

# A Hierarchical Approach to Active Pose Estimation

Jascha Hellwig\*, Mark Baierl\*, João Carvalho<sup>†</sup>, Julen Urain<sup>†</sup>, Jan Peters

**Abstract**—Creating mobile robots which are able to find and manipulate objects in large environments is an active topic of research. These robots not only need to be capable of searching for specific objects but also to estimate their poses often relying on environment observations, which is even more difficult in the presence of occlusions. Therefore, to tackle this problem we propose a simple hierarchical approach to estimate the pose of a desired object. An Active Visual Search module operating with RGB images first obtains a rough estimation of the object’s 2D pose, followed by a more computationally expensive Active Pose Estimation module using point cloud data. We empirically show that processing image features to obtain a richer observation speeds up the search and pose estimation computations, in comparison to a binary decision that indicates whether the object is or not in the current image.

## I. INTRODUCTION

A major task of autonomous robots working in manipulation scenarios is to find objects and estimate their poses. While in recent years there have been major developments in pose estimation methods [1]–[4], these assume the object was already found and is present in the viewing frustum. However, the object might be occluded or even out of the field of view. In these situations, the agent might be required to navigate along and interact with the environment to improve its estimation of the object’s pose. We call Active Pose Estimation (APE) to the problem of actively deciding how to interact with the world in order to minimize the uncertainty with respect to the pose of a target object. In most cases, to estimate the pose of an object, we are required to sense depth information by using depth cameras or Lidars. Nevertheless, the computational requirements of dealing with depth information are high. We remark that in APE, we are expected to decide the next action every step, and thus, we aim to have low computational requirements to increase the control frequency.

We call Active Visual Search (AVS) to the problem of actively deciding how to interact with the environment to search and find a particular object. A possible solution to increase the control frequency is to run a search process with RGB information only and then, once the object is found, run a pose estimation method that uses point clouds, such as the Iterative Closest Point Algorithm (ICP) [5]. While this can lead to good performance in terms of pose estimation, it might not be very robust. If the search procedure terminates without a decent view of the object, the pose estimation quality could decrease significantly. To improve robustness, it feels a more

natural solution to apply an active pose estimation process at the expense of higher computation.

With the aim of having a both computationally light and at the same time robust pose estimator; in this paper, we propose a novel algorithm for occluded or out-of-view object pose estimation. We propose a hierarchical approach that combines a computationally light AVS process with a computationally intense APE process. To avoid the high computational budget of estimating an object pose in high-dimensional point clouds, we combine an AVS running on RGB observations with an APE using pointcloud inputs. The search module uses RGB images to efficiently interact with the environment in order to estimate a rough two-dimensional object pose. When the belief in the object’s 2D pose, computed with the search module, is sufficient, we pass that belief to the pose estimation module and actively move to estimate the 6D pose of the object. We frame both process in terms of a POMDP problem and solve it by adapting the Partially Observable Monte-Carlo Planning (POMCP) algorithm [6] to the problem of pose estimation. Framing both process as POMDP permits sharing the computed belief distributions in the AVS process to the APE process, reducing the computational complexity for the second.

In particular, the main contributions of these work are (i) We propose a unified framework that combines a 2D AVS process with a 3D APE process. Due to the unified framework, we present an approach to transfer the beliefs states from the AVS to the APE process. (ii) We propose two approaches to apply POMCP for both 2D active search and 3D active pose estimation. (iii) We make an extensively ablation study to investigate the performance of our proposed method under multiple hyperparameters. We investigate the required control steps until the pose is properly estimated.

## II. RELATED WORK

AVS is commonly framed as solving a Partially Observable Markov Decision Process (POMDP). In general POMDP-based approaches use observations to update the environment map and compute optimal actions using the current map estimate. POMCP-based Online Motion Planning (POMP) [7] knows the environment map before hand and uses the POMCP algorithm [6] to select optimal actions. Unobserved map cells are marked as *candidate* and observed cells marked according to the observation. The agent explores candidate zones until the object is found, which is determined by a binary object detector that can have errors. A robust visual docking mechanism compensates object detection errors, by returning the agent to a known position if the object said to be found was not encountered. POMP++ [8] extends POMP to the

\*<sup>†</sup>Equal contribution.

