2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Semantic Belief Behavior Graph: Enabling Autonomous Robot Inspection in Unknown Environments

Muhammad Fadhil Ginting[1], David D. Fan[2], Sung-Kyun Kim[2], Mykel J. Kochenderfer[1], and Ali-akbar Agha-mohammadi[2]

***Abstract*— This paper addresses the problem of autonomous robotic inspection in complex and unknown environments. This capability is crucial for efficient and precise inspections in various real-world scenarios, even when faced with perceptual uncertainty and lack of prior knowledge of the environment. Existing methods for real-world autonomous inspections typically rely on predefined targets and waypoints and often fail to adapt to dynamic or unknown settings. In this paper, we introduce the Semantic Belief Behavior Graph (SB2G) framework as a new approach to semantic-aware autonomous robot inspection. SB2G generates a control policy for the robot, using behavior nodes that encapsulate various semantic-based policies designed for inspecting different classes of objects. We design an active semantic search behavior to guide the robot in locating objects for inspection while reducing semantic information uncertainty. The edges in the SB2G encode transitions between these behaviors. We validate our approach through simulation and real-world urban inspections using a legged robotic platform. Our results show that SB2G enables a more efficient object inspection policy, exhibiting similar behaviors comparable to human-operated inspections.**

Fig. 1: Autonomous robot inspection in urban environments. This figure showcases various key semantic-aware behaviors performed by our robot to enable autonomous inspections.

I. Introduction

Consider a robot tasked with inspecting a complex and unknown environment autonomously. The robot needs to examine and interact with different types of objects dispersed throughout the environment, such as high-resolution object inspection, reading gauge measurements, and navigating through stairs and alongside humans (Fig. 1). To accomplish its objective, the robot must identify and reason about the semantic information of these objects and make decisions based on this information. This capability for planning using semantic information is important to a wide range of real-world applications, including urban inspection [1]–[3], monitoring of oil and gas sites [4], exploration of subterranean environments [5]–[8], ocean exploration [9], [10], and planetary exploration [11]–[13]. In this work, we address the problem of planning and decision-making using semantic information for autonomous robot inspection.

However, performing semantic-based planning in unknown environments presents three main challenges, often rendering geometric-based planning methods ineffective. The first challenge involves addressing perceptual uncertainty arising from semantic detection, which can be attributed to limitations in sensing range, false detections, localization errors, and occlusions. The second challenge is in locating the semantic objects with the absence of or limited prior knowledge of the environment. The third challenge is integrating semantic information into the robot's planning and control framework, which traditionally relies on geometric information.

To address these challenges, we develop a framework called SB2G (Semantic Belief Behavior Graph) to enable robots to perform autonomous inspection tasks using semantic information. To account for the perceptual uncertainty of semantic observations in planning, SB2G maintains both geometric and semantic information of objects of interest as a belief state. The behavior nodes within SB2G represent different policies for controlling the robot in performing semantic-based inspection tasks. To assist the robot in locating inspection targets with high confidence, we develop an active semantic search behavior that guides the robot in reducing belief uncertainty, directing it toward areas where it can gather more reliable semantic information. The SB2G edges govern the transitions between the behavior nodes, triggered either by attaining sufficient belief confidence for semantic-based behaviors or task specifications.

Using the SB2G framework, we demonstrate that the robot can autonomously perform inspections in real world environments without prior knowledge of the map or object locations. This framework enables the robot to search for inspection targets and execute precise inspection behaviors efficiently. Additionally, we compare the resulting robot behavior with the manual control exerted by humans during inspection tasks.

In summary, our technical contributions are as follows:

[1]Department of Aeronautics & Astronautics, Stanford University, Stanford, CA, USA {ginting, mykel}@stanford.edu
[2]Field AI, Mission Viejo CA, USA {david, sung, ali}@fieldai.com

1) We introduce the SB2G framework consisting of multiple object-dependent active semantic search types and semantic-based behaviors.
2) We propose an active semantic search algorithm to observe hard-to-detect objects under semantic and perceptual uncertainty actively.
3) We design the behavior transition condition in belief space to enable reliable transitions between behaviors without needing a time-based transition condition.
4) We validate our framework through simulations and real-world demonstrations with a legged robot in various office buildings.

II. Related Work

Planning for autonomous inspection: Our approach is related to planning methods that address the problem of autonomous robot exploration and inspection in real world settings. Coverage planning is a well-studied method to explore the entire area in the environment [14]–[18]. This work addresses the problem where the robot only needs to search and inspect specific objects of interest in some part of the environment. Current state-of-the-practice approaches for object inspection usually rely on predefining routes and observation points or placing identifiable Apriltags or QR codes, making the process labor-intensive for humans [19]. Recent works in the literature combine the object mapping method with coverage planning to tackle the object inspection problem [20]–[22].

