Path Planning Incorporating Semantic Information for Autonomous Robot Navigation

Silya Achat, Julien Marzat, Julien Moras
DTIS, ONERA, Université Paris Saclay, F-91123 Palaiseau, France
silya.achat@onera.fr, julien.marzat@onera.fr, julien.moras@onera.fr

Keywords: Autonomous Robot Navigation, Path Planning, Semantic Scene Understanding

Abstract: This paper presents an approach to take into account semantic information for autonomous robot tasks requiring path planning capabilities. A semantic pointcloud or map serves as input for generating a multi-layered map structure, which can then be exploited to address various navigation goals and constraints. Semantic-aware adaptations of A*, Transition-based RRT and a shortcut algorithm are derived in this framework, and evaluated numerically on an exploration and observation task using a reference dataset with multiple semantic classes as an illustrative test environment.

1 INTRODUCTION

The combined recent progress in learning algorithms and computational power have led to the development of efficient semantic segmentation capabilities based on data from embedded LiDAR and/or RGB-D sensors, processed by (deep) neural networks to produce 2D annotated images (He et al., 2020) or 3D pointclouds (Qi et al., 2017; Qi et al., 2020; Carvalho et al., 2019; Mascaro et al., 2021). The fully-embedded processing of this key perception asset is now being reported for Unmanned Aerial Vehicles (UAV) (Nguyen et al., 2019; Bultmann et al., 2021). This has paved the way for the development of semantic mapping algorithms, where the 3D representation of a surrounding environment also includes the categorization of the perceived 3D points or even 3D object segmentation at a higher level (Jadidi et al., 2017; McCormac et al., 2018; Rosinol et al., 2020; Grinvald et al., 2021). Compared to classical mapping structures which provide binary occupancy information, e.g. Octomap (Hornung et al., 2013) or TSDF-based mapping (Millane et al., 2018), this opens new possibilities for autonomous robot navigation to directly take into account multiple mission objectives and constraints in interaction with the environment.

In this context, we propose a systematic approach to incorporate semantic mapping information into path planning algorithms, with the derivation of semantized versions of the A* and Transition-based Rapidly exploring Random Tree (T-RRT) algorithms as well as a post-processing shortcut procedure. The semantic map is linked with the planning algorithms within a multi-layered structure: the classes are ranked by a user-defined cost representing traversability or observability constraints, and some classes are also identified as of particular interest for observation. The proposed algorithms are evaluated and compared on typical autonomous robot navigation missions, namely waypoint rallying in the presence of obstacles and different types of terrain, and the exploration of an uncharted environment with observation of detected points of interest[1].

2 RELATED WORK

A limited number of previous works have studied the exploitation of these newly available semantic clouds or maps for path planning of autonomous robots. As described in the recent survey paper (Crespo et al., 2020), a lot of effort has been put on defining several semantic map representations for various tasks but there are still few complete semantic navigators linking the proposed maps with planning algorithms, and more extensive simulation and real-world experiments should be conducted.

In the early work by (Sofman et al., 2006), a 2D cost map associated with the traversability of an off-road terrain was obtained by an aerial vehicle to allow the navigation of a ground robot in a natural environment. This map was combined with a local map obtained

[1] Video at https://tinyurl.com/SemanticPlanning

from the on-board sensors of the ground robot, and a path was then computed with a D* algorithm exploiting the local and global traversability data. However, the top view can bias the real traversability of the terrain, for example in the case of dense foliage trees on a flat and passable terrain.

In the context of inspection tasks carried out by an autonomous ground vehicle in a nuclear storage environment, (Wang et al., 2018) proposed to exploit a 2D binary custom map of obstacles, which includes locations and orientations of objects of interest so as to build an enriched map that can be exploited for inspection-oriented path planning.

In (Deeken et al., 2014), a Geographic Information System (GIS) with geometrical and semantic layers is used to build costmaps that are exploited by the ROS *move_base* package to carry out a simplified fetch-and-deliver task with a ground platform. This was a first successful attempt demonstrating that a semantic map can be used within navigation modules and not only for data visualization. In (Suriani et al., 2021), an active-vision approach has been proposed for an indoor exploration mission by a mobile robot so as to generate successive next best views based on the detected segmented objects and associated geometrical priors on their respective sizes.

The navigation of a rover in the framework of a Martian mission has been addressed in (Ono et al., 2015). The onboard sensors provide raw images and a Digital Elevation Map to plan the rover's path over rocky terrain, where each recognized rock is classified. An RRG algorithm derived from RRT allows to define waypoints through the rock types to be avoided by taking into account the positioning of the wheels, and these waypoints are included in a graph to obtain optimal paths with an A* algorithm. These considerations on the rover model and the associated path planning architecture remain however very specific and present a high computational cost for real-time exploration-and-observation tasks.

In (Jeffrey Delmerico and Scaramuzza, 2017), a semantic 2D grid has also been exploited for traversability evaluation, along with a D* path planning algorithm to reach a destination which is designated by a human operator in the context of a rescue mission.

An active perception approach has been derived in (Bartolomei et al., 2020) for the autonomous navigation of a UAV using on-board visual odometry. The idea is to use semantic classes available in a 3D voxel map to evaluate perceptually-informative scenes and therefore maintain a reliable localization. A hierarchical structure is proposed, combining an A* path planner including a penalization related to the information content of the voxels, followed by a B-Spline trajectory optimization which aims at keeping in view the most informative landmarks. The complementary development of a reinforcement learning approach to dynamically identify the most reliable regions for localization from semantic information was further presented in (Bartolomei et al., 2021), where this architecture has been evaluated in photorealistic simulations. Another related topic addressed in (Stache et al., 2021) is to generate relevant trajectories for collecting data so as to obtain a better semantic segmentation in post-processing (e.g., for agricultural applications), based on an accuracy model linking the UAV altitude and the current acquired images.

A semantic planner for a UAV equipped with a RGB camera navigating in an unknown urban environment has been proposed and evaluated experimentally in (Ryll et al., 2020). The acquired images are processed by a convolutional network to obtain a 2D segmentation and then a projected probabilistic map giving preference to roads, which are assumed to be safer to fly over. A high-level long-distance traversability graph is finally inferred by a deep neural network as a combination of pre-defined geometrical primitives and the UAV path is extracted by direct graph search.

