

Multi-Resolution POMDP Planning for Multi-Object Search in 3D

Kaiyu Zhen[†], Yoonchang Suh[†], George Konidaris[‡], Stefanie Tellex[†]

Abstract— Robots operating in households must find objects on shelves, under tables, and in cupboards. In such environments, it is crucial to search efficiently at 3D scale while with limited field of view and the complexity of searching for multiple objects. Principled approaches to object search frequently use Partially Observable Markov Decision Process (POMDP) as the underlying framework for computing search strategies, but constrain the search space in 2D. In this paper, we present a POMDP formulation for multi-object search in a region with a frustum-shaped field-of-view. To efficiently solve this POMDP, we propose a multi-resolution planning algorithm based on online Monte-Carlo tree search. In this approach, we design a novel octree-based belief representation to capture uncertainty of the target objects at different resolution levels, then derive abstract POMDPs at lower resolutions with dramatically smaller state and observation space. Evaluation in a simulated 3D domain shows that our approach finds objects more efficiently and successfully compared to a set of baselines without resolution hierarchy in larger instances under the same computational requirements. We demonstrate our approach on a mobile robot to find objects placed at different heights in a 10m²×2m regions by moving its base and actuating its torso.

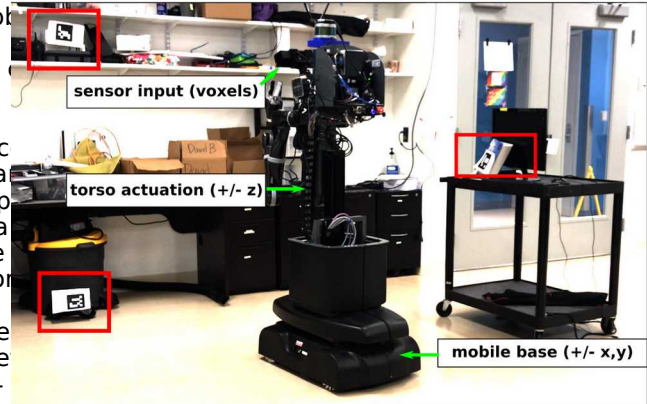


Figure 1: An example of the 3D-MOS problem where a torso-actuated mobile robot is tasked to search for three objects placed at different heights in a lab environment. The objects are represented by paper AR tags marked by red boxes. Note that the robot must actively move itself due to limited field of view, and the objects can be occluded by the attached obstacles if viewed from the side.

I. INTRODUCTION

Robots operating in human spaces must find objects such as glasses, books, or cleaning supplies that could be on the floor, shelves, or tables. This search space is naturally 3D. When multiple objects must be searched for, such as a cup and a mobile phone, an intuitive strategy is to first hypothesize likely search regions for each target object based on semantic knowledge or past experience [1, 2], then search carefully within those regions. Since the latter directly determines the success of the search, it is essential for the robot to produce an efficient search policy within a designated search region under limited field of view (FOV), where target objects could be partially or completely occluded. In this work, we consider the problem setting where a robot must search for multiple objects in a search region by actively moving its camera, with as few steps as possible (Figure 1).

Searching for objects in a large search region requires acting over long horizons under various sources of uncertainty in a partially observable environment. For this reason, previous works have used Partially Observable Markov Decision Process (POMDP) as a principled decision-theoretic framework for object search [34, 5]. However, to ensure the POMDP is manageable to solve, previous works reduce the search space or robot mobility to 2D [1, 2], although objects exist in rich 3D environments. The key challenges lie in the intractability of maintaining belief due to large state space [8], and the high branching factor for planning due to large observation space [10, 9],

In this paper, we introduce 3D Multi-Object Search (3D-MOS), a general POMDP formulation for the multi-object search task with 3D state and action spaces, realistic observation space in the form of labeled voxels within the viewing frustum from a mounted camera. Following the Object-Oriented POMDP (OO-POMDP) framework proposed by Wandzel et al. [6], the state, observation spaces are factored by independent objects, allowing the belief space to scale linearly instead of exponentially in the number of objects. We address the challenges of computational complexity in solving 3D-MOS by developing several techniques that converge to an online multi-resolution planning algorithm. First, we propose a per-voxel observation model which drastically reduces the size of the observation space necessary for planning. Next, we present a novel octree-based belief representation that captures beliefs at different resolutions and allows efficient and exact belief updates. Then, we exploit the octree structure and derive abstractions of the ground problem at different resolution levels leveraging abstraction theory in POMDPs [11, 12]. Finally, a Monte-Carlo Tree Search (MCTS) based online planning algorithm, called Partially-Observable Upper Confidence bounds for Trees (POUCT) [8], is employed to solve these abstract instances in parallel, and the action with highest value in its MCTS tree is selected for execution.

We evaluate the proposed approach in a simulated, discretized 3D domain where a robot with a 6 degrees-of-freedom camera searches for objects of different shapes and sizes randomly generated and placed in a grid environment.

[†]Brown University, Providence, RI. [‡]MIT CSAIL, Cambridge, MA.
Email: {kzheng10, gdk, stefie10}@cs.brown.edu, yooncs8@csail.mit.edu

The results show that the problem scales, our approach

outperforms exhaustive search as well as POMDP baselines without resolution hierarchy under the same computational requirement. We also show that our method is more robust to sensor uncertainty against the POMDP baselines. Finally, we demonstrate our approach on a torso-actuated mobile robot in a lab environment (Figure 6). The robot finds 3 out of 6 objects placed at different heights in two 2D regions in around 15 minutes.

II. BACKGROUND

POMDPs compactly represent the robot's uncertainty in target locations and its own sensor [13]. OO-POMDPs factor the domain in terms of objects, which fits the object search problem naturally [6]. Below, we first provide a brief overview of POMDPs and OO-POMDPs, then we discuss related work in object search.