Semantic active mapping methods have gained popularity for object-based search and mapping [23]. The objective of active mapping is usually defined to maximize information gain [24], cover the meshes [20], and improve the reconstruction quality of the objects [25]. Dang et al. propose a path planning algorithm that explores a new space while improving the object observation resolution on a volumetric map [26]. Similarly, Lu et al. sample candidate trajectories and add semantics-aware cost to improve the object-centric mapping [21]. More relevant to our work, Dharmadhikari et al. present three behaviors for volumetric exploration, semantic hole coverage, and object inspection, and switches between them [22]. Our work addresses a different semantic search and mapping problem. In our inspection scenario, the objects of interest are usually hard to detect without a careful search. The robot often receives false detection or detection with low object classification confidence. This problem requires a different active search method that accounts for the semantic detection uncertainties [27], [28].

Semantic-based task planning is used to switch between object-based inspection behaviors. While some works have proposed semantic-based task planning frameworks in different application domains [29], [30], defining transition conditions for object search and mapping is nontrivial [31]. Due to semantic detection uncertainty, the reactive transition between exploration and object inspection based on object detection signals can be brittle. Consequently, behavior transition for object inspection is usually based on a predefined time or assumes perfect semantic observation [22], [32].

Statement of contribution: With respect to the current literature, our work makes the following contributions. First, instead of combining exploration and object inspection behavior or only using a single object inspection behavior for all objects, our SB2G framework can facilitate multiple object search and inspection behaviors. Searching different objects (e.g., fire extinguishers, doors, stairs) requires different types of sensors and search behavior. Second, we develop a new active semantic search to confirm the object's presence with high confidence before transitioning to object inspection behavior. The active semantic search actively increases the confidence of the belief, making it possible to set the behavior transition condition based on the belief state. Finally, we extensively validate our framework in real world conditions on a legged robot in various office buildings. We highlight the importance of our SB2G framework for reliable semantic-based behaviors.

III. Problem Formulation

We first formulate the problem of semantic-based robot inspection in unknown environments.

Robot and object geo-semantic state: We define a robot state $x_k \in X$ and objects' geo-semantic state $y_k \in Y$ at time k. The robot state $x_k := (x_k^p, x_k^q, x_k^a)$ consists of the position $x_k^p \in \mathbb{R}^3$, orientation $x_k^q \in SO(3)$, and internal state information x_k^a such as robot's locomotion and sensor status. The semantic objects state y_k represents the set of N objects $y_k = \{y_{1,k}, \ldots, y_{N,k}\}$. Each object state $y_{i,k} := (y_i^p, y_i^q, y_i^l, y_{i,k}^a)$ captures the object's position y_i^p, orientation y_i^q, class $y_i^l \in L$, and affordance status $y_{i,k}^a \in A$ of the object. We assume the objects' poses and classes are static, hence we omit the subscript k. In our example, the framework identifies and annotates objects of specific classes such as fire extinguishers, doors, and stairs, also indicating their status like 'to be inspected' or 'to be ascended'.

Robot control and transition model: Let $u_k \in U$ denote the control input, where the d_U-dimensional control space $U \subseteq \mathbb{R}^{d_U}$ accommodates various types of control inputs for both navigation and semantic inspection tasks. Examples of navigation control inputs include velocity commands and velocity limits, as well as robot locomotion modes [33], [34]. For semantic inspection tasks, control inputs may consist of actions like pitching the robot up or down to find objects, or activating sensing and data capture modules. The state evolution model $(x_{k+1}, y_{k+1}) = f(x_k, y_k, u_k, w_k)$ defines how the robot and object geo-semantic states evolve as functions of both robot control inputs and process noise w_k.

Geo-semantic observation variable and model: When the robot operates in unknown environment, the state of the semantic objects y_k is partially observable to the robot. If an object $y_{i,k}$ is visible from the current robot state x_k, the robot can obtain a geo-semantic observation $z_k \in Z$. The observation $z_k := (z_k^p, z_k^q, z_k^l, z_k^s)$ consists of the object's measured position z_k^p and orientation z_k^q, detected class z_k^l and detection confidence score z_k^s [35]. The observation model $z_k = h(x_k, y_k, v_k)$ encodes the relation between (x_k, y_k) and z_k, where v_k is the observation noise.

Belief: A belief state $b_k \in B$ is a conditional probability distribution $b_k := p(x_k, y_k \mid \mathcal{H}_k)$ over robot and objects' geo-semantic states given the history of observations and control inputs up to time k. We use b_k as the basis for decision-making. The belief is updated using a belief evolution model $b_{k+1} = \tau(b_k, u_k, z_{k+1})$, which we can be computed recursively:

$$b_{k+1} = \alpha p(z_{k+1} \mid x_{k+1}, y_{k+1}) \times \iint p(x_{k+1}, y_{k+1} \mid x_k, y_k, u_k) b_k dx_k dy_k, \quad (1)$$

where α serves as a normalization constant.

