maps to reconstruct the geometry of structural elements like walls and rooms, that are utilised for segmentation. Our method is divided into three steps: map cleaning and structural features identification (Section III-A), wall detection (Section III-B), and room detection (Section III-C).

The proposed pipeline is designed to work with regular categorical 2D occupancy grid maps. In this type of map each cell can belong to one of three categories: occupied, free, or unknown. Besides this, our approach is independent of the configuration of the robot system and of the SLAM algorithm used to build the occupancy map.

A. Map cleaning and structural features identification

Typically, for ground robots, occupancy maps M obtained in real-world environments are particularly noisy, since the lidar used as the main sensorial input to build them is commonly placed close to the floor level. As a result, the sensor perceives and adds to the map non-structural objects such as legs of tables and chairs, bags, noise due to reflections caused by mirrors, French windows, and glass walls. Consequently, detecting the structural features, like walls, in such maps could be challenging.

Thus, in the first step of the proposed pipeline we implement a method for differentiating between structural elements of the environment and noise (i.e., clutter, spurious measurements, and non-structural elements of the environment) and further removing noise in order to obtain a *clean map* $\bar{M}$. To do so, we rely on the method that we presented in [14], called ROSE.

ROSE exploits the fact that in human-made environments structural components, like walls, are organised along a limited number of *dominant directions* Ψ.

As a consequence the dominant directions can be observed in the frequency spectrum as sets of radial ridges passing through the center of the frequency image. To identify these ridges ROSE computes a 2D Discrete Fourier Transform (DFT) of M. Then, lines corresponding to the dominant directions are identified. The process starts by computing the cumulative amplitude along each direction of ψ in the frequency image, then those directions that correspond to the most prominent peaks among all directions ψ are selected as Ψ. These peaks define the components of the frequency spectrum that should be retained (structure), while the reminder of the frequency spectrum will be set to zero (clutter). This process is equivalent to automatically generating a band pass filter. However, in contrast to typical image processing, the goal is to retain high-energy parts of the spectrum. Finally, through computation of the inverse DFT for the filtered frequency image, each occupied cell in the map gets a *score* denoting how much it contributes to the structure of the environment.

Note that ROSE does not assume Manhattan environments (with only right angles), but can also handle environments
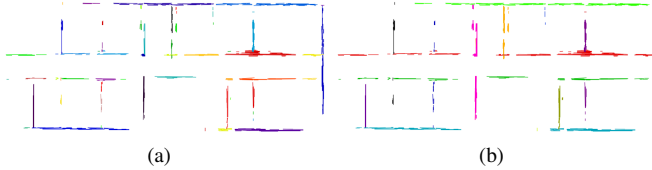
Fig. 2: Line segments obtained from the clean map $\bar{M}$ of the map shown in Fig. 1. Different clusters of collinear line segments are shown with different colors. Note how a single wall in the environment is represented as a series of slightly misaligned small line segments. Collinear clusters of (a) are merged in (b).

with more than two dominant directions. For full details about the method, please refer to [14].

Through discarding cells with a score lower than a threshold (thus identifying them as not belonging to the structural elements) the clean map $\bar{M}$ is generated. The threshold is automatically tuned by checking different threshold values for each map to have a ratio of line segments (obtained using the method from [31]) per free grid map cells in $\bar{M}$ inside a desired interval (experimentally identified), to avoid any further parameters tuning when assessing a new map type.

B. Wall detection

While the process of obtaining a clear map $\bar{M}$ and dominant directions Ψ helps in identifying the main structural components, this is not sufficient to identify walls. This is because, in $\bar{M}$, portions of the same walls are usually represented by small line segments slightly misaligned w.r.t. each other. An example can be seen in Fig. 2, that shows the line segments obtained from the map $\bar{M}$ of Fig. 1b. It can be seen how, despite the clean map $\bar{M}$ (Fig. 1b) just contains some clutter, line segments identified in it are (slightly) misaligned, despite belonging to the same wall. To mitigate this limitation, we identify *walls* by combining together line segments that are close to each other within the map. These line segments are clustered together according to their direction and to the fact that they are lying (or not) on one of the main dominant directions Ψ identified in $\bar{M}$.