In (Sadat et al., 2020), a neural network architecture aims at providing probabilistic semantic occupancy layers including current and predicted locations of vehicles and obstacles for a self-driving vehicle in an urban environment, using voxelized LiDAR data and a prior mapping. The vehicle trajectory is then selected among a set of motion primitives by optimizing a cost function including penalty terms related to safety (computed using the predicted semantic segmentation), and other terms related to driving comfort and traffic rules which are independent from the semantic information.

In (Wolf et al., 2019), semantic information is exploited for motion planning of an autonomous robot in an off-road environment, using an ontology-based behavioural network. This method exploits semantics for safe navigation, avoiding obstacles and potential hazards, and slowing down when approaching them, but does not generate semantically optimised paths.

In (Mozart et al., 2021), a hybrid version of A* relying on a distance map (instead of occupancy) and a vehicle collision model has been proposed for path planning in an urban environment. This allows safe navigation for this specific self-driving car application but it is not generalizable to the navigation of any robotic system in any environment that seeks to optimize the nature of the areas to be traversed.

These previous works mainly focused on a single specific objective or constraint related to the exploitation or data acquisition of semantic information. In

this paper, we propose to incorporate data from the semantic map into a multi-layered structure that can be readily adapted to many autonomous navigation tasks requiring path planning from multi-class semantic segmentation. This makes it possible to derive variations of standard planning algorithms in a more systematic way than in the above-referenced related work, so as to efficiently carry out missions with multiple objectives and constraints.

3 PROPOSED APPROACH

3.1 Problem formulation

A large majority of robotic tasks carried out by autonomous robots can be split into the definition of a high-level goal, followed by the generation of a path that the robot should follow to reach this goal while respecting a set of constraints. Examples of such tasks are the rallying of an arbitrary waypoint with obstacle avoidance or taking into account perception constraints, next-best-view exploration where successive goals are generated on the currently known frontier, fetch-and-deliver tasks where detected objects are defined as targets. The information used in this kind of process is classically included in discretized 2D or 3D maps, such as binary occupancy grids encoding the presence of obstacles, exploration grids storing the explored locations, or more generally costmaps that can encode arbitrary potential functions depending on the state of the vehicle (Jaillet et al., 2010). Multi-layered structures (Lu et al., 2014) can be then be defined to combine different types of data that can be accessed simultaneously to evaluate the quality of the high-level goals or the generated paths.

We define a configuration q in an associated bounded space $\mathbb{S}$ where the task is executed, and the position state components of q are denoted by $\xi \in \mathbb{R}^n$, with $n = 2$ for mobile robots and $n = 3$ for UAVs. The other components of q, left unspecified, could represent orientation and additional multi-body coordinates depending on the task.

A multi-layer map structure $\mathcal{M}$ is defined as a set of n_m layers $\mathcal{M}_i$, which can be used to evaluate an arbitrary cost function $c_j(q, \mathcal{M}_i)$ at any configuration q. A high-level task goal, defined here as a configuration target q_{goal} for the sake of simplicity, can then be optimally selected by evaluating a utility function $u(\mathcal{C}(q_{\text{goal}}, \mathcal{M}))$, where $\mathcal{C}(q, \mathcal{M}) = [c_j(q, \mathcal{M}_i)]^T$ contains all the relevant cost evaluations c_j $(j = 1, ..., n_c)$ from appropriate layers $\mathcal{M}_i$ for this given task.

An associated path P is defined as the ordered list of m successive configurations $q_k \in \mathbb{S}$ from the start q_0 to the goal $q_{m-1} = q_{\text{goal}}$. The problem can then be formulated as finding an optimal path against some criteria (shortest path, no-collision, granting a sufficient safety level, ...) which can also be evaluated from the appropriate set of n_c functions $\mathcal{C}(q, \mathcal{M})$. A standard formulation to evaluate the quality of the path can be derived as the weighted sum:

$$J(P, \mathcal{M}) = \sum_{k=0}^{m-2} \sum_{j=1}^{n_c} c_j(q_{k+1}, \mathcal{M}) \cdot \|\xi_{k+1} - \xi_k\| \quad (1)$$

where each segment of the path is weighted by the sum of appropriate costs extracted from the multi-layered map structure. We study here how the information stored in a semantic pointcloud or map can be further included into this process and how the definition of goals and paths can be adapted from classical algorithms. It is assumed that the semantic information on the environment can be represented by a closed frame of discernment $\Omega = \{l_1, l_2, ..., l_n\}$ containing all the labels considered l_i. This allow us to create a semantic grid $\mathcal{S}$ of the environment (in 2D or 3D, as detailed in Section 3.2), where each cell of index j of the map contains a label value $\mathcal{S}_j \in \Omega$. This semantic grid will be used as a proxy to define semantic-aware map layers. The following example of application with a map structure containing 3 layers is addressed in Figure 1.

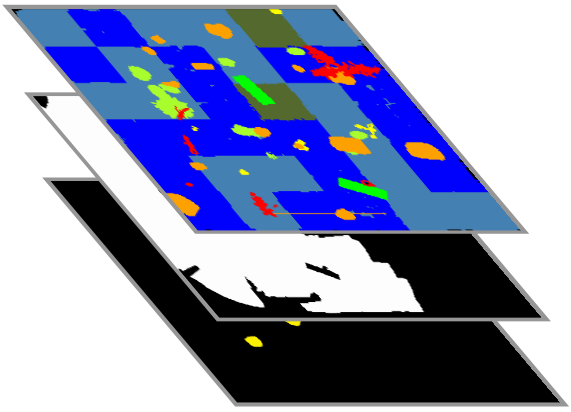

Figure 1: Example of a 3-layer semantized map structure:

- a costmap $\mathcal{M}_1$ derived from the semantic pointcloud to incorporate traversability or observation constraints.
- a classical exploration grid $\mathcal{M}_2$ with binary states $\{unknown\ ;\ explored\}$, independent from semantic information.
- a binary grid $\mathcal{M}_3$ to monitor the observation of specific semantic classes during navigation or exploration.