Policy, reward, and cost: Given the current belief b_k, the robot generates action according to a policy $u_k = \pi(b_k)$, where $\pi \in \Pi$. To find an optimal policy for the robot, we define semantic task rewards $r_l(b_k, u_k)$, for different object classes $l \in L$, and costs $c(b_k, u_k)$. The robot gets a reward when it successfully inspects objects, climbs stairs, or enters doors, while the cost includes distance traveled by the robot.

Given the preceding descriptions and formulations, we can formally define the problem:

Problem 1 (Semantic-based robotics inspection in unknown environment): Given a current belief b_0 and semantic object tasks defined by $r_l(b_k, u_k)$, find an optimal inspection policy π

$$\pi = \arg\max_{\pi \in \Pi} \mathbb{E}\left[\sum_{k=0}^{\infty} \sum_{l \in L} r_l(b_k, \pi(b_k)) - c(b_k, \pi(b_k))\right] \quad (2)$$

$$\text{s.t. } b_{k+1} = \tau(b_k, \pi(b_k), z_k)$$
$$z_k \sim p(z_k \mid x_k, y_k),$$
$$x_{k+1} = f(x_k, y_k, \pi(b_k), w_k).$$

IV. Semantic Belief Behavior Graph (SB2G)

To solve Problem 1, we propose the Semantic Belief Behavior Graph (SB2G) framework. In SB2G, nodes represent distinct robot behaviors, and edges denote transition conditions to switch between behaviors. This framework is similar to finite state machines. SB2G enables the robot to efficiently navigate and inspect the environment by combining semantic understanding, geo-semantic state uncertainty, and behavioral decision making. The architecture of SB2G is shown in Fig. 2.

In this section, we first detail the belief representation b and prediction model τ that are used for decision making in SB2G (Section IV-A). Then, we discuss three types of SB2G behavior for policy π to control the robot: geometric coverage, semantic-based behaviors (Section IV-B) and active semantic search behaviors (Section IV-C). Active semantic search enables the robot to control the uncertainty of semantic beliefs for a reliable and informed transition to semantic-based behaviors. Finally, we discuss how to design transition conditions to switch between behaviors and use SB2G for planning (Section IV-D).

A. Belief Representation and Prediction

Belief representation: We represent the robot belief and semantic object belief separately. The belief of robot pose $p(x^p, x^q)$ is represented as a Gaussian distribution. We use a LIDAR-based odometry method to estimate $p(x^p, x^q)$ [36]. Meanwhile, we represent the semantic object belief $p(y)$ as an array of the belief of detected object $p(y_i)$. The object pose belief $p(y_i^p, y_i^q)$ is represented as a Gaussian distribution and estimated with an object-based localization using LIDAR and a color camera. The object class belief $p(y_i^l)$ is represented as a categorical distribution with $|L|$ categories. The class belief is estimated with a YOLO-based object detection [37]. Both the robot's other internal states (e.g., locomotion gaits, robot's sensor status), x^a, and the object's inspection status, y^a, are known to the robot.

Observation model: The semantic observation of an object y_i can be derived from a detection model and observation likelihood. We assume the robot can distinguish semantic observation z from every object y_i.

The detection model quantifies the likelihood of the robot x detecting an object y_i. We model the detection probability for an object within the robot's field of view, $\mathrm{FoV}(x)$, using a decaying function based on distance

$$p_d(y_i, x) := \begin{cases} p_{d,l} \exp\left\{-\frac{|m_i - d(x,y_i)|}{v_i}\right\} & \text{if } y_i^p \in \mathrm{FoV}(x) \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

where $d(x, y_i)$ denotes distance between x and y_i. The constants $p_{d,l}$, m_l, and v_l specify the base detection probability, optimal detection distance, and the decay rate for y_i^l.

When the robot detects an object, the observation likelihood models the probability of obtaining the observation z. Since z is conditioned on the object state y_i, the observation measurements can be assumed to be independent of each other. We model the pose measurement likelihood $p_{pq}(z^p, z^q \mid y_i, x)$ as a Gaussian distribution with mean (y_i^p, y_i^q) and covariance (Σ_i^p, Σ_i^q). From experiments, we observe the noise of pose measurement and detection score increases with the distance to object $d(x, y_i)$ and true bearing angle $\beta(y_i, x)$. The noise also depends on the object class y_i^l. We model (Σ_i^p, Σ_i^q) as

$$\Sigma_i^p = \mathrm{diag}((\sigma_{p,d} d(x, y_i) + \sigma_{p,\beta} \beta(y_i, x) + \sigma_{p,l} \gamma(y_i^l))^2, \quad (4)$$
$$\Sigma_i^q = \mathrm{diag}((\sigma_{q,d} d(x, y_i) + \sigma_{q,\beta} \beta(y_i, x) + \sigma_{q,l} \gamma(y_i^l))^2.$$

The constant σ and function γ are learned from data.