The first operation on $\bar{M}$ is the detection of line segments $s \in S$ using the probabilistic Hough line transform [31]. The line segments S are then clustered together to identify those that belong to a portion of the same wall W. To do so, we rely on the method we described in [16]. At first, we cluster together all line segment with a similar angular coefficient. After that, we identify a wall $W = \{s_1, \ldots, s_n\}$ as a cluster of line segments (obtained using DBSCAN, [32]) that are spatially close to each other (i.e., there is a continuity between them) and have a similar angular coefficient. For full details about this step, please refer to [16].

We then align all line segment clusters W to dominant directions Ψ, by associating each line segment $s \in W$ with its closest dominant direction ψ. Note that, as line segment s are obtained from the clean map $\bar{M}$ from which ψ is computed, all line segments follow one of the dominant

directions. Thus, for each line segment s, we compute a line segment $\bar{s}$ which is a projection of s on a line with direction ψ passing through the middle point p of s, thus aligning all line segments in $\bar{M}$ to a dominant direction in Ψ. As a result, for each wall cluster W (of unaligned line segments s), we obtain the corresponding aligned cluster $\bar{W}$ (of line segments $\bar{s}$ aligned with a dominant direction).

As walls are shared by different rooms, we have often the case where the same wall in the environment is split into multiple clusters W, as in the case of the walls along the main corridor of Fig. 2a. To solve this, we merge together collinear wall clusters $\bar{W}$, using the following procedure. Given a wall $\bar{W} = \{\bar{s}_1, \ldots, \bar{s}_n\}$, we represent it with its central point P. The central point P is selected as the median point across all middle points p_i of the line segments $\bar{s}_i \in \bar{W}$, along a direction perpendicular to the dominant direction of W. Given two walls $\bar{W}$ and $\bar{W}'$, we call l and l' the parallel lines passing through the middle points P and P', both with dominant direction coefficient ψ. If the distance between l and l' is less than a threshold (intuitively, closer than the width of a doorway), then the two corresponding walls $\bar{W}$ and $\bar{W}'$ are merged together in $\bar{W}''$ (for which a new middle point P'' is computed). The result at the end of this process for Fig 2a is shown in Fig. 2b, where all line segments belonging to the same wall along the central corridor are assigned to the same cluster.

At this point, for each cluster W_k we obtained, a *representative line* l_k is assigned as a line whose angular coefficient is equal to a dominant direction ψ and passing through p_k. Each representative line, in red in Fig. 1c, indicates the direction of a wall within the building, and is projected across the whole map M.

The convex area delimited by intersections of representative lines is called a *face*. We call L the set of representative lines l, F the set of faces f, and we define the edges $e \in E$ as the portions of representative lines common between two faces f, f'. Note how there is a relation among all of these objects: a wall cluster W is composed of line segments s; for each wall cluster W exists a representative line l; l indicates the direction of that wall W.

We define an edge weight $w(e)$ that represents how much of e is "covered" by a projection on it of line segments s of a wall W corresponding to the representative line l on which e lies. Intuitively, if an edge e is fully covered by line segments (in $\bar{M}$), it has $w(e) \approx 1$; an edge e' located in an empty part of $\bar{M}$ has $w(e') \approx 0$. We perform then a filtering process: we remove all representative lines whose cumulative weight of their edges is below a threshold (empirically set to 0.1, so that less than the 10% of them is covered by an obstacle in $\bar{M}$) so to remove all representative lines that are caused by a local feature that may be not common to the whole environment. The rationale behind this is to identify a set of *few meaningful* representative lines that cover the walls as observed by the robot in the map M. An example is shown in Fig. 1c. To avoid to loose interesting local structural features with the filtering process, we keep only the edges $\underline{e}$ of a removed representative line $\underline{l}$ that are almost entirely
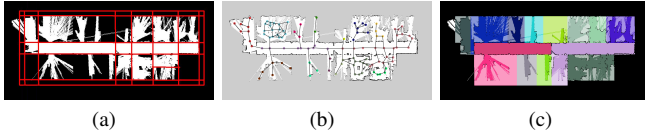
Fig. 3: Representative lines (a), Voronoi Graph (b), and resulting floor plan $\mathcal{F}$ (c) in a case where structure is not fully observed in the map.

covered (so with a large weight $w(\underline{e})$).