The semantic grid $\mathcal{S}$ is used to generate the first layer $\mathcal{M}_1$, where each cell j is assigned a cost value $c(q_j, \mathcal{M}_1)$ as a function of the corresponding semantic label l_i in $\mathcal{S}$ at the location q_j corresponding to

the cell j of $\mathcal{M}_1$. As an application example, we consider traversability constraints defined in the following way:

$$c(q_j, \mathcal{M}_1) = c(\xi_j, \mathcal{M}_1) = \begin{cases} \frac{v_{ref}}{v_i^{\max}} & \text{if } l_i \text{ traversable} \\ +\infty & \text{if } l_i \text{ not traversable} \end{cases} \quad (2)$$

where $v_i^{\max}$ is the maximum velocity reachable in safe conditions for a robot located at position ξ_j with label l_i, and $v_{ref} = \max_i(v_i^{\max})$. This can translate, e.g., constraints related to the motion of a ground robot on a specific soil type or the presence of sufficient textures in the scene for vision-based localization, as in (Bartolomei et al., 2020) for the latter. For a simple navigation task to an arbitrary waypoint considering only the information from this map layer $\mathcal{M}_1$, the proposed cost function (1) simplifies with a single cost $c(\xi_k, \mathcal{M}_1)$ as defined in (2). Thus, for a path section within a label area i such that $v_{ref} = v_i^{\max}$, the cost associated to this label would be equal to 1, so the cost for this path section would be equivalent to its Euclidean distance. It should then be minimized by a candidate planning algorithm to incorporate this semantic information in relation with the mission carried out. As a byproduct, it can be directly used to evaluate the paths obtained and thus compare the efficiency of different planners for such a task (see Section 4). Note that the infeasibility character of a path is directly obtained from the definition of non-traversable infinite costs in (2).

The second layer $\mathcal{M}_2$ is used to monitor a standard exploration mission, where the next best view (NBV) is evaluated by considering the current frontier between explored and known cells (see Section 3.4). Then, the path to this NBV is generated using the process defined above, which takes into account traversability via layer $\mathcal{M}_1$. Finally, we consider an independent subset $\mathcal{V} \subset \Omega$ containing a list of labels that should be observed during exploration, with a binary visit status updated in the corresponding layer $\mathcal{M}_3$. Note that other types of map layers could be considered for alternative tasks.

3.2 Semantic data representation

The construction of a discretized metric-semantic map is a mandatory requirement for the presented planning approach. Although the real-time construction of such a map is outside the scope of this work, since the focus here is on the comparison of planning strategies, it is reasonable to consider the availability of such a map in the near future. Indeed, several recent works in this area have given promising results. The mapping approaches proposed by (Rosinol et al., 2020; Grinvald et al., 2019) fuse directly image semantic segmentation produced by deep learning algorithms such as (Ronneberger et al., 2015; He et al., 2020) to build 3D TSDF semantic maps. Others works as (Thomas et al., 2019; Landrieu and Simonovsky, 2018) on pointcloud semantic segmentation reach a high level of accuracy and could also be used to produce such a map.

For the purpose of this work, we have used a semantic pointcloud taken from the *Garden* outdoor dataset of the 3DRMS reference challenge (Tylecek et al., 2018), as a basis for the numerical experiments reported in Section 4. To obtain a discrete semantic map structure, the labeled pointcloud is voxelized following a transformation similar to the one proposed in (Guiotte et al., 2019). For each voxel, its label $l \in \Omega$ is assigned following a majority vote including all the points belonging to this voxel. If no point belongs to a voxel then its label is assigned to *free*. Let h be the resolution of a voxel, a point $\{x, y, z\}$ is part of a voxel $\{i, j, k\}$ if it satisfies:

$$\begin{cases} x \text{ s.t. } ih \leq x < (i+1)h \\ y \text{ s.t. } jh \leq y < (j+1)h \\ z \text{ s.t. } kh \leq z < (k+1)h \end{cases} \quad (3)$$

In the case of a ground robot, a 2D grid is usually sufficient and also more suitable for traversability representation since it implicitly models the ground surface. This grid is computed by projecting along the z axis taking the label of the higher *non-free* voxel as the label of the 2D cell that does not exceeds a threshold z_{th}. This threshold, above which the voxels are not taken into account, should be set to be of the order of magnitude of the height of the robot. Taking the z_{th} threshold into account makes it possible to avoid, for example, the robot bypassing the projection of the branches of a tree if this class is considered non-traversable. Figure 2 illustrates the result of this pre-processing, where the left view shows the 3D voxelized map and the right view shows the 2D projected grid exploited in the remainder of the paper, each color representing a distinct label (see Table 1).

3.3 Planning algorithms

Modified versions of A^*, T-RRT and a shortcut procedure taking into account the mapping structure described above are provided. Note that other graph-based or sampling-based path planning algorithms can be adapted in a similar fashion.

1) A weighted A* algorithm (Ebendt and Drechsler, 2009) is used to take into account the semantic-aware traversability layer $\mathcal{M}_1$ defined in Section 3.1 to compute a path P from a starting node q_0 to a destina-

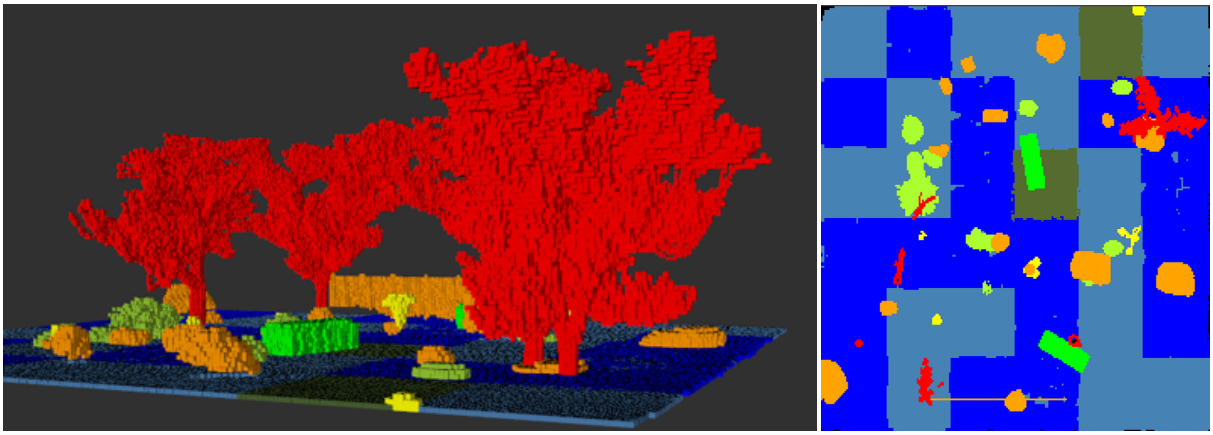

Figure 2: Voxelized 3D semantic grid (left) and 2D projection for mobile robot path planning (right) based on the 3DRMS dataset (Tylecek et al., 2018).

tion q_{goal}, with an initially unknown path length (i.e., the number of intermediate points m), minimizing the cost function (1). The navigation graph is generated with the same discretization as the map cells.