All authors are with the Intelligent Autonomous Systems, Technische Universität Darmstadt, Germany - {jascha.hellwig, mark.baierl}@stud.tu-darmstadt.de, {joao,julen}@robot-learning.de, jan.peters@tu-darmstadt.de  
Correspondence to joao@robot-learning.de

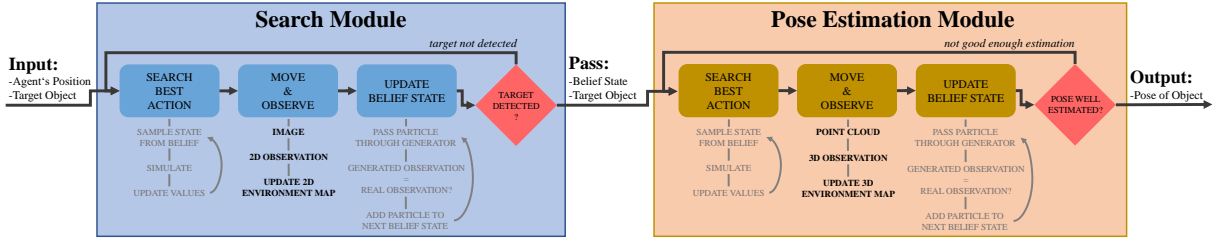


Fig. 1: POMCP Search & Pose Estimation Modules. In the Search Module a target object is searched within a 2-dimensional space using RGB images, producing a belief state of the object’s pose in 2D. This belief acts a prior for the Pose Estimation Module, where the full pose is computed using point clouds.

case where the map is unknown and is dynamically created, which enlarges the search space to navigate the environment. [9] presents two approaches that extend the general POMCP algorithm to the continuous domain, Continuous Real-time POMCP and Adaptive Highest Belief Continuous Real-time POMCP Follower, which are validated in real-life scenarios by navigating over 3km. [10] uses semantic environment models in combination with various search strategies, and does not use the POMCP algorithm.

In contrast to POMDPs, deep RL methods learn a parametrized policy to navigate through environment. This allows faster control frequencies in deployment, as there is no planning phase. Generalizable Approaching Policy Learning (GAPLE) [11] uses depth information for policy learning, whereas [12] builds an episodic semantic map of the environment to explore efficiently. [13] combines two deep neural network modules, one that learns to detect given objects in its view and another that learns to approach the target object. [14] uses a Conditional Random Field (CRF) to build a prediction map representing the most promising locations for the searched object. While these approaches showed good performance, they require a large amount of training data to be able to generalize to unseen situations, and one of the reasons for focusing on solving POMDPs instead.

### III. BACKGROUND

Our proposed method is framed as a POMDP problem. Then, we solve the POMDP problem by adapting the POMCP algorithm to the pose estimation problem. In the following, we introduce the POMDP problem and we briefly explain the POMCP.

#### A. Partially Observable Markov Decision Processes

A POMDP [15] is described as a tuple  $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \Omega, \mathcal{O}, \gamma \rangle$ , where  $\mathcal{S}$  is a finite set of states,  $\mathcal{A}$  a finite set of actions,  $\mathcal{T} : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$  is a state transition model,  $\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$  is a reward function,  $\Omega$  is a finite set of observations,  $\mathcal{O} : \mathcal{S} \times \mathcal{A} \rightarrow \Omega$  is an observation model, and  $\gamma \in [0, 1)$  is a discount factor. At time step  $t$  an agent is in state  $s_t$ , performs an action  $a_t$  according to its policy  $\pi$ , obtains an immediate reward  $r_t$ , transitions to state  $s_{t+1}$  and obtains an observation  $o_t$ . All the variables are available to the agent, except for the internal state  $s$ , which it cannot observe. Therefore, and opposite to a Markov Decision Process (MDP), the policy is conditioned on the history of

Symbol	Meaning
$s_t$	state at time $t$
$a_t$	action at time $t$
$o_{s_t, a_t}$	observation for state $s_t$ and action $a_t$
$\mathbf{M}$	environment map
$V$	observed environment state
$\mathbf{p}^{agent}$	agent’s state
$\mathbf{P}^{object}$	object’s state belief
$\mathbf{I}$	observed RGB image
$\mathbf{PC}$	observed pointcloud
$\mathcal{G}_I$	simulation environment

TABLE I: Used Notation.

previous interactions with the environment and not on the current state  $\pi(a_t | h_t)$ , where  $h_t = \{o_0, a_0, \dots, o_{t-1}, a_{t-1}\}$ . Since the true belief state for a history  $h_t$  is not known, it is approximated using particles  $B_t^i$ . Each particle corresponds to a sample state and the sum of all particles builds the current belief state  $\mathcal{B}(h_t)$ . The goal of a planning agent is to obtain an optimal policy to maximize the expected sum of discounted rewards  $J_\pi = \mathbb{E}_{\pi, \mathcal{T}, \mathcal{O}} [\sum_{t=0}^{\infty} \gamma^t \mathcal{R}(s_t, a_t)]$ .