C. Room detection

The third and last step is to reconstruct the shape of the rooms in the environment. Rooms are found by clustering faces according to the following two rules: (i) adjacent faces whose common edge corresponds to a wall should belong to different rooms, (ii) adjacent faces that share an edge not corresponding to any wall should be part of the same room. To do so, we partition F in $\{F_1, F_2, \dots F_n\}$, in which each F_i collects the faces of a room, by clustering together faces using DBSCAN [32] as the clustering method, and using the weight $w(e_{f,f'})$ (as defined in the previous section) of the edge adjacent to faces f and f' as the metric, similarly to what we did in [16].

The faces belonging to each F_i are then merged together, obtaining a polygonal representation of the room r_i, in which the borders of the polygon are assumed to be the external walls of the room. The floor plan $\mathcal{F} = \{r_1, r_2, \dots, r_n\}$ is finally retrieved by considering together all rooms r, as displayed in Fig. 1e.

The reconstruction of the floor plan $\mathcal{F}$ is a key step in our method. It extracts the aligned shape and location of the walls as observed by the robot and, simultaneously, the border of a room in $\mathcal{F}$ predicts the presence of walls that complete the rooms even when they are not observed (yet) by the robot (e.g., due to occlusion). Note that the estimated floor plan can be used to infer the actual shape of a partially observed room, as in Fig. 1e. This step differs from the one of [16], where two different methods are used to first extract the shape of fully-mapped rooms [15], and then to predict the shape of partially observed ones.

Finally, we obtain the segmented map $\check{M}$, where empty cells in the map M are assigned to the corresponding room as identified in $\mathcal{F}$.

The output of our method is thus the set $\langle \bar{M}, L, F, \mathcal{F}, \check{M} \rangle$ which identifies the environment's structure.

D. Integrating missing structural knowledge

In cluttered environments, it can happen that neither side of a wall that divides two different rooms r and r' is directly observed, due to partial mapping or occlusion. In these cases, we cannot rely directly on walls to separate those two rooms and identify their structure.

Instead, we consider the building topology by computing a Voronoi topological graph $G = (N, T)$ of the environment, as shown in Fig. 3b, using the method of [9], where N and T

are the set of nodes and edges, respectively. In this way, we can check, for each room r, that all nodes N_r belonging to r are connected to each other (i.e., the sub-graph G_r containing all nodes N_r is a connected graph). If this condition does not hold, we split accordingly the room r into two rooms r' and r'' according to the separated components of the graph G_r. To estimate the shape of r' and r'' we rely on the symmetry of the building. If a representative line l that divides $G_{r'}$ from $G_{r''}$ exists, we use it to separate those rooms in the floor plan $\mathcal{F}$. If such a line does not exist, we divide the two rooms by using a line with the same direction of the dominant direction Ψ that can separate nodes in $G_{r'}$ from $G_{r''}$. An example is shown in Fig. 3, where Fig. 3a shows the representative lines, Fig. 3b shows the Voronoi segmentation, and Fig. 3c shows the final result with the segmented rooms and the floor plan $\mathcal{F}$ where the added lines are shown.

IV. EXPERIMENTAL EVALUATION

In this section we evaluate the capabilities of ROSE[2] to identifying semantically meaningful structure of cluttered 2D environments.

To do so, we focus on the task of *room segmentation*. Evaluation is performed both visually and quantitatively, comparing the segmented map $\check{M}$ with the actual ground-truth segmentation of the rooms M_{Gt} (obtained with manual labelling and using the environment floor plan, when available, as a reference) of the same map M. Given a room $\hat{r}$ obtained from $\check{M}$ and that of its ground-truth counterpart r obtained from M_{Gt}, we compute their *Intersection over Union* (IoU) as $\text{IoU}(\hat{r}, r) = \hat{r} \cap r / \hat{r} \cup r$.

Intuitively, we consider the area $r \setminus \hat{r}$ as a false negative, the area $\hat{r} \setminus r$ as a false positive, and the area of $\hat{r} \cap r$ as a true positive. Then, given the segmented room $\hat{r}$ of a room in $\check{M}$, we look for the corresponding room r in M_{Gt} that maximizes the overlap with $\hat{r}$ using the same approach of [10]. We scale IoU in range [0-100].