A. POMDPs and OO-POMDPs

A POMDP models a sequential decision making problem where the environment state is not fully observable by the agent. It is formally defined as a tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{O}, T, O, R, \gamma \rangle$, where $\mathcal{S}, \mathcal{A}, \mathcal{O}$ denote the state, action and observation spaces and the functions $T(s', a, s) = \Pr(s' \clubsuit a | s)$, $O(s', a, o) = \Pr(o \clubsuit a | s')$, and $R(s, a) \in \mathbb{R}$ denote the transition, observation and reward models. The agent takes an action $a \in \mathcal{A}$ that causes the environment state to transition from $s \in \mathcal{S}$ to $s' \in \mathcal{S}$. The environment in turn returns the agent an observation $o \in \mathcal{O}$ and reward $r \in \mathbb{R}$. A history $h_t = (ao)_{1:t-1}$ captures all past actions and observations. The agent maintains a distribution over states given current history $b_t(s) = \Pr(s \clubsuit h_t)$. The agent updates its belief after taking an action and receiving an observation by $b_{t+1}(s') = \eta \Pr(o \clubsuit a) \sum_s \Pr(s' \clubsuit a) b_t(s)$ where $\eta = \sum_s \sum_{s'} \Pr(o \clubsuit a) \Pr(s' \clubsuit a) b_t(s)$ is the normalizing constant. The task of the agent is to find a policy $\pi(b_t) \in \mathcal{A}$ which maximizes the expectation of future discounted rewards $V^\pi(b_t) = \mathbb{E} \sum_{k=0}^{\infty} \gamma^k R(s_{t+k}, \pi(b_{t+k})) \clubsuit b_t$ with a discount factor γ .

An Object-Oriented POMDP (OO-POMDP) (generalization of OO-MDP [14]) is a POMDP that considers the grid cell in the environment state and observation spaces to be factored by a set of attributes of objects where each belongs to a class with a set of attributes. A simplifying assumption is made for the 2D MOS domain that objects are independent so that the belief space scales linearly rather than exponentially in the number of objects. $b_t(s) = \prod_i b_t^i(s_i)$. We make this assumption for the same computational reason.

Offline POMDP solvers are often too slow to be practical for large domains [15]. State-of-the-art online POMDP solvers leverage sparse belief sampling and MCTS to scale up to domains with large state spaces and to address the curse of history [8, 16, 9]. POMCP [8] is one such algorithm that combines particle belief representation with Partially Observable UCT (POUCT) which extends the UCT algorithm [17] to POMDPs and is proved to be asymptotically optimal [8]. We build upon POUCT due to its optimality and simplicity of implementation.

B. Related Work

Previous work primarily address the computational complexity of object search by hypothesizing likely regions based on object co-occurrence [18], semantic knowledge [2] or language [6] reducing the state space from 3D to 2D [6, 19, 20, 21], or constrain the sensor to be stationary [5, 2]. Our work focuses on multi-object search within a 3D region where the robot actively moves the mounted camera, for example through pan or tilt, by moving itself.

Several works explicitly reason over the arrangement of occluded objects based on given geometry models of clutter [3, 21, 23]. Our approach considers occlusion as part of the observation that contains no information about target locations and we do not require geometry models. Many works formulate object search as a POMDP. Notably, Aydemir et al. [2] first infer a room to search in then perform search by calculating candidate viewpoints in a 2D plane. Li et al. [7] plan sensor movements online, assume objects are placed at the same surface level in a container with partial occlusion. Xiao et al. [3] address object fetching on a cluttered tabletop where the robot's FOV fully covers the scene and that occluding obstacles are removed permanently during search. Wandze et al. [6] formulates the multi-object search (MOS) task on a 2D map using the proposed Object-Oriented POMDP (OO-POMDP). We extend that work to 3D and tackle additional challenges by proposing a new observation model and belief representation and a multi-resolution planning algorithm. In addition, our POMDP formulation allows fully occluded objects and can be in principle applied on different robots such as mobile robots or drones.

III. MULTI-OBJECT SEARCH IN 3D

The robots tasked to search for static target objects (e.g. cup and book) of known type but unknown location in a search space that also contains static non-target obstacles. We assume the robot has access to detectors for the objects that it is searching for. The search region is a 3D grid map

environment denoted by \mathcal{G} . Let $g \in \mathcal{G} \subseteq \mathbb{R}^3$ be a 3D cell. We use G^l to denote a grid resolution level $l \in \mathbb{N}$, and $g^l \in G^l$ to denote a grid cell at this resolution level. When omitted, it is assumed that the domain resolution level is $l = 0$. We introduce the 3D-MOS domain that objects are independent so that the belief space scales as follows:

State space. An environment state $s \in \mathcal{S}$ is factored in an object-oriented way as $s = (s_r, s_o)$ where $s_r \in \mathcal{S}_r$ is the state of the robot and $s_o \in \mathcal{S}_o$ is the state of target objects.

A robot state is defined as $(p, \mathcal{F}) \in \mathcal{S}_r$ where p is the 6D camera pose and \mathcal{F} is the set of found objects. The robot state is assumed to be observable to the robot. We consider the object state to be specified by one attribute of the object, its center of mass c corresponding to a point in the grid G . We denote a state $s_i^l \in \mathcal{S}_i^l$ to be an object state at resolution level l where $\mathcal{S}_i^l = G^l$.

Observation space. The robot receives an observation o through a viewing frustum projected from a mounted camera. The viewing frustum forms the FOV of the robot.

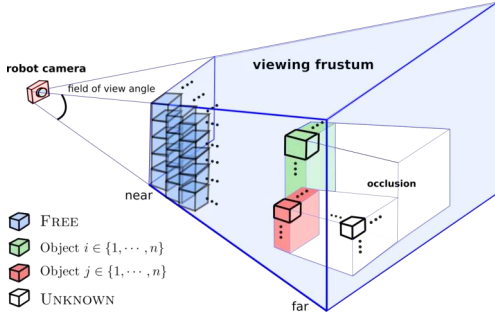


Fig. 2: Illustration of the viewing frustum and volumetric observation. The viewing frustum V consists of voxels where each $v \in V$ can be labeled as $i \in \{1, \dots, n\}$, FREE, or UNKNOWN.

by V , which consists of voxels. Note that the resolution of a voxel should be no lower than that of a 3D grid cell g . We assume both resolutions to be the same in this paper. For notational convenience, let $V \subseteq G$, but in general a voxel with higher resolution can be easily mapped to the corresponding grid cell.

For each voxel $v \in V$, a *detection function* $d(v)$ labels the voxel to be either an object $i \in \{1, \dots, n\}$, FREE, or UNKNOWN (Figure 2). FREE denotes that the voxel is a free space or an obstacle. We include the label to take into account occlusion incurred by target objects or static obstacles. In this case, the corresponding voxel does not give any information about the environment. An observation $o = \{(v, d(v)) \mid v \in V\}$ is defined as a set of voxel-label tuples. This can be thought of as the result of voxelization and object segmentation given a raw point cloud.