#### B. Partially Observable Monte-Carlo Planning

A widely used online planning method for POMDPs is the POMCP algorithm [6]. It extends Monte-Carlo tree search (MCTS) [16] to POMDPs and combines it with a Monte-Carlo update of the agent’s belief state. POMCP addresses the intractability of online planning in large POMDPs by utilizing Monte-Carlo sampling for belief state updates and planning. In addition it requires only a black-box simulator of the POMDP, rather than explicit probability distributions, in order to sample state transitions and observations. This algorithm can be divided in two major steps:

- I) Select the best action using an extension of the UCT algorithm [17] for POMDPs - Partially Observable UCT.
- II) Update the belief state after an action is performed and an observation is received, using Monte-Carlo belief state updates.

I) *Partially Observable UCT*: To select the action, three procedures are used:

- 1) SEARCH - search for the best action.
- 2) SIMULATE - determine the value of a state.
- 3) ROLLOUT - extend the search tree.

The SEARCH procedure starts with sampling a state  $s$  from the current belief  $\mathcal{B}(h_t)$ .  $s$  and the history  $h_t$  is given to the SIMULATE procedure in order to evaluate the value

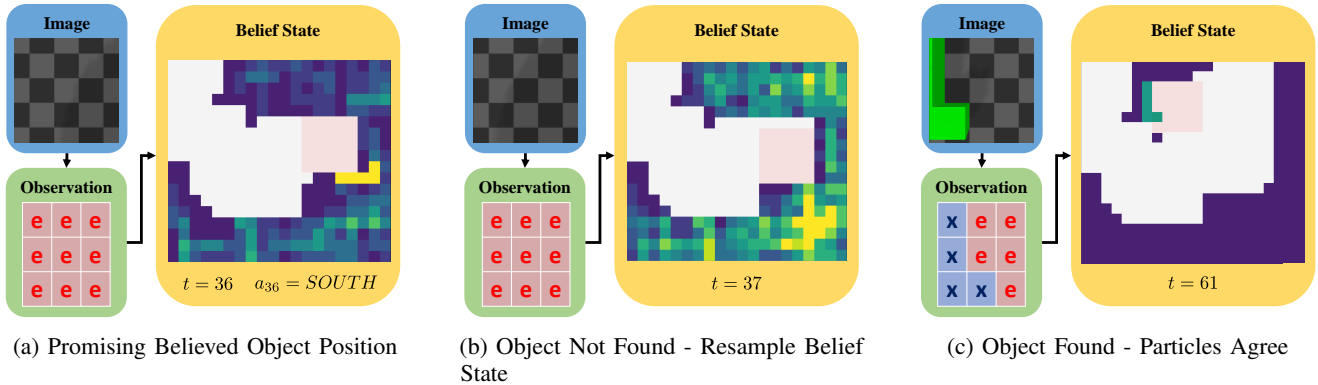


Fig. 2: Illustration of the belief and object position over time steps (indicated with  $t$ ). The light pink square represents the current agent’s position. Parts of the map the agent has already observed are marked in white. a) At time-step  $t = 36$  many particles agree on the object 2D pose, given by the marked yellow  $L$  shape, and therefore the agent moves *SOUTH*. b) Because the object is not the location the agent believed at  $t = 36$ , particles are resampled, resulting in a new belief state. c) At time-step  $t = 61$  the agent finds the object and all particles agree.

of different actions. This process is repeated  $N_{\text{sim}}$  times before selecting the next action to update the search tree and approximate the action values.