For comparison with [10], we also report the metrics of *precision* and *recall*. Precision is defined as the maximum overlapping area of a segmented room with the corresponding ground truth room, divided by the area of the segmented room; recall is defined the maximum overlap of a ground truth room with the corresponding segmented room divided by the area of the ground truth room. Note that these two metrics have complementary bias [12]; an undersegmented map (fewer rooms than the actual ones) can have high recall and low precision; conversely, a oversegmented map (more rooms than the actual ones) can have high precision and low recall). We consider the IoU as the most relevant metric as it is not affected by this bias.

We compare our results (label ROSE[2]) against publicly available methods used in the room segmentation survey of [10]: Voronoi-based (Voronoi), morphological (Morph), and distance-based segmentation (Dist). As these methods, that are described in Section II, are based on features extracted directly from the map M, their performance varies when the maps contain significant noise and clutter. Full
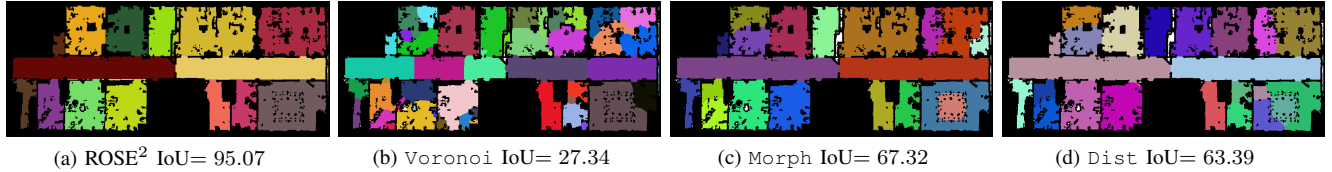
(a) ROSE² IoU= 95.07　　(b) Voronoi IoU= 27.34　　(c) Morph IoU= 67.32　　(d) Dist IoU= 63.39

Fig. 4: Room segmentation of our method and of methods from [10] on the map of Fig. 1.

	ROSE²	Morph	Dist	Voronoi
precision	**89.02** (8.39)	84.98 (6.12)	88.34 (7.85)	85.86 (9.84)
recall	**93.93** (4.21)	82.3 (11.16)	79.79 (12.15)	71.92 (13.25)
IoU	**73.3** (17.83)	51.24 (12.25)	54.65 (14.73)	28.65 (10.55)

TABLE I: Average results over 10 cluttered maps.

results from all considered maps and the code of the implemenentation of our method are available online*.

A. Results on publicly available maps

In this section, we present the results obtained in 10 fully-explored maps available from the datasets of [33], [34], as the one of Fig. 1. Those maps are processed by ROSE² regardless of the algorithm used for performing SLAM and of their size, and without changing any parameter. As those maps are obtained in real environments, like offices, they naturally present a certain degree of clutter and artefacts.

Table I presents the results. Our method clearly outperforms the other ones from [10]. The increase in performance is particularly clear when we consider the IoU, which is the metric that better describes how our method can produce a robust and stable segmentation. Fig. 4 compares the results obtained by our method with those obtained by other methods in segmenting the map of Fig. 1, while Fig 5 shows results in a map with a non-Manhattan floor plan. In both examples, not only our method performs better than the others in terms of IoU, but also the errors of our method in segmenting the map are less critical, as the segmentation is coherent with the structure of the environment. At the same time, the methods of [10] often segment the environment in a way that is not semantically coherent with its structure, as for example in Figs. 4b and 5c.

Fig. 6 shows results in a map where several artifacts due to perception issues (e.g., glass walls) are clearly visible. Nevertheless, our method, differently from those of [10], compensates for those map inaccuracies and provides a meaningful segmentation of the environment, identifying its structure.