We can factor by objects in the following way. First, given the robot state s_r at which o is received, the voxels in V have known locations. Under this condition, V can be reduced to exclude voxels labeled UNKNOWN while still maintaining the same information. Then, V can be decomposed by objects into $V = \bigcup_{i=1}^n V_i$ where for any $v \in V_i$, $d(v) \in \{i, \text{FREE}\}$ which retain the same information as o . For a given robot state s_r , the observation can be written as $o_i = \{(v, d(v)) \mid v \in V_i\}$.

Action space. Searching for objects generally requires three basic capabilities: *moving*, *looking*, and *declaring* an object to be found at some location. Formally, the action space consists of these three types of primitive actions: $\text{MOVE}(s_r, g)$ action moves the robot from pose in s_r to destination $g \in G$ stochastically, $\text{LOOK}(\theta)$ changes the camera pose to look in the direction specified by θ , and projects a viewing frustum, and $\text{FIND}(i, g)$ declares object i to be found at location g . The implementation of these actions may vary depending on the type of search space of the robot. Note that this formulation allows macro actions, as “look after move” to be composed for planning.

Transition function. Target objects and obstacles are static objects, thus $\Pr(s_i \mid s_{i-1}, a) = 1(s_i = s_{i-1})$. For the robot, the actions $\text{MOVE}(s_r, g)$ and $\text{LOOK}(\theta)$ change the camera location and direction and the following a domain-specific

stochastic dynamics function. The action $\text{FIND}(i, g)$ adds to the set of found objects in the robot state only if g is within the FOV determined by θ .

Reward function. The correctness of declarations can only be determined by the target object or additional interactions with the object, therefore we considered declarations to be expensive. The robot receives $R_{\max} \gg 0$ if an object is correctly identified by an action, otherwise the FIND action incurs a $R_{\min} \ll 0$ penalty. MOVE and LOOK receive a negative step cost $c_{\text{step}} < 0$ dependent on the robot state and the action itself. This is a sparse reward function.

A. Observation Model

We have previously described how a volumetric observation can be factored by objects into $o = \bigcup_{i=1}^n o_i$. Here, we describe a method to model $\Pr(o_i \mid a)$, the probabilistic distribution over an observation object.

Modeling a distribution over a 3D volume is a challenging problem [24]. To develop an efficient model, we make the simplifying assumption that an object is contained within a single voxel located at the grid cell $g = s_i$. We address the case of searching for objects of unknown sizes with our planning algorithm (Section IV). It plans at multiple resolutions in parallel.

Under this assumption, $d(s_i) = \text{FREE}$ deterministically for s_i and the uncertainty is reduced to the uncertainty of $d(s_i)$. As a result, $\Pr(o_i \mid a)$ can be simplified to $\Pr(d(s_i) \mid a)$. When $s_i \notin V_i$, either $d(s_i) = \text{UNKNOWN}$ (occlusion) or $s_i \notin V$ (not in FOV). In this case, there is no information regarding the value of $d(s_i)$. When $s_i \in V_i$, that is, the non-occluded region within the FOV covers s_i , the case of $d(s_i) = i$ indicates correct detection while $d(s_i) = \text{FREE}$ indicates sensing error. We let $\Pr(d(s_i) = i \mid a) = \alpha$ and $\Pr(d(s_i) = \text{FREE} \mid a) = \beta$. It should be noted that α and β do not necessarily sum to one because the belief update equation does not require the observation model to be normalized as explained in Section II-A. Thus, hyperparameters α and β independently control the reliability of the observation model.

IV. OCTREE BELIEF REPRESENTATION

Particle belief representation [6, 8] suffers from particle depletion under large observation spaces. Moreover, if the resolution of o is dense, it may be possible that most of 3D grid cells do not contribute to the behavior of the robot.

We represent a belief state for objects as an octree, referred to as an *octree belief*. It can be constructed incrementally as observations are received and it tracks the belief of object state at different resolution levels. Furthermore, it allows efficient belief sampling and belief update using the per-voxel observation model (ISA).

An octree belief consists of an octree and a normalizer. An octree is a tree where every node has at most 8 children. In our context, a node represents a grid cell where the resolution level, such that it covers a cubic volume of ground-level grid cells, the ground resolution level given

¹The FOV V is fixed for a given camera pose in the robot state, therefore excluding UNKNOWN voxels does not lose information.

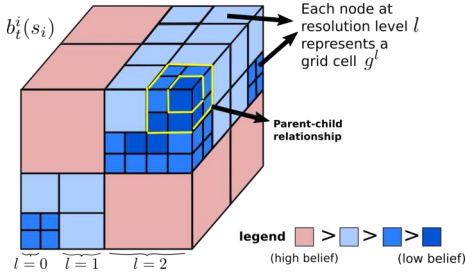


Fig. 3: Illustration the octree belief representation. The color on a node g indicates the belief $\text{VAL}_t^i(g)$ that the objects located within. The highlighted grid cells indicate parent-child relationship between a grid cell at resolution level $l = 1$ (parent) and one at level $l = 0$.

by $l = 0$. The 8 children of the node equally subdivide the volume a^l into smaller volumes at resolution level

$l - 1$ (Figure 3). Each node stores a value $\text{VAL}_t^i(g^l) \in \mathbb{R}$, which represents the unnormalized belief at $s_i^l = g^l$, that is, object i is located at grid cell g^l . We denote the set of nodes at resolution level $< l$ that reside in a subtree rooted at g^l by $\text{CH}^k(g^l)$. By definition $b_t^i(g^l) = \Pr(g^l \clubsuit_t) = \sum_{c \in \text{CH}^k(g^l)} \Pr(c \clubsuit_t)$. Thus, with a normalizer $\text{NORM}_t = \sum_{g \in G} \text{VAL}_t^i(g)$, we can rewrite the normalized belief as:

$$b_t^i(g^l) = \frac{\text{VAL}_t^i(g^l)}{\text{NORM}_t} = \sum_{c \in \text{CH}^k(g^l)} \frac{\text{VAL}_t^i(c)}{\text{NORM}_t} \quad (1)$$

which means $\text{VAL}_t^i(g^l) = \sum_{c \in \text{CH}^k(g^l)} \text{VAL}_t^i(c)$. In words, the value stored in a node is the sum of values stored in its children. The normalizer equals to the sum of values stored in the nodes at the ground resolution level.