In the A* process, the transition function g from the current node q_{cur} (defined at the start as q_0 with value $g(q_0, \mathcal{M}_1) = 0$) to a candidate node q_{nei} among the eight neighbours is taken as the sum of the cumulated cost at q_{cur} and the distance between the corresponding positions ξ_{nei} and ξ_{cur} weighted by the cost corresponding to the semantic label of the node q_{nei}, as

$$g(q_{\text{nei}}, \mathcal{M}_1) = g(q_{\text{cur}}, \mathcal{M}_1) + c(q_{\text{nei}}, \mathcal{M}_1) \cdot \|\xi_{\text{nei}} - \xi_{\text{cur}}\| \quad (4)$$

The classical Euclidean distance heuristic for attraction to the goal q_{goal} is then applied without weight,

$$f(q_{\text{nei}}, q_{\text{goal}}, \mathcal{M}_1) = g(q_{\text{nei}}, \mathcal{M}_1) + \|\xi_{\text{goal}} - \xi_{\text{nei}}\| \quad (5)$$

Note that the cost $c(q_{\text{nei}}, \mathcal{M}_1)$ as defined in (2) is never less than 1, so the Euclidean distance $\|\xi_{\text{goal}} - \xi_{\text{nei}}\|$ is always lower than or equal to the actual optimal path cost, which guarantees the admissibility of the chosen heuristic.

2) Transition-based RRT (T-RRT) (Jaillet et al., 2010) is an extension of the Rapidly-Exploring Random Tree path finding algorithm which probabilizes the conservation of the new nodes of a tree, and thus the transitions of this tree, according to a costmap so as to favor the valleys and saddle points. The proposed semantic-aware version of this algorithm probabilizes the tree transitions depending on the costmap $\mathcal{M}_1$ derived from the semantic data. The transition probability p from a new sampled node q_{new} to the nearest node included in the tree q_{near} of respective costs $c_{\text{new}} = c(q_{\text{new}}, \mathcal{M}_1)$ and $c_{\text{near}} = c(q_{\text{near}}, \mathcal{M}_1)$ is then taken as

$$p(q_{\text{new}}, q_{\text{near}}) = \begin{cases} 1 & \text{if } c_{\text{new}} < c_{\text{near}} \\ \exp\left(\dfrac{c_{\text{near}} - c_{\text{new}}}{T \cdot \|\xi_{\text{near}} - \xi_{\text{new}}\|}\right) & \text{else} \end{cases} \quad (6)$$

where T is a temperature parameter. It can be noted that if q_{new} is not a traversable node (i.e. $c_{\text{new}} = +\infty$), then the transition probability p from q_{near} to q_{new} is equal to 0. In the implemented version, the *minExpandControl* introduced in the classical formulation of the algorithm to maintain a minimal rate of expansion toward unexplored regions has been deactivated and replaced by a bias towards the target with uniform probability 5%, in which case the tree is not expanded towards a random node q_{rand} of the configuration space, but towards the target node q_{goal}.

3) Paths generated by sampling-based algorithms usually require to be post-processed by a shortcut strategy (Campana et al., 2016) to remove intermediate nodes that were useful to explore the space but result in an increase of the overall length, which usually does not take into account the terrain traversability or any other related characteristic that could be obtained from semantic information. In the present context, nodes can be removed from a generated path P obtained either with the A* or the T-RRT strategy if two successive nodes of the path correspond to the same class label in the semantic grid (which also contains obstacle classes in our case). The paths are stored such that each node points to the next in the path (a node thus corresponds to a waypoint, and the last node of the path q_{m-1} points to *NULL*). The proposed semantic-aware shortcut algorithm iterates on the path with three node pointers i, j_{prev} and j which initially point respectively to q_0, q_1 and q_2. Then, as long as there is no obstacle or label change between the nodes q_i and q_j (which is verified by a line iterator travers-

ing $[q_i, q_j]$ in the semantic grid $\mathcal{S}$), then the pointers j and j_{prev} are moved forward to the next nodes of the path. When a label change is detected between i and j, then a shortcut is made between the nodes pointed by i and j_{prev} such that q_i points to $q_{j_{\text{prev}}}$ in the shortcutted path P_s. The procedure is then iterated until j points to the target node q_{m-1}.

3.4 Next-Best-View Exploration

An exploration task consisting in the observation of the full bounded space by an autonomous robot using a sensor with a field of view (FOV) limited in detection angle and range (90 deg and 3 m in our experiments) has been studied as a case study including the computation of high-level goals and path planning from semantic information. The multi-layered map $\mathcal{M}$ defined in Section 3.1 is initially unknown and discovered during exploration. A standard *Next-Best-View* (NBV) strategy (González-Banos and Latombe, 2002; Darmanin and Bugeja, 2016) has been defined, where the map layer $\mathcal{M}_2$ monitors the binary status (observed or not) of the cells of the environment. This layer is eroded with a circular structuring element larger than the radius of the robot to take into account its dimensions for obstacle avoidance. A complementary sub-task has been considered to observe objects belonging to a class of interest when they are detected during exploration (similar to the *exploration and observation* problem addressed in (Okada and Miura, 2015)), whose observation is monitored using the map layer $\mathcal{M}_3$ to store the binary observation status of this specific class (positions of the objects are updated from the semantic grid $\mathcal{S}$ when the corresponding cell has been observed). The NBV is primarily selected to observe the closest points of interest of the latter layer at a close distance in the robot FOV. When there is no remaining point of interest in $\mathcal{M}_3$ at a given instant, a random number of NBV candidates are sampled such that their position is on the *border* cells of the exploration layer $\mathcal{M}_2$ with a yaw reference given by the normal to the border (computed using a Sobel filter). The NBV is then chosen as the candidate with the maximum number of unobserved cells that could be seen in the FOV. After the NBV has been calculated, a path P between the current robot position and this NBV is computed using either the previously presented semantic-aware A* or the T-RRT algorithms followed by the semantic shortcut post-processing strategy, all relying on the map layer $\mathcal{M}_1$ containing the traversability costs and with the additional constraint that unobserved areas (in map layer $\mathcal{M}_2$) are not traversable either.