The measurement of semantic class z^l can be assumed to be independent of x due to the scale and orientation invariance of the object detection module. The semantic class likelihood $p_l(z^l \mid y_i^l)$ is derived from the confusion matrix of the object detector and can be learned from data.

Meanwhile, experimentally we observe the score measurement z^s decreases with $d(y_i, x)$ and we model it as

$$p_s(z^s \mid y_i^l, x) = p_{s,l} \exp\left\{-\frac{|m_l - d(x, y_i)|}{v_l}\right\}. \quad (5)$$

Finally, the observation likelihood can be defined as

$$p_z(z \mid y_i, x) := p_{pq}(z^p, z^q \mid y_i, x) p_l(z^l \mid y_i^l) p_s(z^s \mid y_i^l, x). \quad (6)$$

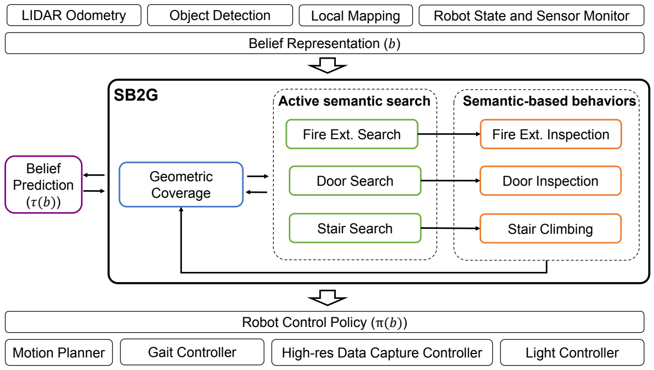

Fig. 2: The system architecture of SB2G. SB2G gathers belief state b from various perception modules. SB2G selects a behavior based on b and uses belief prediction τ to compute a robot control policy π. The policy controls robot locomotion, high-resolution data capture, and lighting modules.

State transition model: For the transition model of the robot pose $p_x(x^p_{k+1}, x^q_{k+1} \mid x_k)$, We use a unicycle model that captures longitudinal and lateral velocities of the system, which are characteristics often found in legged robots. We assume deterministic transitions for other state variables x^a, such as the robot's locomotion mode and the sensor in use.

Belief prediction model: Based on the state transition and observation models, we formulate the belief transition function τ according to Eq. (1).

B. Semantic-based Behavior

Semantic-based behaviors enable a robot to perform specific semantic tasks. When a semantic-based behavior is triggered in SB2G, it executes control actions according to a policy π^{i-l} to accomplish a specific task i for a particular semantic class l. To gain rewards r_l for the task, the robot needs to execute the behavior in a certain range of the belief state B^l. For example, B^l can be defined as a set of beliefs when the estimated object location $p(y^p)$ is in close proximity to the robot with a low uncertainty and the semantic class confidence $p(y^l)$ is high.

The design of a semantic-based behavior policy depends on the task. In this work, we use three different types of behavior to enable a complete autonomous inspection: object inspection, stair climbing, and geometric coverage.

Object inspection: This behavior enables the robot to perform a close-range inspection task. In the urban inspection scenario, we are interested in inspecting fire extinguishers and doors. This object inspection involves controlling the robot according to a policy $\pi^{\text{inspect}-\text{FE}}$ or $\pi^{\text{inspect}-\text{Door}}$. For fire extinguisher inspection, the robot needs to read the pressure gauge. For door inspection, the robot needs to check whether the door is closed or open and measure the dimension of the door. The robot achieves its inspection objective when it successfully measure the correct object. Performing an incorrect inspection in the absence of the correct object will waste operational time.

Stair climbing: This behavior is crucial for autonomous inspection in multi-level environments. The stair climbing control policy $\pi^{\text{climb}-\text{Stairs}}$ is different between mobility system [2], [38], [39]. For legged robots, the robot first needs to estimate the pose of the stair in the proximity of the robot using dense point cloud data. After positioning the robot in front of the stairs, the robot needs to switch its gait locomotion policy to a stair climbing mode before traversing the stairs. Successful stair climbing expands the robot's inspection capabilities considerably. However, failed execution due to incorrect stair localization can be catastrophic.

Geometric coverage: When the robot is not performing semantic-based behaviors and active semantic search, the robot explore the unknown environment using the geometric coverage behavior. We use a coverage planning algorithm π^{coverage} to explore the obstacle-free space in the environment [14]. The policy plans robot trajectory to sweep the free space with the sensor footprint. Geometric coverage ensures the robot to cover the unknown environment efficiently while searching for inspection targets and stairs.

C. Active Semantic Search Behavior

To execute the semantic-based behaviors successfully, the robot needs to have a low belief uncertainty about the object state. However, in real world operations, achieving high-confidence estimates of the object state is challenging due to perceptual uncertainties To address this challenge, we develop an active semantic search behavior to guide the robot to perform actions to increase the confidence of the belief.

When active semantic search is triggered, the robot execute a policy ρ^l to reduce the belief uncertainty of y_i for the expected class l. The policy ρ^l is parameterized by the object class l to account for the varying strategies required for locating different types of objects.