The octree does not need to be constructed fully in order to query the probability at any grid cell. This can be achieved by setting a default value $\text{VAL}_0^i(g) = 1$ for all ground grid cells $g \in G$ not yet present in the octree. Then, any node corresponding to g has a default value of $\text{VAL}_0^i(g) = \sum_{c \in \text{CH}^1(g^1)} \text{VAL}_0^i(c) = \clubsuit \text{CH}^1(g^1) \clubsuit$.

A. Belief Update

We have defined a per-voxel observation model for $\Pr(o_i \clubsuit a)$, which is reduced to $\Pr(d(s_i) \clubsuit a)$ if $s_i \in V_i$, or a uniform distribution \clubsuit if $s_i \notin V_i$. This suggests that the belief update need only happen for voxels that are FOV to reflect the information in the observation.

Upon receiving observation within the FOV V_i , belief is updated according to Algorithm 1. This algorithm updates the value of the ground-level node corresponding to each voxel $v \in V_i$ as $\text{VAL}_{t+1}^i(g) = \Pr(d(v) \clubsuit a) \text{VAL}_t^i(g)$. The normalizer is updated to make sure normalized

Lemma 1 The normalizer NORM_t at time t can be correctly updated by adding the incremental update of $\text{VAL}_{t+1}^i(g)$ in Algorithm 1.

Proof: The normalizer must equal to the sum of node values at the ground level for the next belief valid (Equation 1). That is, $\text{NORM}_{t+1} = \sum_{s_i \in G} \text{VAL}_{t+1}^i(s_i)$. This sum can be decomposed into two cases where s_i is inside of V_i and outside of V_i . For object locations

Algorithm 1: OctreeBeliefUpdate(b_t^i, o_i) $\rightarrow b_{t+1}^i$

```

input :  $b_t^i$ : octree belief for object action taken
        by robot,  $\clubsuit = \Pr(d(v) \clubsuit a) \in V_i \diamond$  factored
        observation for object
output :  $b_{t+1}^i$ : updated octree belief
// Let  $\Psi(b)$  denote the octree underlying  $b$ 
for  $v \in V_i$  do
     $s_i \leftarrow v$ ; // State at grid cell corresponding to voxel  $v$ 
    if  $s_i \notin \Psi(b_t^i)$  then
        Insert node  $a$  to  $\Psi(b_t^i)$ ;
    end
     $\text{VAL}_{t+1}^i(s_i) \leftarrow \Pr(d(v) \clubsuit a) \text{VAL}_t^i(s_i)$ ;
     $\text{NORM}_{t+1} \leftarrow \text{NORM}_t + \text{VAL}_{t+1}^i(s_i) - \text{VAL}_t^i(s_i)$ ;
end

```

at $s_i \in V_i$, the *unnormalized* observation model is uniform, thus $\text{VAL}_{t+1}^i(s_i) = \Pr(d(s_i) \clubsuit a) \text{VAL}_t^i(s_i) = \text{VAL}_t^i(s_i)$. Therefore, $\text{NORM}_{t+1} = \sum_{s_i \in V_i} \text{VAL}_{t+1}^i(s_i) + \sum_{s_i \notin V_i} \text{VAL}_t^i(s_i)$. Note the set $\clubsuit \in V_i \diamond$ is equivalent to $\clubsuit \in G \cap V_i \diamond$. Using this fact and the definition of NORM_t , we obtain $\text{NORM}_{t+1} = \text{NORM}_t + \sum_{s_i \in V_i} \text{VAL}_{t+1}^i(s_i) - \text{VAL}_t^i(s_i)$ which proves the lemma. ■

This belief update is therefore exact since the objects are static. The complexity of this algorithm is $\mathcal{O}(\clubsuit \log(\clubsuit))$. Inserting nodes and updating values of nodes can be done by traversing the tree depth-wise.

B. Belief Sampling

Octree belief affords exact belief sampling at any resolution level in logarithmic time complexity with respect to the size of the search space \clubsuit , despite not being completely built. Given resolution level l , we sample from S_i^l by traversing the octree in a depth-first manner until we reach the maximum resolution level for the search space. Let s_i^{des} be the *desired* resolution level at which a state is sampled. If s_i^{des} is sampled then all nodes in the octree that have s_i^{des} , i.e. $s_i^{\text{max}} \leq s_i \leq s_i^{\text{des}} + 2_{cs}^{\text{des}} + 1$, must also be implicitly sampled. Also, the event that is sampled is independent from other samples given that is sampled. Hence, the task of sampling s_i^{des} is translated into sampling a sequence of samples $s_i^{\text{max}} \leq s_i \leq s_i^{\text{des}} + 2_{cs}^{\text{des}} + 1$, each according to the distribution $\Pr(s_i \clubsuit^{l+1} h_t) = \frac{\text{VAL}_t^i(s_i^l)}{\text{VAL}_t^i(s_i^{l+1})}$. Sampling from this probability distribution is efficient, as the sample space of the children of nodes s_i^{l+1} is only of size 8. Therefore, this sampling scheme yields a sample s_i^{des} according to $\Pr(s_i^{\text{des}})$ with time complexity $\mathcal{O}(\log(\clubsuit))$.

V. MULTI-RESOLUTION PLANNING VIA ABSTRACTIONS

POUCT expands an MCTS tree using a *generative function* $(s' \circ r) \sim \mathcal{G}(s \circ a)$, which is straightforward to acquire since we explicitly define the 3D-MOS model. However, directly applying POUCT is subject to high branching factor due to the large observation space in our domain. Our intuition is that tree belief imposes a spatial abstraction which can be used to derive an abstraction over

observations reducing the branching factor for planning. Below, we formulate an *abstract 3D-MOS* with smaller spaces, and propose our multi-resolution planning algorithm.

A. Abstract 3D-MOS

We adopt the abstraction scheme in [11] where in general the abstract transition and reward functions are weighted sum of the original problem's transition and reward functions, respectively with weights that sum up to 1. We define an abstract 3D-MOS $\langle \hat{S}, \hat{A}, \hat{O}, \hat{T}, \hat{O}, R, \gamma, l \rangle$ at resolution level \hat{d} as follows.