To test the robustness and performance stability of ROSE², we report the results on the dataset used as a benchmark in the survey on room segmentation methods of [10]. The dataset is composed of 20 clean maps and 20 maps with artificial additional noise of large-scale structured environments†. Those maps are significantly different (and simpler)
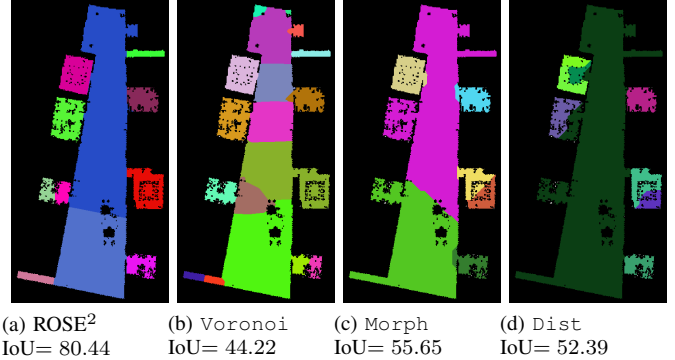
*https://github.com/goldleaf3i/
declutter-reconstruct
†http://wiki.ros.org/ipa_room_segmentation

(a) ROSE²　　(b) Voronoi　　(c) Morph　　(d) Dist
IoU= 80.44　　IoU= 44.22　　IoU= 55.65　　IoU= 52.39

Fig. 5: Room segmentation of our method and of methods from [10] on a non-Manhattan map.

(a) ROSE² IoU= 38.15　　(b) Voronoi IoU= 10.23

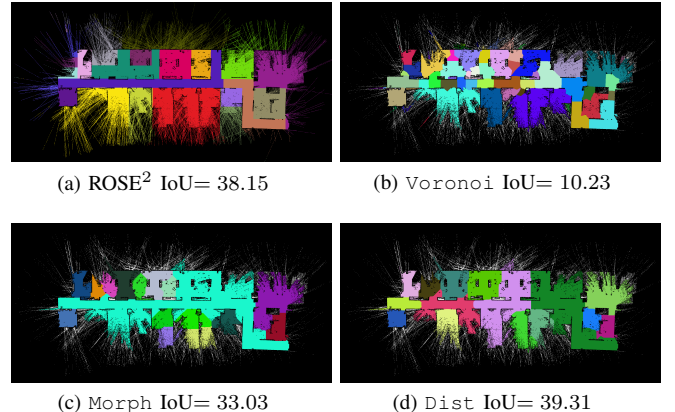(c) Morph IoU= 33.03　　(d) Dist IoU= 39.31

Fig. 6: Room segmentation of our method and of methods from [10] on a map with several perception artifacts.

from real-world maps, as they do not present clutter nor noise typically present in real-world maps, but are empty environments (no furniture) or with geometric noise added (furnished).

On these maps, ROSE² obtains results that are similar to those obtained in cluttered maps (and that are better than those of [10] and of our previous work of [16]), with a 93.54 (5.46) precision and 91.03 (3.46) recall on furnished maps, and 93.26 (6.4) precision and 97.58 (2.2) recall on no furniture, without changes in parameters. Full results are reported in the repository.
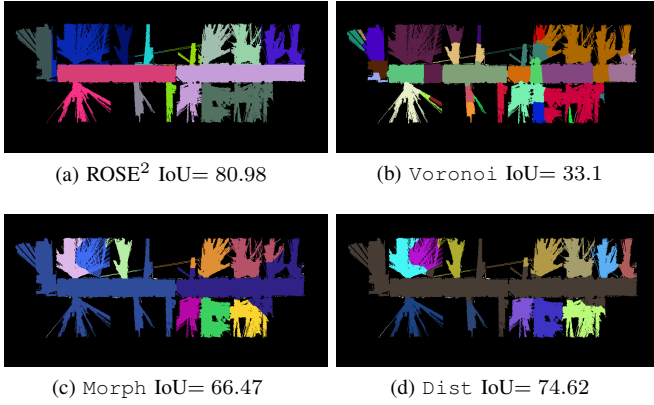
(a) ROSE^2 IoU= 80.98   (b) `Voronoi` IoU= 33.1

(c) `Morph` IoU= 66.47   (d) `Dist` IoU= 74.62

Fig. 7: Room segmentation of our method and of methods from [10] on a partial map obtained during exploration.