4 NUMERICAL EXPERIMENTS

Repeated simulations have been carried out to evaluate the strategies proposed in the previous section, based on the 2D projection of the semantic grid defined in Section 3.2 from the 3DRMS reference semantized pointcloud, which contains 243×256 cells with a 5×5 cm^2 resolution. Table 1 summarizes the semantic classes, with the user-defined cost they have been assigned in the map layer $\mathcal{M}_1$ corresponding to traversability (3 classes with different allowed speeds, the other ones being non-traversable) and the observation interest which is used to fill the map layer $\mathcal{M}_3$, here on the single *flower* class. It could be noted that this class is non-traversable but is of interest for observation, which further motivates the use of the proposed multi-layered structure. Table 1 also lists the colors associated to the classes, which are used in all the figures displaying 2D or 3D views. The developed algorithms have been implemented in C++ and run within a Ubuntu 18.04 Virtual Machine with 4096MB RAM on a Intel Core i5 8th generation CPU.

4.1 Path planning unitary tests

A first evaluation campaign was carried out to evaluate path planning performance with respect to the semantic-weighted cost function (1) and computation time. The weighted A* and T-RRT algorithms with shortcutting have been compared, along with a standard A* procedure (which does not use the costmap $\mathcal{M}_1$) as a baseline. For this purpose, 100 pairs of sufficiently distant start and goal positions have been sampled in the free space of the map. As the distances between each pair of start and goal positions vary, the weighted path length have been normalized by the distance between the two positions of each pair (see Table 2). A first result is that the semantic-aware weighted A* always obtains shorter weighted path lengths compared to the standard A* algorithm computed with an Euclidean distance heuristic.

Table 1: Semantic labels and related costs

Label l_i	Label Name	Color	Cost c_i ($\mathcal{M}_1$)	Visit interest ($\mathcal{M}_3$)
l_1	grass	dark blue	1.0	0
l_2	ground	pale blue	2.0	0
l_3	paving	dark green	3.0	0
l_4	hedge	bright green	∞	0
l_5	topiary	light green	∞	0
l_6	flower	yellow	∞	1
l_7	stone	orange	∞	0
l_8	tree	red	∞	0

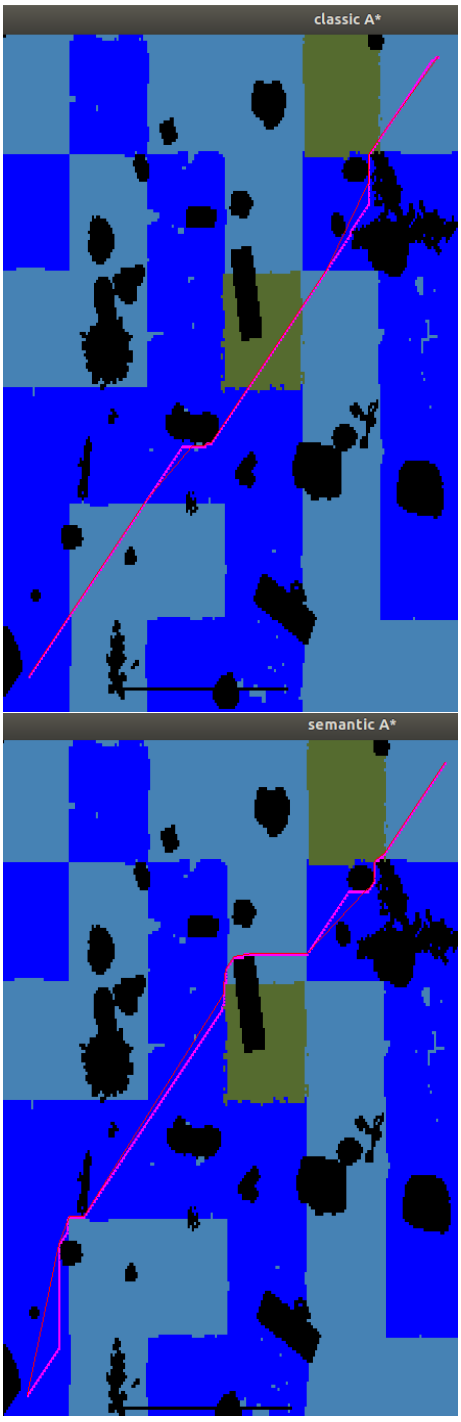

Figure 3: Paths generated with classical (up) and semantic-aware (down) A*. Initial paths are displayed in pink and semantic-based shortcuts in red.

On the other hand, the weighted version takes on average more than twice as long to execute, and is faster in only 5% of cases. This is due to the fact that the paths obtained with a classical A* are shorter in terms of Euclidean distance, and thus fewer nodes of the semantic grid $\mathcal{M}$ need to be evaluated before the algorithm reaches the target. Two respective paths generated by both algorithms are displayed in Figure 3. It can be seen that the less costly semantic class (dark blue) is favored by the weighted version. The T-RRT algorithm is on average faster to execute than the A* ones (in 82% of the runs), however the A* procedures provide shorter paths. This is a well-known trade-off when graph-based and sampling-based algorithms are compared on the same task, and this is also related to the relative simplicity of the test environment which presents many traversable areas. In 3D more complex environments, the complexity of the graph-based A* strategies would pose more computational challenges and the T-RRT algorithm will probably be more robust to dimension increase.

Table 2: Comparison of path planners with semantic information. 100 runs, averaged results (± std.)