State space \hat{S} . For each object, an abstraction function $\phi_i : S_i \rightarrow \hat{S}_i$ transforms the ground-level object state to an abstract object state at resolution level \hat{d} . The abstraction of the full state is $\hat{s} = \phi(s) = \phi_r \diamond \bigcup_i \phi_i(s_i)$ where the robot state is kept as is. The *inverse image* (ϕ_i^{-1}) is the set of ground states that correspond to ϕ_i [11].

Action space \hat{A} . Since state abstraction lowers the resolution of the search space, consider macro move actions that move the robot over longer distance at each planning step. Each macro move action $\text{MOP}(s_r \diamond g)$ is an *option* [25] that moves to goal location using multiple primitive actions. The primitive `GO` and `FIND` actions are kept.

Transition function \hat{T} . Targets and obstacles are still static and the robot state still transitions according to the ground-level transition function. However, the transition of the found set from \mathcal{F} to \mathcal{F}' is special since the action $\text{FIND}(i \diamond g)$ operates at the ground level while a lower resolution $\hat{d} > 0$. Let f_i be the binary state variable that is true if and only if object $i \in \mathcal{F}$. Because the action $\text{FIND}(i \diamond g)$ affects f_i based only on whether object i is located at g , and that the problem is no longer Markovian due to state abstraction, the transitions to following

$$\begin{aligned} \Pr(f_i \clubsuit s_i \clubsuit h_t \diamond \text{FIND}(i \diamond g)) \\ = \sum_{s_i \in \phi_i^{-1}(s_i^!)} \Pr(f_i \clubsuit f_i \diamond \text{FIND}(i \diamond g)) \Pr(s_i \clubsuit h_t) \end{aligned} \quad (2)$$

The above is consistent with the abstract transition function in the works [11, 12] where the first term corresponds to the ground-level deterministic transition function and the second term $\Pr(s_i \clubsuit h_t)$, stored in the octree belief, is the weight that sums up to 1 for all $s_i \in S_i$.

Observation space \hat{O} and function \hat{O} . For the purpose of planning, we again use the assumption that objects contained within a single voxel at resolution level \hat{d} . Then, given state, the abstract observation is regarded as a voxel-label pair $(s_i^! \diamond d(s_i^!))$. Since it is computationally expensive to sum all object states, we approximate the observation model by ignoring objects other than

$$\Pr(o_i \clubsuit a \clubsuit h_t) = \Pr(d(s_i^!) \clubsuit a \clubsuit h_t) \quad (4)$$

$$\approx \Pr(d(s_i^!) \clubsuit s_r \diamond a \clubsuit h_t) \quad (5)$$

$$= \sum_{s_i \in \phi_i^{-1}(s_i^!)} \Pr(d(s_i^!) \clubsuit s_r \diamond a) \Pr(s_i \clubsuit h_t) \quad (6)$$

This resembles the abstract transition function, where $\Pr(d(s_i^!) \clubsuit s_r \diamond a)$ is the ground observation function, and

Algorithm 2: MR-POUCT($\mathcal{P} \diamond b_t \diamond d$) $\rightarrow \hat{a}$

input : \mathcal{P} : a set of abstract 3D-MOS instances at different resolution levels
 t ; d : planning depth
output: \hat{a} : an action in the action space of some $P_1 \in \mathcal{P}$

Procedure Plan(\mathcal{P})
 foreach $P_1 \in \mathcal{P}$ in parallel do
 // Recall that $P = \langle \hat{S}, \hat{A}, \hat{O}, \hat{T}, \hat{O}, R, \gamma, l \rangle$
 $G \leftarrow \text{GenerativeFunction}(P)$
 $Q_P(b_t \diamond \hat{a}) \leftarrow \text{POUCT}(G \diamond h_t \diamond d)$
 end
 $\hat{a} \leftarrow \text{argmax}_{\hat{a}} \{Q_P(b_t \diamond \hat{a}) \mid P \in \mathcal{P}\}$
 return \hat{a}

$\Pr(s_i \clubsuit h_t)$ is again the weight.

For practical POMDP planning, it can be inefficient to sample from this abstract observation model if \hat{d} is large. In our implementation, we approximate this distribution by Monte Carlo sampling [26]: We sample k states from $\phi_i^{-1}(s_i^!)$ according to their weights. If we set $d(s_i^!) = i$ if the majority of these samples have i , and $d(s_i^!) = \text{FREE}$ otherwise. A similar approach is used for sampling from the abstract transition model.

Reward function R . The reward function is the same as the one in ground 3D-MOS, since computing the reward only depends on the robot state which is not abstracted and the abstract action space consists of the same primitive actions as 3D-MOS. Therefore, solving an abstract 3D-MOS is solving the same task as the original 3D-MOS.

B. Multi-Resolution Planning Algorithm

Abstract 3D-MOS is smaller than the original 3D-MOS which may provide benefit for online planning. However, it may be difficult to define a single resolution level, due to the uncertainty of the size or shape of objects, and the unknown distance between the robot and these objects.

Therefore, we propose to solve a number of abstract 3D-MOS problems in parallel, and select an action with the highest value for execution. The algorithm is formally presented in Algorithm 2.

The set of abstract 3D-MOS problems \mathcal{P} , can be defined based on the dimensionality of the search space and the particular object setting. Then, it is straightforward to define a *generative function* $G(\hat{s} \diamond \hat{a}) \rightarrow (\hat{s}' \diamond \hat{o} \diamond r)$ from an abstract 3D-MOS instance P using its transition, observation and reward functions.

POUCT uses G to build a search tree and plan the next action. Thus, all problems in \mathcal{P} are solved online in parallel, each by a separate POUCT. The final action with the highest value $Q_P(b_t \diamond \hat{a})$ in its respective POUCT search tree is chosen as the output (see [8] for details on POUCT). We call this algorithm Multi-Resolution POUCT (MR-POUCT).

²We tested $k = 10$ and $k = 40$ and observed similar search performance. We used $k = 10$ in our experiments.