The policy ρ^l drives the current belief b to a set of target beliefs B^{target}. The belief target for ρ^l is a union of two belief set, $B^{target} = B^l \cup B^a$. The first set B^l represents beliefs with high confidence in the semantic object y_i. The second set B^a represents the alternative outcome of the semantic search where y^l_i is not l with a high confidence.

To drive the belief to B^{target}, we use the entropy $\mathbb{H}(y_{i,0:T} \mid z_{1:T}, b_0)$ of the target object y_i conditioned over the future observations. The conditional entropy is an appropriate objective function because it quantifies the amount of information needed to describe the belief of the target object y_i given the probabilistic value of semantic observation z. The active semantic search behavior solves the following problem:

Problem 2 (Active semantic search): Given a current belief b_0, the target object belief $p(y_i)$ with expected class l, compute a policy ρ^l:

$$\begin{aligned} \rho^l_{0:T}(p(y_i)) = &\arg\min_{\rho_{0:T}} \mathbb{H}(y_{i,0:T} \mid z_{1:T}, b_0) \quad (7)\\ &\text{s.t. } b_{k+1} = \tau(b_k, \pi(b_k), z_k)\\ &\quad z_k \sim p(z_k \mid x_k, y_k),\\ &\quad x_{k+1} = f(x_k, y_k, \pi(b_k), 0). \end{aligned}$$

The policy ρ^l is executed until the belief b reaches B^{target}.

To reduce the computational cost of solving Problem 2 in real-time on the robot, we make two assumptions. First, we assume the transition model is deterministic ($w = 0$) because the transition noise of the robot pose over a short horizon is small enough for the planning problem. Second, we employ a sparse sampling method to reduce the search branching factor of the robot's action space. Our approach narrows the search on identifying the next poses that are both situated in obstacle-free zones and contribute to entropy reduction.

We compute the active semantic search policy by performing a branch and bound search in a receding horizon manner [40]. At every time step, an optimal policy $\rho^l_{0:T}(p(y_i))$ is solved and only ρ^l_0 is executed. After the belief is updated with the actual z, we compute a new ρ^l.

D. Behavior Transition in SB2G

In SB2G, behavior transitions are represented by the graph edges $e \in \mathcal{E}$. The trigger condition to transition between behaviors is governed by the SB2G transition policy $\pi^t : \mathcal{N} \times B \rightarrow \mathcal{N}$. Then, based on the selected behavior, the SB2G graph policy π^g returns a control policy for the robot $\pi = \pi^g(n_k, b_k)$.

Trigger condition: There are four approaches to designing the trigger condition in SB2G. First, we consider the edges that transition from active semantic search behaviors to semantic-based behaviors π^{i-l}. The trigger condition for these edges occurs when the belief b_k is in the set of beliefs B^l that can ensure the robot performs the semantic-based behaviors successfully. Second, edges transitioning from active semantic search behaviors to geometric coverage are triggered when $b_k \in B^a$. Next, we consider the edges that go from geometric coverage to active semantic search behaviors. Here, the robot switches to active semantic search if it detects an uninspected object y_i and the current belief is in $B^{search-l}$. The set $B^{search-l}$ consists of beliefs having low-confidence object probabilities $p(y_i)$, enabling the active search to effectively reduce the uncertainty. Finally, transitions from semantic-based behaviors to geometric coverage are triggered when the robot has successfully completed its semantic tasks.

Semantic-based planning with SB2G: We summarize the process of performing robotic inspection using SB2G until a terminal condition $B^{terminal}$ (e.g., inspected N objects) in Algorithm 1.

Algorithm 1 Planning with SB2G

Require: b_0, $SB2G = (\mathcal{N}, \mathcal{E})$, π^t, π^g, $B^{terminal}$
- $b_k \leftarrow b_0$
- $n_k \leftarrow$ 'geometric coverage'
- **while** $b_k \notin B^{terminal}$ **do**
  - Select a SB2G behavior $n_{k+1} \leftarrow \pi^t(n_k, b_k)$
  - Compute a control policy $\pi_k \leftarrow \pi^g(n_{k+1}, b_k)$
  - Apply the action $u_k = \pi_k(b_k)$ to the system
  - Observe the actual observation z_{k+1}
  - Update the belief $b_{k+1} \leftarrow \tau(b_k, u_k, z_{k+1})$
- **end while**

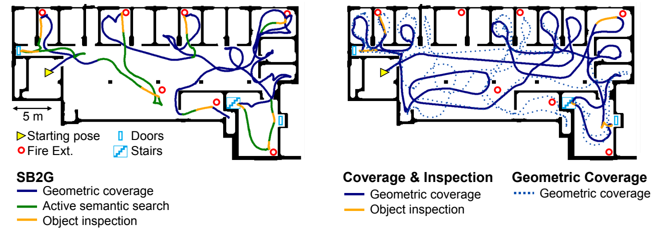

Fig. 3: Comparison of robot paths for object inspections using SB2G and baseline methods in a simulation. Our method successfully locates and inspects all objects while following a shorter trajectory.