Planning method	Path weighted length (normalized)	Computation Time (ms/pix)
Standard A*	1.70 (± 0.259)	0.0687 (± 3.31e-02)
Weighted A*	1.40 (± 0.140)	0.153 (± 5.91e-02)
T-RRT	2.02 (± 0.407)	0.0968 (± 0.191)

4.2 Exploration-and-Observation Task

The NBV exploration and observation scheme has then been evaluated using the above weighted A* and T-RRT procedures (Figure 4). Note that in the simulation, the robot speed depends on the traversability class label at its current position, according to the cost values listed in Table 1 and the inversely proportional rule defined by (2). The following two missions have been investigated: the case of a simple exploration of the map (pure exploration) and the case of an exploration with observation of objects of interest (layer $\mathcal{M}_3$), in this case those labeled l_6. 100 simulations with random initialization have been performed with each path planning algorithm, half of which aimed at visiting the flowers in the semantic map as soon as they were detected. The exploration performance indicators corresponds to the number of observed cells as a function of the number of algorithm iterations for both missions, and the additional number of iterations to visit all objects of interest for the observation task. Table 3 summarizes the performance indices obtained, while the cumulated performance indices during the missions are displayed in Figures 5 and 6. All the tasks have been success-

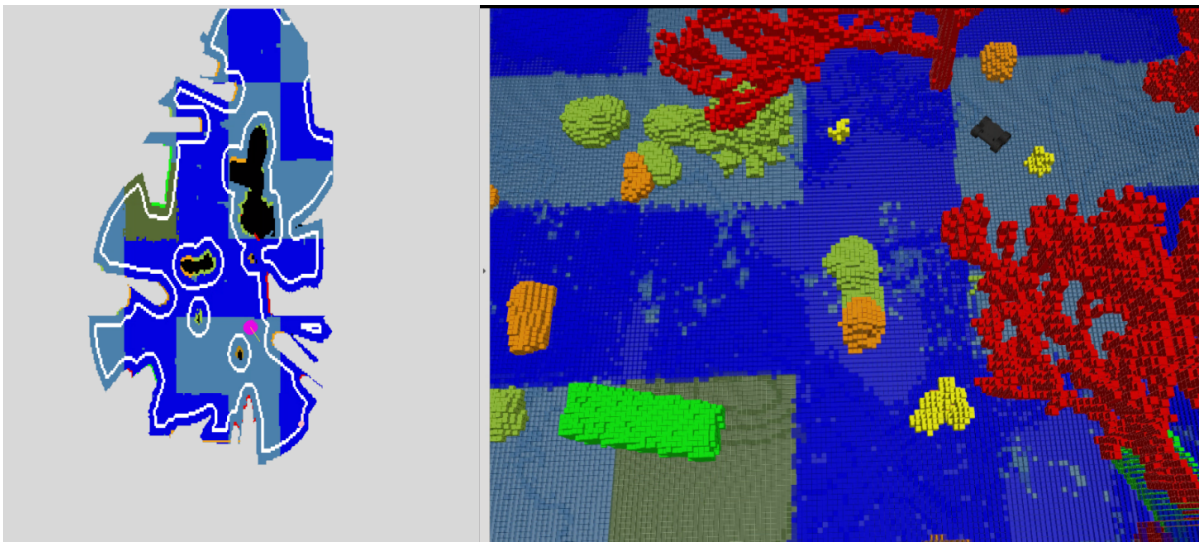

Figure 4: Example of semantic-aware exploration and observation. Left: Real-time update of coverage layer (robot position in pink, frontier nodes in white). Right: Ground truth visualization of the semantic 3D voxel grid from which information is extracted. Colors of semantic classes according to Table 1. Video available at https://tinyurl.com/SemanticPlanning.

fully completed (with a stopping coverage criterion at 90% of the free space, which corresponds to 70% of the full space) and the convergence curves are globally consistent. As a consequence of the unitary path planning tests, it follows that although the T-RRT planner is faster in calculating paths, exploration is more efficient in terms of number of iterations with the weighted A* algorithm because computation time is compensated by the optimality of the path lengths. Indeed, the robot takes less time to reach the targets and this largely compensates for the slight additional path calculation time in this 2D evaluation case. The same trend is observed in the convergence rates to the maximum number of visited objects of interest.

Table 3: Performance indices for Exploration/Observation tasks. 50 runs, averaged results (± std.)

Path planner for NBV rallying	Weighted A*	T-RRT
Nb of iterations for 90% coverage, without visiting a class of interest	1381 (± 223)	1520 (± 438)
Nb of iterations for 90% coverage, while visiting a class of interest	1581 (± 321)	1682 (± 357)
Nb of iterations to visit all objects of interest	854 (± 361)	944 (± 401)

5 CONCLUSIONS

An approach has been proposed in this paper to incorporate information available from semantic point-clouds or maps into autonomous robot navigation tasks that can be decomposed into high-level goals followed by path planning. An example of construction of a multi-layered map incorporating semantic information to define costs or observation targets has been proposed, and it has been shown how the classical A* and T-RRT planning algorithms can be adapted. This has been evaluated numerically for waypoint rallying and exploration-and-observation tasks, with promising results.

The design and testing of a fully embedded process for mobile or aerial robots including semantic learning, map building and autonomous navigation will be the focus of future work. This includes the analysis of the on-line construction of semantic maps and the best way to represent this information and associated uncertainty for autonomous navigation. The extension to multiple robots or heterogeneous teams, with the fusion of viewpoints to integrate semantic data into distributed 3D maps and use them in such planning algorithms will also be challenging problems.

ACKNOWLEDGMENTS

This research was partially supported by the French Grant ANR DELICIO (ANR-19-CE23-0006).