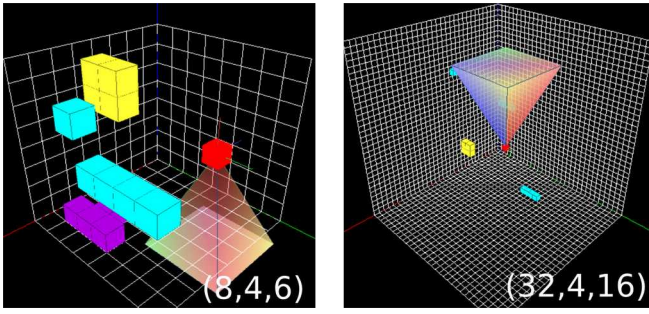


Fig. 4: Simulated environment for 3D object search. The robot (represented as a red cube) can project a viewing frustum to observe the search space where objects are represented by sets of cubes. Search space size scales from 4^3 to 32^3 . The tuple (m, n, d) at lower-right defines the problem instance.

VI. EXPERIMENTS

We assess the hypothesis that our approach, MR-POUCT, improves the robot’s ability to efficiently and successfully find objects especially in large search spaces. We conduct a simulation evaluation and a study on a real robot.

A. Simulation Evaluation

Setup. We implement our approach in a simulated environment designed to reflect the essence of the 3D-MOS domain (Figure 4). Each simulated problem instance is defined by a tuple (m, n, d) , where the search region has size $C = m^3$ with n randomly generated, randomly placed objects. The on-board camera projects a viewing frustum with 45 degree FOV angle, 1.0 aspect ratio, a minimum range of 1 grid cell, and a maximum range of grid cells. Hence, we can increase the difficulty of the problem by increasing m and n , or by reducing the percentage of voxels covered by a viewing frustum through reducing the FOV range. Occlusion is simulated using perspective projection and treating each grid cell as a point.

There are two primitive actions per axis (e.g., $+x$, $-x$) that each moves the robot along that axis by one grid cell. There are two LOOK actions per axis, one for each direction. Finally, a FIND action is defined that declares all not-yet-found objects within the viewing frustum as found. Thus, the total number of primitive actions MOVE and LOOK actions have a step cost of 1. A successful FIND receives +1000 while a failed attempt receives 0. A FIND action is successful if part of a new object lies within the viewing frustum. If multiple new objects are present within one viewing frustum when the FIND is taken, only the maximum reward of 1000 is received. The task terminates either when the total planning time limit is reached or FIND actions are taken.

Baselines. We compare our approach (MR-POUCT) with the following baselines. POUCT uses the octree belief but solves the ground POMDP directly using the original POUCT algorithm. Options+POUCT uses the octree belief and a resolution hierarchy, but only the motion action abstraction (i.e., MOVE options) is used, meaning that the agent can move for longer distances per planning step but

not make use of state and observation abstractions. POMCP uses a particle belief representation which is subject to particle deprivation. A uniform random rollout policy is used for all POMDP-based methods. Exhaustive uses a hand-coded exhaustive policy where the agent traverses every location in the search environment. At every location, the agent takes a sequence of $60k$ actions, one in each direction. Finally, Random executes actions at uniformly at random.

Each algorithm begins with uniform prior and is allowed a maximum of 3.0s for planning each step. The total amount of allowed planning time plus time spent on belief update is 120s, 240s, 360s, and 480s for environment sizes of $(4, 8, 16, \text{or } 32)$, respectively. Belief update is not necessary for Exhaustive and Random. The maximum number of planning steps is 500. The discount factor is set to 0.99. For each (m, n, d) setting, 40 trials (with random world generation) are conducted.

Results. We evaluate the scalability of our approach with 4 different settings of search space size $\{4^3, 8^3, 16^3, 32^3\}$ and 3 settings of number of objects $\{2, 4, 6\}$, resulting in 12 combinations. The FOV range is chosen such that the percentage of the grids covered by one projection of the viewing frustum decreases as the world size increases. The sensor is assumed to be near-perfect, with $\beta = 0$. We measure the discounted cumulative reward, which reflects both the search efficiency and effectiveness, as the number of objects found per trial.

Results are shown in Figure 5. Particle deprivation happens quickly due to large observation space, the behavior degenerates to a random guessing POMCP to perform poorly. In small-scale domains, the Exhaustive approach works well, outperforming the POMDP-based methods. We find that in those environments, the FOV can capture a significant proportion of the environment, making exhaustive search desirable. The POMDP-based approaches are competitive or better in the two largest search environments (16^3 and $m = 32$). In particular, MR-POUCT outperforms Exhaustive in all test cases in the large environments, with greater margin in discounted cumulative reward. Exhaustive takes more search steps, but it is less efficient. When the search space contains fewer objects, MR-POUCT and POUCT show more resilience than Options+POUCT. With MR-POUCT performing consistently better. This demonstrates the benefit of planning with the resolution hierarchy in octree belief, especially in large search environments.

We then investigate the performance of our method with respect to changes in sensing uncertainty, controlled by the parameters α and β of the observation model. According to the belief update algorithm in Section IV-A, a noisy but functional sensor should increase the belief $bel_t(g)$ for object g if an observed voxel x is labeled, while decrease the belief if labeled FREE. This implies that a properly working sensor should satisfy $\beta < 1$. We investigate on 5 settings of $\beta \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$ and 2 settings of $\alpha \in \{0.3, 0.8\}$. A fixed problem difficulty

³The maximum FOV coverage for $m = 4, 8, 16$, and 32 is 17.2% ($d = 6$), 4.7% ($d = 10$), and 2.6% ($d = 16$), respectively.