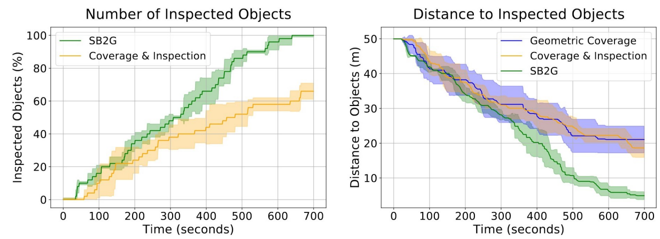

Fig. 4: The number of inspected objects (left) and the sum of the closest distance to the all 10 objects over time (right). The closest distance to each object is initially set to 5 m.

TABLE I: SB2G parameters used in the experiments.

Parameter	**Value**
$B^{search-l}, \forall l \in L$	$p(y^l_i) > 0.7$ and $\sum^p_i < 5\,\text{m}$
$B^l, \forall l \in L$	$p(y^l_i) > 0.9$, $\sum^p_i < 1\,\text{m}$, and $\mathbb{E}[d(x, y_i)] < 2.5\,\text{m}$
T	$8\,\text{s}$

V. Experimental Results

We evaluate the SB2G framework for autonomous inspection in simulation and real-world environments.

A. Simulation Results

We perform simulations in a representative office environment with a size of $38 \times 20 \times 10$ m. The simulation is carried out using the Gazebo simulator with a Boston Dynamics' Spot model. Ten semantic objects comprised of fire extinguisher, closed doors, and stairs are placed in the environment shown in Fig. 3. The robot is tasked to perform object inspections. The environment map and the true state of the objects are unknown to the robot. The parameters used for the experiments are summarized in Table I. The parameters are set based on the empirical characteristic of our object detection model. The simulations were performed on a laptop with an Intel i9-11950H CPU.

We compare our method to two methods:

1) Geometric coverage: this method only perform coverage π^{coverage} to explore uncovered free space in the environment. We compare our approach with a rollout-based coverage planning method that predicts future coverage using a sensor model similar to our sensor setup [14].

Fig. 5: Experimental results of real-world autonomous inspections using SB2G. The left figure compares the SB2G's robot paths with paths manually operated by a human during the search and inspection of fire extinguishers. The middle figures show our legged robot performing geometric coverage, semantic search, fire extinguisher inspection, and stair climbing. The right figures provide the camera view of the robot, demonstrating its capability to inspect the gauge of the fire extinguisher.

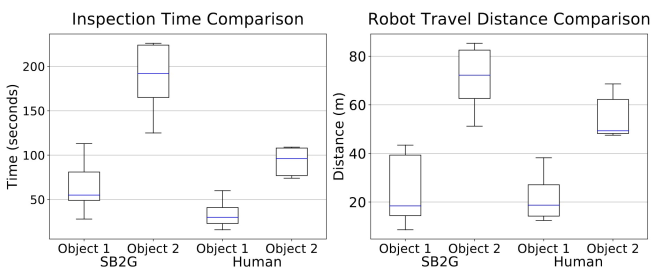

Fig. 6: The comparison of inspection time and travel distance as the robot locates and inspects each object across 5 runs.

2) Coverage and inspection: this method performs coverage π^{coverage} and object inspection ($\pi^{\text{inspect}-\text{FE}}$, $\pi^{\text{inspect}-\text{Door}}$, $\pi^{\text{inspect}-\text{Stair}}$) without the active semantic search. The robot switches to object inspection behavior when the robot has high confidence in performing the inspection ($b \in B^l$).

For each method, we run 5 simulation runs starting from the same location for 700 s.

Fig. 3 shows the comparison of the robot path and the behavior transition. The 'coverage and inspection' method inspects fewer objects than SB2G. This occurs because the robot rarely approaches the inspection targets closely and get reliable semantic observation for performing inspections. Meanwhile, SB2G generates a more efficient path by guiding the robot towards inspection targets using the active semantic search. The active search policy ρ^l controls the robot towards area where it can gather better semantic measurements.

The inspection performance is presented in Fig. 4. The SB2G method outperforms the baseline methods considerably in the number of inspection over time. In all methods, the geometric coverage behavior helps the robot in rapidly approaching inspection targets, especially at the beginning of the run when semantic observations are not yet available. However, without the active semantic search, the robot often fails to approach inspection targets closely enough, and consequently, is unable to execute inspection behaviors.

B. Field Test Results

We tested and deployed our solution in various office buildings in California, USA. In this paper, we focus on the results from experiments conducted in the Field AI office. We conducted the experiment using a Boston Dynamics Spot robot. The robot is equipped with a LIDAR and 3 cameras for navigation and semantic observation. We use a YOLO-based model to detect fire extinguishers and a point cloud-based model for stair detection. The detection model is not trained using the data gathered in the office. Two fire extinguishers are randomly placed in the environment. The robot does not have prior knowledge of the objects' true locations.