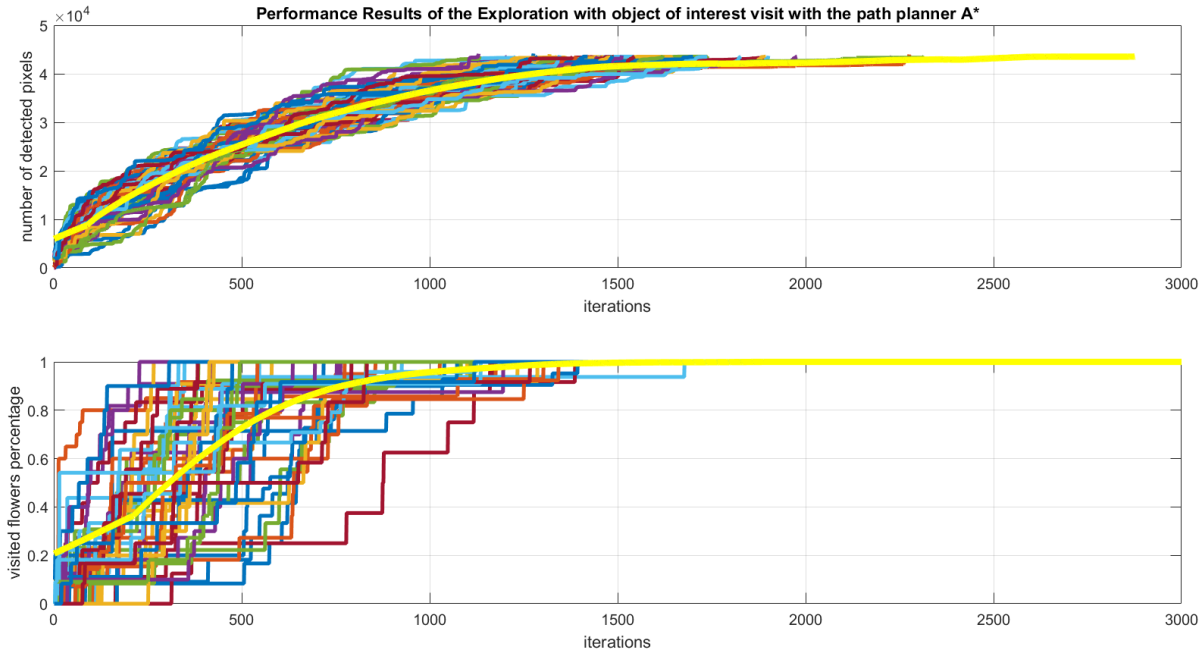

Figure 5: Cumulated coverage and visit of classes of interest during Exploration task with semantic-aware A* planner to reach NBVs (50 runs). The yellow curves correspond to the mean values.

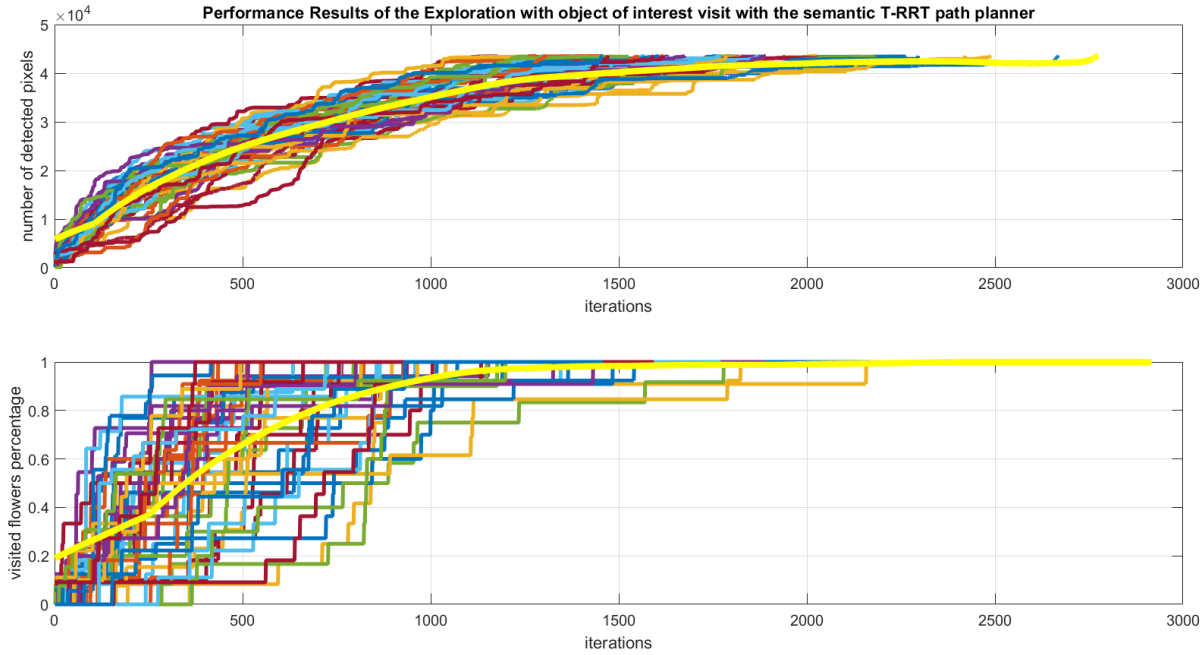

Figure 6: Cumulated coverage and visit of classes of interest during Exploration task with semantic-aware T-RRT planner to reach NBVs (50 runs). The yellow curves correspond to the mean values.

REFERENCES

Bartolomei, L., Pinto Teixeira, L., and Chli, M. (2021). Semantic-aware active perception for UAVs using deep reinforcement learning. In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).

Bartolomei, L., Teixeira, L., and Chli, M. (2020). Perception-aware path planning for UAVs using semantic segmentation. In 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 5808–5815.

Bultmann, S., Quenzel, J., and Behnke, S. (2021). Real-time multi-modal semantic fusion on unmanned aerial vehicles. In European Conference on Mobile Robots (ECMR).

Campana, M., Lamiraux, F., and Laumond, J.-P. (2016). A gradient-based path optimization method for motion planning. Advanced Robotics, 30(17-18):1126–1144.

Carvalho, M., Ferrera, M., Boulch, A., Moras, J., Le Saux, B., and Trouvé-Peloux, P. (2019). Technical Report: Co-learning of geometry and semantics for online 3D mapping. arXiv:1911.01082.

Crespo, J., Castillo, J. C., Mozos, O. M., and Barber, R. (2020). Semantic information for robot navigation: A survey. Applied Sciences, 10(2):497.

Darmanin, R. and Bugeja, M. (2016). Autonomous exploration and mapping using a mobile robot running ROS. In International Conference on Informatics in Control, Automation and Robotics (ICINCO), pages 208–215.

Deeken, H., Puetz, S., Wiemann, T., Lingemann, K., and Hertzberg, J. (2014). Integrating semantic information in navigational planning. In 41st International Symposium on Robotics, pages 1–8.

Ebendt, R. and Drechsler, R. (2009). Weighted A* search–unifying view and application. Artificial Intelligence, 173(14):1310–1342.

González-Banos, H. H. and Latombe, J.-C. (2002). Navigation strategies for exploring indoor environments. The International Journal of Robotics Research, 21(10-11):829–848.