The SIMULATE procedure uses two different policies to traverse the search tree, depending on whether the given history is already present in the search tree or not. If the history is present, a tree policy  $\pi_{\text{tree}}$  is used to select actions while navigating through the search tree. This tree policy uses an extension of the UCB1 algorithm [18] for POMDPs in order to compute the value of performing an action given the current history. The best action is then passed to the black box simulator  $\mathcal{G}$  to generate an observation, a reward and the next state  $(s', o, r) \sim \mathcal{G}(s, a)$ . The next state  $s'$  is used to recursively call the SIMULATE procedure and accumulate the returns in a discounted fashion until a terminal state or a maximum depth is reached. If the history is not present during the SEARCH procedure, a new node  $T(ha)$  in the tree is created for each possible action. Then the ROLLOUT procedure is called to determine the value of the given history.

The ROLLOUT procedure selects an action according to a random rollout policy  $\pi_{\text{rollout}}$  that is given to the black box simulator to generate an observation, a reward and the next state. The next state is then used to recursively call the ROLLOUT procedure, until a terminal state or a maximum search depth is reached and a discounted reward is accumulated.

Finally the real action to be executed by the agent  $a_t$  can be selected using the values returned by the SEARCH procedure and a real observation  $o_t$  is received, which is used with Monte-Carlo belief state updates to update the belief state before selecting the next action  $a_{t+1}$ .

*II) Monte-Carlo Belief State Update:* In order to update the belief state,  $K$  particles are used. Each particle  $B_t^i$  represents a randomly selected state  $s$  from the current belief state  $B_t$ , and is then passed with the selected real action  $a_t$  to the black box simulator to receive an observation  $o$  and the next state  $s'$ . The observation  $o$  is used to check, whether it matches

the real observation from the environment  $o_t$ . If they match  $o = o_t$ , the state  $s'$  is added to the next belief state. This process is repeated until  $K$  particles are added to the next belief state  $B_{t+1}$ .

#### IV. COMBINING ACTIVE VISUAL SEARCH AND ACTIVE POSE ESTIMATION IN POMDPs

The problem we address with our method is the one of estimating the pose of an object in an environment using visual information. This means that the agent has no access to the true state of the environment, but to observations.

We consider the situation in which the object might not be in our field of view at first, but it might be occluded or even in a different area of the environment. To deal in a computationally efficient way with this problem, we split the process in two stages. First, the agent navigates along the environment in order to get a decent view for estimating its pose. We call to this stage, the Search stage and we apply AVS methods to search efficiently. Second, once the object is found, we apply APE methods to estimate the pose of the object. Namely, we call it Pose Estimation stage. An overview of our method that combines those two stages, is shown in Fig. 1.

AVS has been shown to work with RGB images and therefore we do not use point cloud data when searching the object. Because of this, we obtain a computationally more efficient approach. Hence, we merge AVS using RGB images and APE using point clouds into a single approach. While it is possible to first use AVS and then do pose estimation independently, in our approach we use the environment information from the search module as a prior for the pose estimation module improving the computational efficiency. In the following, we introduce extensively the algorithm we run on each module.

##### A. Search Module

We solve the AVS problem with POMDPs methods. We propose to solve the search problem in a two-