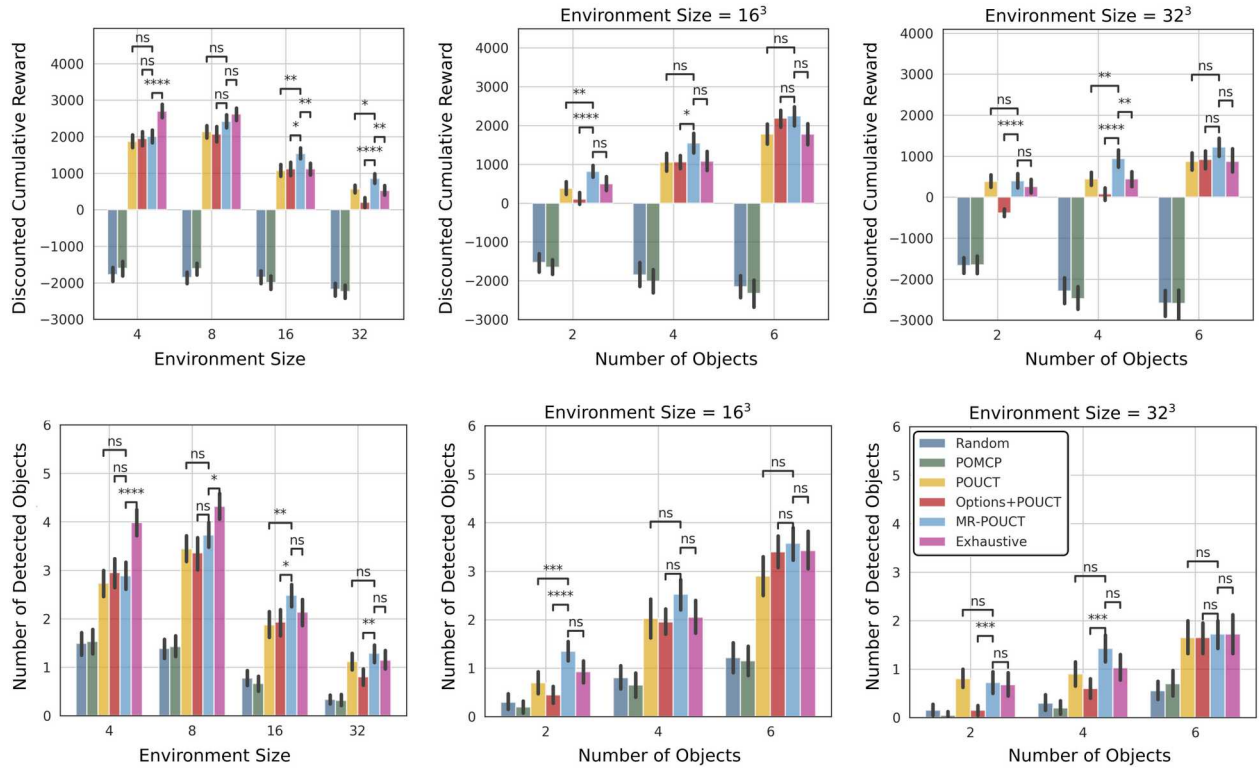


Fig. 5: Discounted cumulative reward and number of detected objects as the environment size increases and as the number of objects (n) increases. Exhaustive search performs well in small-scale environments (4 and 8) where the exploration strategy is taken advantage of. In large environments, our method MR-POUCT performs better than the baselines in most cases. The error bars are 95% confidence intervals. The level of statistical significance is shown by comparing MR-POUCT against POUCT, Options+POUCT and Exhaustive respectively, indicated by ns ($p > 0.05$), * ($p \leq 0.05$), ** ($p \leq 0.01$), *** ($p \leq 0.001$), **** ($p \leq 0.0001$).

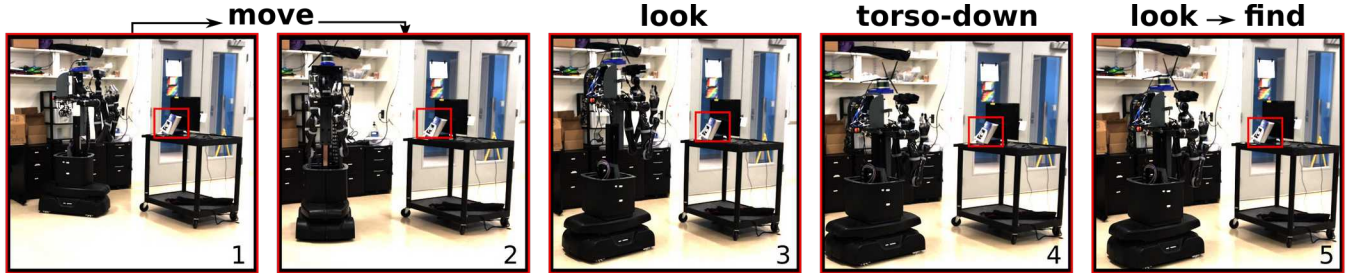


Fig. 6: Example action sequence produced by the proposed approach. The mobile robot first navigates in front of a portable table, then takes a look action to observe the space in front of it, and (3) the target is observed since the torso is lowered. The robot then decides to lower its torso (takes another look action in the same direction), and finally to mark the object as found (5). This sequence of actions demonstrate that our algorithm can produce efficient search strategies in real world scenarios.

of $(16 \times 2 \times 10)$ is used to conduct this experiment. Results in Figure 7 show that MR-POUCT is consistently better than other algorithms in a lab environment which is decomposed into two search regions G_1 and G_2 of size roughly 10×2 m. The robot operates in a lab environment which is decomposed into two search regions G_1 and G_2 of size roughly 10×2 m. The robot is tasked to look for n_{G_1} and n_{G_2} objects in each search region sequentially, where objects are represented by paper AR tags that could be in clutter or not detectable at an angle. The robot instantiates an instance of the 3D-MOS problem over this 3D-MOS implementation, where actions are implemented based on a topological map of the environment. The neighbors of the robot are represented by a metric occupancy grid map.

B. Demonstration on a Torso-Actuated Mobile Robot

We demonstrate that our approach is scalable to real world settings by implementing the 3D-MOS problem as well as MR-POUCT for a mobile robot setting. We use the Kinova MOVO Mobile Manipulator robot, which has an actuated torso with an extension range between around 0.05m

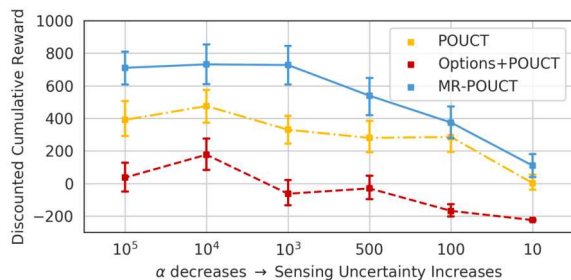


Fig. 7: Discounted cumulative reward with 95% confidence interval as the sensing uncertainty increases, aggregating over the β settings.

graph node form the motion action space when the robot is at that node. The robot can take 4 action in 4 cardinal directions in place and receive volumetric observations. A volumetric observation is a result of downsampling and thresholding points in the corresponding cloud. The robot was able to find 3 out of 6 total objects in the two search regions in around 15 minutes. One sequence of actions (Figure 6) shows that the robot decides to lower its torso order to look and find an object. A failure mode is that the object may not be covered by any viewpoint and thus not detected; this can be improved with a denser topological map, or by considering destination MOVE actions sampled from the continuous search region.