B. Results on partial maps

In this section, we evaluate the results obtained in several maps coming from the incremental exploration of a building. In this way we test the robustness of ROSE^2 to extract the structural features from different types of maps. For example, maps obtained in the early stages of exploration have few rooms that are only partially mapped by the robot. Data are taken from [33] using two robot runs obtained in the Freibug Building 79 (`FR79`) and in the University of Bremen Cartesium building (`Cartesium`). We relied on GMapping [35] as SLAM method. We used the last map to compute the coverage percentage of the map from the start of exploration (coverage = 0) to the full exploration of the environment (coverage = 1).

Fig. 7 shows the results of ROSE^2 compared with the methods of [10] in a partial map obtained in `FR79`. ROSE^2 can segment the environment in a meaningful way also when the structure of the building can be identified by only a few walls as partially perceived by the robot. Conversely, the results of the methods from [10] show several artifacts. Fig. 3 shows the features we extract, namely the floor plan $\mathcal{F}$ and representative lines l, obtained in the same environment of Fig 7. Note how $\mathcal{F}$ is a reliable representation of the actual shape of the environment, which could be used by the robot to have a better understanding of the shape of rooms which are only partially perceived by the robot.

Fig. 8 shows the IoU at different coverage levels for the two environments `FR79` and `Cartesium`. The performance of our methods are consistent (IoU around 80) at all degrees of map completion. At the same time, qualitatively speaking, the segmentation performed by other methods varies greatly as the map of the environment changes, while also achieving a lower quantitative performance.

C. Discussion

The results from Fig. 8 show how our method is robust in different map conditions; this is due to the fact that the steps described in Section III allow the extraction of semantically
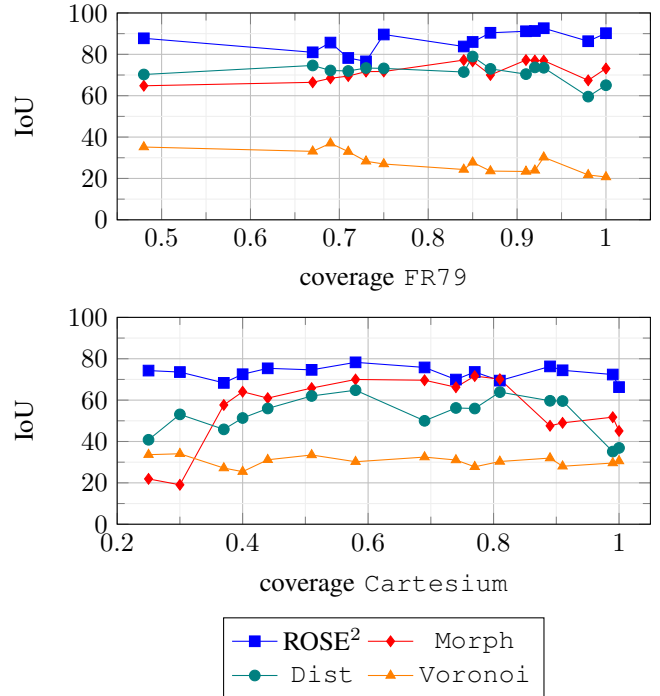

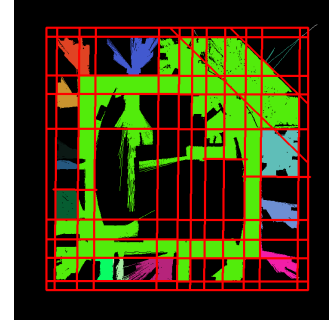Fig. 8: IoU vs. percentage of coverage for the `FR79` (top) `Cartesium` (bottom).

Fig. 9: Features extracted (representative lines, room segmentation) from a partial map of `INTEL` (IoU= 75.98).

meaningful structural features even from partial maps or from particularly cluttered ones.

Another example of this is in Fig. 9, where we show the segmentation and the representative lines retrieved from a partial map obtained from the Intel Lab (`INTEL`) [33]. Also in this environment, which is complex due to its large size and its non-Manhattan layout, our method gives a semantically meaningful reconstruction of the environment, while providing the directions of the main walls.