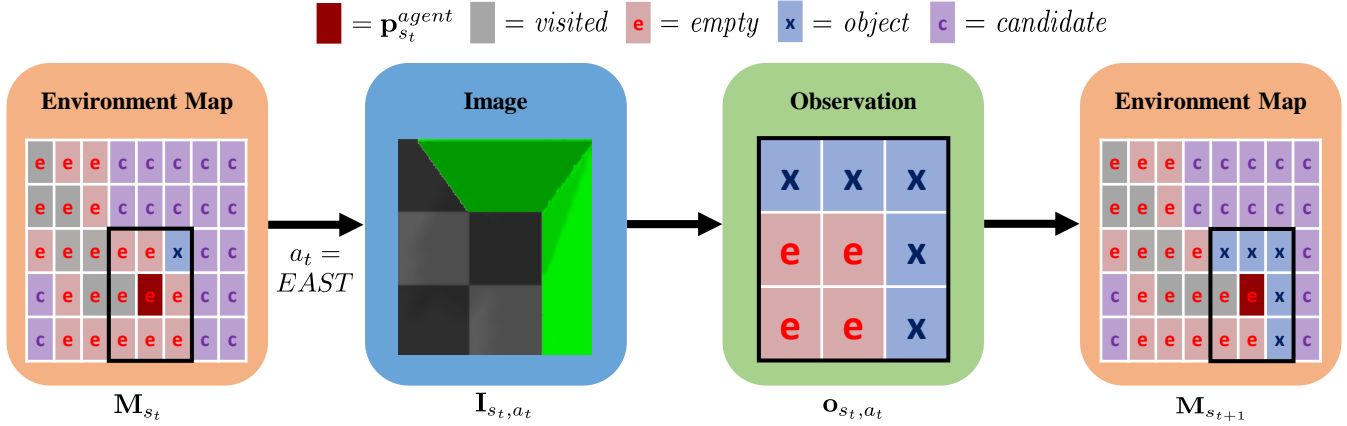


Fig. 3: Environment map update between two steps  $\mathbf{M}_t \rightarrow \mathbf{M}_{t+1}$  in the 2D Search Module. The leftmost figure shows the current map and position of the agent (center of black borders). After executing the real action  $a_t$  in the current state  $s_t$ , an RGB image  $\mathbf{I}_{s_t, a_t}$  is received (2nd figure), which is processed into an observation (3rd figure). With this observation the agent updates his internal map of the environment accordingly. Notice the area inside the black borders.

dimensional projection of the environment. This projection has been previously shown to be enough [7]. We stick to this approach and therefore define our environment map as  $\mathbf{M} \in V^{x \times y}$  where  $V$  are the possible values of a cell and  $x, y$  the discrete cell position in the environment.  $V$  can take 5 possible values, representing the knowledge we have about the observed environment,  $V := \{\text{empty}, \text{candidate}, \text{object}, \text{other\_object}, \text{blocked}\}$ . Finally, for each time step  $t$  the agent is in a specific state  $s_t$ , which represents both the knowledge of the environment and the belief states of our search object. This state is a tuple  $s_t = (\mathbf{M}, \mathbf{p}_{s_t}^{agent}, \mathbf{P}_{s_t}^{object})$ , where  $\mathbf{M}_{s_t}$  is the state's map of the environment, the agents position  $\mathbf{p}_{s_t}^{agent} \in \mathbb{N}^2$  and the believed object position  $\mathbf{P}_{s_t}^{object} \in \mathbb{N}^{obj\_size \times 2}$ . Our definition of the believed object position allows to define objects corresponding to the occupied cells in the environment map. Therefore, each object is described by locations in  $\mathbf{M}$  according to the level of discretization. For example, a cube might take a single cell whereas an edge-like shape might occupy three cells.

In a POMDP, observations are the only way an agent can reason about the state of the environment. In our case we use RGB images to produce observations which are then used to update the gathered information about the environment, as depicted in Fig. 3. We use observations to replace *candidate* values of the environment map with observed values. Those observed values are generated by transforming an image  $\mathbf{I}_{s_t, a_t}$  to an observation  $\mathbf{o}_t$  after an action  $a_t$  in state  $s_t$  is performed. The observation  $\mathbf{o}_t \in V^{w \times h}$  represents, in which cells of the image the searched object is observed, where  $w$  is the amount of cells observed along the x-axis and  $h$  the y-axis ones. Various methods can be used to transform an image into this type of observation, e.g. color based decisions can be sufficient. Those observations are introducing noise and misidentified cells as can be seen in Fig. 5. However, we can show that they are expressive enough and allow to

successfully find objects.

Most important, those type of observations allow the agent to reason about the environment and the object by himself. This means if an object is identified in some part of the image, the agent will explore that region until it observes the full object. This exploration is the major benefit towards the binary observations mentioned earlier. There is more information that is passed to the agent and therefore used for action selection. These observations allow to move to believed object positions in the received images, rather than simply exploring until the object is found in an image.

Since our method is using POMCP, a black box simulator  $\mathcal{G}_{\mathbf{I}}(s_t, a_t)$  is required to perform simulations. It returns for a given state  $s_t$  and action  $a_t$  an successor state  $s'$  and the corresponding generated observation  $\mathbf{o}_{s_t, a_t}$  along with the reward  $r_t$ . Since, we cannot generate images  $\mathbf{I}_{s_t, a_t}$  in order to generate the observation  $\mathbf{o}_{s_t, a_t}$ , the black box simulator has to be capable of surpassing the need for images. We solve this by directly creating the observation  $\mathbf{o}_{s_t, a_t}$  by generating it from the environment map  $\mathbf{M}_{s_t}$  of the given state  $s_t$ . We know the position and orientation of the agent as well as the characteristics of the camera used to take images, which allows us to calculate in which parts of the image, an object should be seen. It is important to notice, that this calculation is independent of the true position of the searched object, but only dependent on the believed object position  $\mathbf{P}_{s_t}^{object}$  of the state  $s_t$ . This means, after a true observation from the environment is received, we can pass all particles to the black box simulator to generate observations. Those generated observations are used to discard the particles which believed in a contradicting object position. For example, if a particle believes the object to be seen, when taking the action  $a_t$ , but taking that action in the environment results not in observing the object, the particle is thrown away and a new particle is sampled.