Grinvald, M., Furrer, F., Novkovic, T., Chung, J. J., Cadena, C., Siegwart, R., and Nieto, J. (2019). Volumetric Instance-Aware Semantic Mapping and 3D Object Discovery. IEEE Robotics and Automation Letters, 4(3):3037–3044.

Grinvald, M., Tombari, F., Siegwart, R., and Nieto, J. (2021). TSDF++: A multi-object formulation for dynamic object tracking and reconstruction. In International Conference on Robotics and Automation (ICRA).

Guiotte, F., Lefèvre, S., and Corpetti, T. (2019). Attribute filtering of urban point clouds using max-tree on voxel data. In International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing, pages 391–402.

He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2020). Mask R-CNN. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(2):386–397.

Hornung, A., Wurm, K. M., Bennewitz, M., Stachniss, C., and Burgard, W. (2013). OctoMap: An efficient probabilistic 3D mapping framework based on octrees. Autonomous Robots, 34(3):189–206.

Jadidi, M. G., Gan, L., Parkison, S. A., Li, J., and Eustice, R. M. (2017). Gaussian processes semantic map representation. arXiv preprint arXiv:1707.01532.

Jaillet, L., Cortés, J., and Siméon, T. (2010). Sampling-based path planning on configuration-space costmaps. IEEE Transactions on Robotics, 26(4):635–646.

Jeffrey Delmerico, Elias Mueggler, J. N. and Scaramuzza, D. (2017). Active autonomous aerial exploration for ground robot path planning. In IEEE Robotics and Automation Letters, volume 2, pages 664–671.

Landrieu, L. and Simonovsky, M. (2018). Large-scale point cloud semantic segmentation with superpoint graphs. IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 4558–4567.

Lu, D. V., Hershberger, D., and Smart, W. D. (2014). Layered costmaps for context-sensitive navigation. In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 709–715.

Mascaro, R., Teixeira, L., and Chli, M. (2021). Diffuser: Multi-view 2D-to-3D label diffusion for semantic scene segmentation. In IEEE International Conference on Robotics and Automation (ICRA).

McCormac, J., Clark, R., Bloesch, M., Davison, A., and Leutenegger, S. (2018). Fusion++: Volumetric object-level SLAM. In International Conference on 3D Vision (3DV), pages 32–41.

Millane, A., Taylor, Z., Oleynikova, H., Nieto, J., Siegwart, R., and Cadena, C. (2018). C-blox: A scalable and consistent TSDF-based dense mapping approach. In IEEE/RSJ International Conference on Intelligent Robots and Systems, Madrid, Spain, pages 995–1002.

Mozart, A., Moraes, G., Guidolini, R., Cardoso, V. B., Oliveira-Santos, T., de Souza, A. F., and Badue, C. S. (2021). Path planning in unstructured urban environments for self-driving cars. In International Conference on Informatics in Control, Automation and Robotics (ICINCO).

Nguyen, T., Shivakumar, S. S., Miller, I. D., Keller, J., Lee, E. S., Zhou, A., Özaslan, T., Loianno, G., Harwood, J. H., Wozencraft, J., Taylor, C. J., and Kumar, V. (2019). Mavnet: An effective semantic segmentation micro-network for MAV-based tasks. IEEE Robotics and Automation Letters, 4(4):3908–3915.

Okada, Y. and Miura, J. (2015). Exploration and observation planning for 3D indoor mapping. In IEEE/SICE International Symposium on System Integration (SII), pages 599–604.

Ono, M., Fuchs, T. J., Steffy, A., Maimone, M., and Yen, J. (2015). Risk-aware planetary rover operation: Autonomous terrain classification and path planning. In IEEE Aerospace Conference, Big Sky, MT, USA, pages 1–10.

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE Conference

on Computer Vision and Pattern Recognition, pages 652–660.

Qi, X., Wang, W., Liao, Z., Zhang, X., Yang, D., and Wei, R. (2020). Object semantic grid mapping with 2D LiDAR and RGB-D camera for domestic robot navigation. Applied Sciences, 10(17):5782.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI).

Rosinol, A., Abate, M., Chang, Y., and Carlone, L. (2020). Kimera: an open-source library for real-time metric-semantic localization and mapping. In IEEE International Conference on Robotics and Automation (ICRA), pages 1689–1696.

Ryll, M., Ware, J., Carter, J., and Roy, N. (2020). Semantic trajectory planning for long-distant unmanned aerial vehicle navigation in urban environments. In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 1551–1558.

Sadat, A., Casas, S., Ren, M., Wu, X., Dhawan, P., and Urtasun, R. (2020). Perceive, predict, and plan: Safe motion planning through interpretable semantic representations. In European Conference on Computer Vision (ECMR), pages 414–430.

Sofman, B., Lin, E., Bagnell, J. A., Cole, J., Vandapel, N., and Stentz, A. (2006). Improving robot navigation through self-supervised online learning. Journal of Field Robotics, 23(11-12):1059–1075.

Stache, F., Westheider, J., Magistri, F., Popović, M., and Stachniss, C. (2021). Adaptive path planning for UAV-based multi-resolution semantic segmentation. In European Conference on Mobile Robots (ECMR).

Suriani, V., Kaszuba, S., Sabbella, S. R., Riccio, F., and Nardi, D. (2021). S-AVE: Semantic active vision exploration and mapping of indoor environments for mobile robots. In European Conference on Mobile Robots (ECMR).

Thomas, H., Qi, C. R., Deschaud, J.-E., Marcotegui, B., Goulette, F., and Guibas, L. J. (2019). Kpconv: Flexible and deformable convolution for point clouds. Proceedings of the IEEE International Conference on Computer Vision (ICCV).

Tylecek, R., Sattler, T., Le, H.-A., Brox, T., Pollefeys, M., Fisher, R. B., and Gevers, T. (2018). The second workshop on 3D reconstruction meets semantics: Challenge results discussion. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops.

Wang, M., Long, X., Chang, P., and Padlr, T. (2018). Autonomous robot navigation with rich information mapping in nuclear storage environments. In IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR).

Wolf, P., Ropertz, T., Feldmann, P., and Berns, K. (2019). Combining onthologies and behavior-based control for aware navigation in challenging off-road environments. In International Conference on Informatics in Control, Automation and Robotics (ICINCO), pages 135–146.