To evaluate the efficiency of our approach in real-world inspections, we compare SB2G with a robot manually operated by a human. The human operator also lacks prior knowledge of the object locations. We repeated the experiment five times with different object locations. Each run lasts for 5 minutes or until two objects had been inspected.

Using the SB2G framework, the robot can autonomously inspect fire extinguishers and climb the stairs (Fig. 5). Despite numerous false-positive detection, our method effectively locates inspection targets by actively reducing semantic belief uncertainty. The robot accurately switches between different SB2G behaviors to accomplish a fully autonomous inspection task.

The comparison between our SB2G approach and a human-controlled robot is summarized in Fig. 5 and Fig. 6. Qualitatively, as shown in Fig. 5, our approach and the human operator produce similar inspection orders and paths. According to the statistics in Fig. 6, the human-controlled robot performs inspections more quickly and with less travel distance. The primary reason for this efficiency is that humans can more easily identify inspection targets, allowing for more direct paths and quicker inspections. Although the human operator lack prior knowledge of the office map, the intuitive understanding of potential room layouts allows the operator to search rooms efficiently for the objects. These findings can serve as guidelines for the development of more efficient and semantically aware behaviors.

VI. Conclusions

We introduced the SB2G framework that enables semantic-aware autonomous robotic inspection in uncertain and unknown environments. SB2G uses semantic information to compute a control policy that guides robots through various inspection tasks. To enhance both the efficiency and accuracy of inspections, SB2G utilizes semantic belief uncertainty during planning and proactively mitigates this uncertainty before task execution. Through simulations and real-world experiments, we showed that our approach achieves more efficient inspection behaviors and is comparable to human-performed inspections. We believe this work represents an important step toward enabling more sophisticated semantic-aware behaviors in real-world inspection tasks.

References

[1] Y. Tan, S. Li, H. Liu, P. Chen, and Z. Zhou, "Automatic inspection data collection of building surface based on BIM and UAV," *Automation in Construction*, vol. 131, p. 103 881, 2021.

[2] A. Bouman, M. Ginting, N. Alatur, *et al.*, "Autonomous Spot:long-range autonomous exploration of extreme environments with legged locomotion," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, 2020.

[3] D. Lattanzi and G. Miller, "Review of robotic infrastructure inspection systems," *Journal of Infrastructure Systems*, vol. 23, no. 3, 2017.

[4] C. Gehring, P. Fankhauser, L. Isler, *et al.*, "Anymal in the field: Solving industrial inspection of an offshore hvdc platform with a quadrupedal robot," in *Field and Service Robotics: Results of the 12th International Conference*, Springer, 2021, pp. 247–260.

[5] A. Agha, K. Otsu, B. Morrell, *et al.*, "NeBula: TEAM CoSTAR's Robotic Autonomy Solution that Won Phase II of DARPA Subterranean Challenge," *Field Robotics*, vol. 2, pp. 1432–1506, 2022.

[6] M. Tranzatto, T. Miki, M. Dharmadhikari, *et al.*, "Cerberus in the darpa subterranean challenge," *Science Robotics*, vol. 7, no. 66, 2022.

[7] N. Hudson and et al., "Heterogeneous ground and air platforms, homogeneous sensing: Team CSIRO data61's approach to the DARPA subterranean challenge," *Field Robotics*, vol. 2, no. 1, pp. 595–636, 2022.

[8] S. Scherer, V. Agrawal, G. Best, C. Cao, *et al.*, "Resilient and modular subterranean exploration with a team of roving and flying robots," *Field Robotics*, vol. 2, pp. 678–734, 2022.

[9] O. Khatib, X. Yeh, G. Brantner, *et al.*, "Ocean one: A robotic avatar for oceanic discovery," *IEEE Robotics & Automation Magazine*, vol. 23, no. 4, pp. 20–29, 2016.

[10] J. J. Leonard and A. Bahr, "Autonomous underwater vehicle navigation," in *Springer Handbook of Ocean Engineering*, Springer, 2009.

[11] B. Balaram, T. Canham, C. Duncan, *et al.*, "Mars helicopter technology demonstrator," in *AIAA Atmospheric Flight Mechanics Conference*, 2018.

[12] A. Agha, K. Mitchell, and P. Boston, "Robotic exploration of planetary subsurface voids in search for life," in *AGU Fall Meeting Abstracts*, 2019.

[13] T. Touma *et al.*, "Mars dogs: Biomimetic robots for the exploration of mars, from its rugged surface to its hidden caves," in *AGU Fall Meeting*, 2020.

[14] A. Bouman, J. Ott, S.-K. Kim, *et al.*, "Adaptive coverage path planning for efficient exploration of unknown environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Kyoto, Japan, 2022.