VII. CONCLUSION

We present a POMDP formulation of multi-object search in 3D with volumetric observation space and solve it with a novel multi-resolution planning algorithm. Our evaluation demonstrates that challenging POMDPs can be solved online efficiently and scalably with practicality by real robot by extending existing general POMDP solvers with domain-specific structure and belief representation.

One limitation of the presented work is that the assumption of object independence, though beneficial computationally, may discard useful object dependence information in some cases. Optimal search for correlated objects becomes important. In addition, we do not explicitly reason over object geometry in the observation model. Considering belief over geometrical appearances is a challenging future direction. Finally, incorporating a heuristic rollout policy may be a promising direction for more realistic object search problems while sacrificing optimality.

ACKNOWLEDGEMENTS

The authors thank Selena Ling for help with the simulator. This work was supported by the National Science Foundation under grant number IIS1652561, ONR under grant number N00014-17-1-5806, the US Army under grant number W911NF1920145, Intel Labs, STRAC Institute, and Hyundai.

REFERENCES

[1] T. Kollar and N. Roy, "Utilizing object-object and object-scene context when planning to find things," in *IEEE International Conference on Robotics and Automation*, IEEE, 2009, pp. 2168–2173.

⁴Video footage with visualization of volumetric observations and belief update is available at <https://zkytony.github.io/3D-MOS/>.

[2] A. Aydemir, A. Pronobis, M. Göbel, and P. Jensfelt, "Active visual object search in unknown environments using uncertain semantics," *IEEE Transactions on Robotics (T-RO)*, vol. 29, no. 4, pp. 986–1002, Aug. 2013.

[3] Y. Xiao, S. Katt, A. ten Pas, S. Chen, and C. Amato, "Online planning for target object search in clutter under partial observability," in *Proceedings of the International Conference on Robotics and Automation*, 2019.

[4] N. Atanasov, B. Sankaranarayanan, L. Ny, G. Pappas, and K. Daniilidis, "Nonmyopic view planning for active object classification and pose estimation," *IEEE Transactions on Robotics (TRO)*, 2014.

[5] M. Danielczuk, A. Kurenkov, A. Balakrishna, M. Matl, D. Wang, R. Mann-Martin, A. Garg, S. Savarese, and K. Goldberg, "Mechanical search: Multi-step retrieval of a target object occluded by clutter," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 1614–1621.

[6] A. Wandzely, A. Wang, M. Fishman, N. Kumar, and S. Tellex, "Multi-Object Search using Object-Oriented POMDPs," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019.

[7] J. K. Li, D. Hsu, and W. S. Lee, "Act to see and see to act: POMDP planning for objects search in clutter," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016.

[8] D. Silver and J. Veness, "Monte-Carlo planning in large POMDPs," in *Advances in neural information processing systems*, 2010.

[9] Z. N. Sunberg and M. J. Kochenderfer, "Online algorithms for POMDPs with continuous state and observation spaces," in *Twenty-Eighth International Conference on Automated Planning and Scheduling*, 2018.

[10] N. P. Garg, D. Hsu, and W. S. Lee, "DESPOT- α : Online POMDP planning with large state and observation spaces," in *Robotics: Science and Systems*, 2019.

[11] L. Li, T. J. Walsh, and M. L. Littman, "Towards a unified theory of state abstraction for MDPs," in *IS2016*, 2016.

[12] A. Bai, S. Srivastava, and S. J. Russell, "Markovian state and action abstractions for MDPs via hierarchical MCTS," in *ICML*, 2016.

[13] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial intelligence*, vol. 101, no. 1-2, pp. 99–134, 1998.

[14] C. Diuk, A. Cohen, and M. L. Littman, "An object-oriented representation for efficient reinforcement learning," in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 240–247.

[15] S. Ross, J. Pineau, S. Paquet, and B. Chaib-Draa, "Online planning algorithms for POMDPs," *Journal of Artificial Intelligence Research*, vol. 32, pp. 663–704, 2008.

[16] A. Somani, N. Ye, D. Hsu, and W. S. Lee, "DESPOT: Online POMDP planning with regularization," *Advances in neural information processing systems*, 2013, pp. 1772–1780.

[17] L. Kocsis and Cs. Szepesvári, "Bandit-based Monte-Carlo planning," in *European conference on machine learning*, Springer, 2006.

[18] S. Edelkamp, W. Wixson, and D. H. Ballard, "Using intermediate objects to improve the efficiency of visual search," *International Journal of Computer Vision*, vol. 12, no. 2-3, pp. 209–230, 1994.

[19] C. Wang, J. Cheng, J. Wang, X. Li, and M. Q.-H. Meng, "Efficient object search with belief road map using mobile robot," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3081–3088, 2018.

[20] A. Sarmiento, R. Murrieta, and S. A. Hutchinson, "An efficient strategy for rapidly finding an object in a polygonal world," in *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)* (Cat. No. 03CH37453), 2003.

[21] X. Nie, L. L. Wong, and L. P. Kaelbling, "Searching for physical objects in partially known environments," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016.

[22] M. Pradeep, M. C. Koval, A. Tallavajhula, and S. S. Srinivasa, "Object search by manipulation," *Autonomous Robots*, 2016.

[23] L. L. Wong, L. P. Kaelbling, and T. Lozano-Perez, "Manipulation-based active search for occluded objects," in *2013 IEEE International Conference on Robotics and Automation (ICRA)*, 2013, pp. 2814–2819.

[24] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "DeepSDF: Learning continuous signed distance functions for representation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 165–174.

[25] R. S. Sutton, D. Precup, and S. Singh, "Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning," *Artificial intelligence*, vol. 112, no. 1-2, pp. 181–211, 1999.

[26] A. Shapiro, "Monte Carlo sampling methods," *Handbooks in operations research and management science*, vol. 10, pp. 353–425, 2003.