The retrieval of the floor plan $\mathcal{F}$ and the identification of walls and representative lines can provide meaningful insights on the actual shape of the environment also in cases the robot has not a full knowledge of the entire map. An example is shown in Fig. 10 with the inpainted room shape from the floor plan $\mathcal{F}$ for two partial maps obtained from `FR79` and `Cartesium`. Despite that there are several parts of the environment not observed yet by the robot, our method can give an accurate representation of its structure.
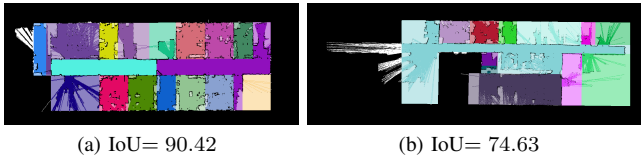
(a) IoU= 90.42   (b) IoU= 74.63

Fig. 10: Floor plan $\mathcal{F}$ and room segmentation of two partial maps from `FR79` (a) and `Cartesium` (b).

This knowledge could be used to complement the map, by estimating the actual shape of the rooms, but also to give extra knowledge and awareness of the environment to the robot [6].

V. CONCLUSIONS

In this work, we have presented a method to extract the structure of an environment from its (cluttered) 2D occupancy grid map, while performing a segmentation of the map into a set of rooms. Our method starts to identify the main directions of lines as observed in the occupancy map, and cleans the map from clutter and noise. After that, it estimates the presence of the walls in the environment and of their directions with a set of representative lines; these lines are combined to obtain a geometrical floor-plan-like representation that is used to segment the map in rooms. Results show that our method can be successfully applied to occupancy grid maps, even in the case of severe noise and clutter and when the map represents only a part of the environment, as during exploration. Future work involves the use of structural information to improve long-term mapping of changing environments.