[15] O. Peltzer *et al.*, "Fig-op: Exploring large-scale unknown environments on a fixed time budget," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022.

[16] C. Cao, H. Zhu, H. Choset, and J. Zhang, "Tare: A hierarchical framework for efficiently exploring complex 3d environments," in *Robotics: Science and Systems*, 2021.

[17] B. Charrow, G. Kahn, S. Patil, *et al.*, "Information-theoretic planning with trajectory optimization for dense 3d mapping," in *Robotics: Science and Systems*, Rome, vol. 11, 2015, pp. 3–12.

[18] A. Bircher, K. Alexis, M. Burri, *et al.*, "Structural inspection path planning via iterative viewpoint resampling with application to aerial robotics," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015.

[19] Boston Dynamics, *Autonomy Technical Summary - Spot 3.3.2 documentation*, `https://dev.bostondynamics.com/docs/concepts/autonomy/graphnav_tech_summary`, Accessed: 2024-02-02, 2023.

[20] T. Dang, C. Papachristos, and K. Alexis, "Autonomous exploration and simultaneous object search using aerial robots," in *IEEE Aerospace Conference*, 2018.

[21] L. Lu, Y. Zhang, P. Zhou, *et al.*, "Semantics-aware receding horizon planner for object-centric active mapping," *IEEE Robotics and Automation Letters*, 2024.

[22] M. Dharmadhikari and K. Alexis, "Semantics-aware exploration and inspection path planning," in *IEEE International Conference on Robotics and Automation (ICRA)*, London, United Kingdom, 2023.

[23] A. Asgharivaskasi and N. Atanasov, "Semantic octree mapping and shannon mutual information computation for robot exploration," *IEEE Transactions on Robotics*, 2023.

[24] R. P. De Figueiredo, J. le Fevre Sejersen, J. G. Hansen, M. Brandão, and E. Kayacan, "Real-time volumetric-semantic exploration and mapping: An uncertainty-aware approach," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021.

[25] S. Papatheodorou, N. Funk, D. Tzoumanikas, C. Choi, B. Xu, and S. Leutenegger, "Finding things in the unknown: Semantic object-centric exploration with an mav," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

[26] M. Grinvald, F. Furrer, T. Novkovic, *et al.*, "Volumetric instance-aware semantic mapping and 3d object discovery," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 3037–3044, 2019.

[27] N. Atanasov, M. Zhu, K. Daniilidis, and G. J. Pappas, "Localization from semantic observations via the matrix permanent," *The International Journal of Robotics Research*, vol. 35, no. 1-3, pp. 73–99, 2016.

[28] D. Morilla-Cabello, J. Westheider, M. Popovic, and E. Montijano, "Perceptual factors for environmental modeling in robotic active perception," *arXiv preprint arXiv:2309.10620*, 2023.

[29] Z. Wang, G. Tian, and X. Shao, "Home service robot task planning using semantic knowledge and probabilistic inference," *Knowledge-Based Systems*, vol. 204, p. 106 174, 2020.

[30] C. Galindo, J.-A. Fernández-Madrigal, J. González, and A. Saffiotti, "Robot task planning using semantic maps," *Robotics and Autonomous Systems*, vol. 56, no. 11, pp. 955–966, 2008.

[31] J. A. Placed, J. Strader, H. Carrillo, *et al.*, "A survey on active simultaneous localization and mapping: State of the art and new frontiers," *IEEE Transactions on Robotics*,

[32] J. Chen, G. Li, S. Kumar, B. Ghanem, and F. Yu, "How to not train your dragon: Training-free embodied object goal navigation with semantic frontiers," *arXiv preprint arXiv:2305.16925*, 2023.

[33] M. F. Ginting, S. Kim, *et al.*, "Safe and efficient navigation in extreme environments using semantic belief graphs," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

[34] M. Brandao *et al.*, "Multi-controller multi-objective locomotion planning for legged robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019.

[35] X. Lei, T. Kim, N. Marchal, *et al.*, "Early recall, late precision: Multi-robot semantic object mapping under operational constraints in perceptually-degraded environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022.

[36] A. Reinke *et al.*, "Locus 2.0: Robust and computationally efficient lidar odometry for real-time 3d mapping," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9043–9050, 2022.

[37] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[38] M. Hutter, C. Gehring, D. Jud, *et al.*, "Anymal-a highly mobile and dynamic quadrupedal robot," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2016, pp. 38–44.

[39] E. Mihankhah, A. Kalantari, E. Aboosaeedan, H. D. Taghirad, S. Ali, and A. Moosavian, "Autonomous staircase detection and stair climbing for a tracked mobile robot using fuzzy controller," in *IEEE International Conference on Robotics and Biomimetics*, 2009.

[40] M. J. Kochenderfer, T. A. Wheeler, and K. H. Wray, *Algorithms for Decision Making*. MIT Press, 2022.