REFERENCES

[1] H. Moravec and A. Elfes, "High resolution maps from wide angle sonar," in *Proc. ICRA*, vol. 2, 1985, pp. 116–121.
[2] O. Mozos, R. Triebel, P. Jensfelt, A. Rottmann, and W. Burgard, "Supervised semantic labeling of places using information extracted from sensor data," *Robot Auton Syst*, vol. 55, no. 5, pp. 391–402, 2007.
[3] M. Magnusson, T. P. Kucner, S. G. Shahbandi *et al.*, "Semi-supervised 3d place categorisation by descriptor clustering," in *Proc. IROS*, Sep. 2017, pp. 620–625.
[4] S.-Y. An and J. Kim, "Extracting statistical signatures of geometry and structure in 2d occupancy grid maps for global localization," *IEEE RA-L*, pp. 1–1, 2022.
[5] H. Howard-Jenkins, J.-R. Ruiz-Sarmiento, and V. A. Prisacariu, "Lalaloc: Latent layout localisation in dynamic, unvisited environments," in *Proc. CVPR*, 2021, pp. 10 107–10 116.
[6] M. Luperto, L. Fochetta, and F. Amigoni, "Exploration of indoor environments through predicting the layout of partially observed rooms," in *Proc. AAMAS*, 2021, pp. 836–843.
[7] M. Luperto, M. Antonazzi, F. Amigoni, and N. A. Borghese, "Robot exploration of indoor environments using incomplete and inaccurate prior knowledge," *Robot Auton Syst*, vol. 133, p. 103622, 2020.
[8] A. Quattrini Li, R. Cipolleschi, M. Giusto, and F. Amigoni, "A semantically-informed multirobot system for exploration of relevant areas in search and rescue settings," *Auton Robot*, vol. 40, no. 4, pp. 581–597, 2016.
[9] F. Amigoni, V. Castelli, and M. Luperto, "Improving repeatability of experiments by automatic evaluation of SLAM algorithms," in *Proc. IROS*, 2018, pp. 7237–7243.
[10] R. Bormann, F. Jordan, W. Li, J. Hampp, and M. Hägele, "Room segmentation: Survey, implementation, and analysis," in *Proc. ICRA*, 2016, pp. 1019–1026.
[11] R. Ambruş, N. Bore, J. Folkesson, and P. Jensfelt, "Meta-rooms: Building and maintaining long term spatial models in a dynamic world," in *Proc. IROS*, 2014, pp. 1854–1861.
[12] 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 *Proc. ICRA*, 2018, pp. 4993–4999.
[13] I. Armeni, O. Sener, A. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese, "3D semantic parsing of large-scale indoor spaces," in *Proc. CVPR*, 2016, pp. 1534–1543.
[14] T. P. Kucner, M. Luperto, S. Lowry, M. Magnusson, and A. J. Lilienthal, "Robust frequency-based structure extraction," in *Proc. ICRA*, 2021, pp. 1715–1721.
[15] M. Luperto and F. Amigoni, "Extracting structure of buildings using layout reconstruction," in *Proc. IAS-15*, 2018.
[16] M. Luperto, V. Arcerito, and F. Amigoni, "Predicting the layout of partially observed rooms from grid maps," in *Proc. ICRA*, 2019, pp. 6898–6904.
[17] M. Luperto, A. Quattrini Li, and F. Amigoni, "A system for building semantic maps of indoor environments exploiting the concept of building typology," in *Proc. RoboCup*, 2013, pp. 504–515.
[18] S. Thrun, "Learning metric-topological maps for indoor mobile robot navigation," *Artif Intell*, vol. 99, no. 1, pp. 21–71, 1998.
[19] P. Buschka and A. Saffiotti, "A virtual sensor for room detection," in *Proc. IROS*, 2002, pp. 637–642.
[20] A. Diosi, G. Taylor, and L. Kleeman, "Interactive SLAM using laser and advanced sonar," in *Proc. ICRA*, 2005, pp. 1103–1108.
[21] 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, 2021.
[22] Z. Liu and G. von Wichert, "A generalizable knowledge framework for semantic indoor mapping based on Markov logic networks and data driven MCMC," *Future Gener Comp Sy*, vol. 36, pp. 42–56, 2014.
[23] S. Ochmann, R. Vock, R. Wessel, and R. Klein, "Automatic reconstruction of parametric building models from indoor point clouds," *Comput Graph*, vol. 54, pp. 94–103, 2016.
[24] R. Ambruş, S. Claici, and A. Wendt, "Automatic room segmentation from unstructured 3-D data of indoor environments," *IEEE RA-L*, vol. 2, no. 2, pp. 749–756, 2017.
[25] S. Oesau, F. Lafarge, and P. Alliez, "Indoor scene reconstruction using feature sensitive primitive extraction and graph-cut," *ISPRS J Photogramm*, vol. 90, pp. 68–82, 2014.
[26] A. Pronobis and P. Jensfelt, "Large-scale semantic mapping and reasoning with heterogeneous modalities," in *Proc. ICRA*, 2012, pp. 3515–3522.
[27] N. Sünderhauf, F. Dayoub, S. McMahon, B. Talbot, R. Schulz, P. Corke, G. Wyeth, B. Upcroft, and M. Milford, "Place categorization and semantic mapping on a mobile robot," in *Proc. ICRA*, 2016, pp. 5729–5736.
[28] C. Jin, A. Elibol, P. Zhu, and N. Y. Chong, "Semantic mapping based on image feature fusion in indoor environments," in *Proc. ICCAS*, 2021, pp. 693–698.
[29] Z. He, H. Sun, J. Hou, Y. Ha, and S. Schwertfeger, "Hierarchical topometric representation of 3d robotic maps," *Auton Robot*, vol. 45, no. 5, pp. 755–771, 2021.
[30] M. Luperto, F. Amadelli, and F. Amigoni, "Completing robot maps by predicting the layout of roms behind closed doors," in *Proc. ECMR*, 2021.
[31] N. Kiryati, Y. Eldar, and A. M. Bruckstein, "A probabilistic Hough transform," *Pattern Recogn*, vol. 24, no. 4, pp. 303–316, 1991.
[32] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. KDD*, 1996, pp. 226–231.
[33] A. Howard and N. Roy, "The robotics data set repository (Radish)," 2003. [Online]. Available: http://radish.sourceforge.net/
[34] N. Hawes, C. Burbridge, F. Jovan, L. Kunze, B. Lacerda, L. Mudrova, J. Young, J. Wyatt, D. Hebesberger, T. Kortner *et al.*, "The Strands project: Long-term autonomy in everyday environments," *IEEE RAM*, vol. 24, no. 3, pp. 146–156, 2017.
[35] G. Grisetti, C. Stachniss, and W. Burgard, "Improved techniques for grid mapping with Rao-Blackwellized particle filters," *IEEE T Robot*, vol. 23, pp. 34–46, 2007.