Using those generated observations to update the belief state results in more correct believed object positions, which

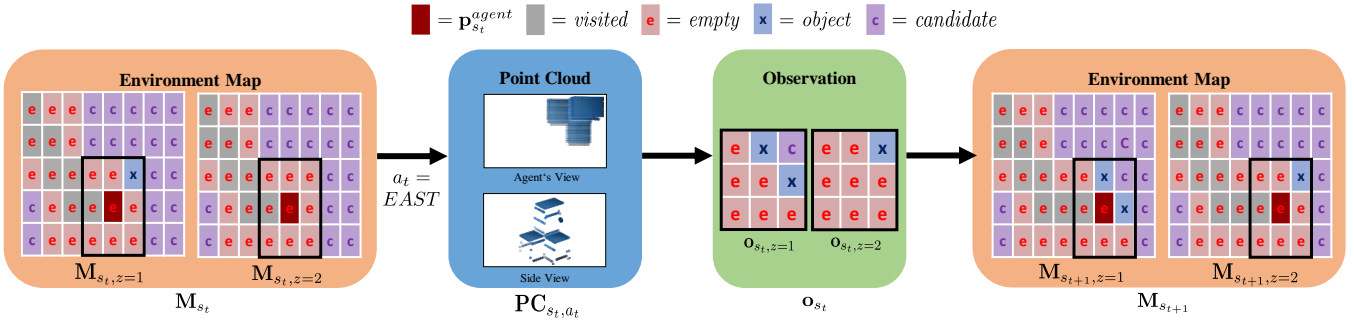


Fig. 4: Environment map update between two steps  $M_t \rightarrow M_{t+1}$  in the 3D Pose Estimation Module. The leftmost figure shows two levels of the current map ( $z = 1$  and  $z = 2$ ) and position of the agent (center of black borders). After executing the real action  $a_t$  in the current state  $s_t$ , a point cloud  $PC_{s_t, a_t}$  is received (2nd figure), which is processed into an observation  $o_t$ . The observation has the  $Z$ -levels, where each is used to update the corresponding level in the environment  $M_{t, z=i} \rightarrow M_{t+1, z=i}$ ,  $i = 1, 2, \dots, Z$ . With this observation the agent updates his internal map of the environment accordingly. Notice the area inside the black borders.

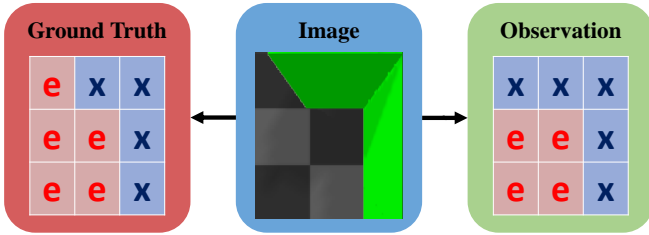


Fig. 5: Mapping from RGB images to observations. Creating observations from images can lead to mistakes, e.g. due to perspective distortion. The upper left corner in the observation is marked with *object*, while the ground truth is *empty*.

will finally result in finding the object at the believed object position, as can be seen in Fig. 2. A terminal state  $s_t$  has  $\forall(x, y) \in \mathbf{P}_{s_t}^{object} : M_{s_t}(x, y) = object$ . This can be interpreted as a state, whose believed object position matches its past observations from the environment. When such a terminal state is reached, the current belief state in terms of particles is passed to the pose estimation module and the search module terminates.

### B. Pose Estimation Module

The pose estimation module follows the same concept as the search module, with two major differences: (1) the environment map is now three-dimensional:  $M \in V^{x \times y \times z}$ , and (2) the observations are produced using point clouds  $PC_{s_t, a_t}$ .

In order to use the prior belief state, each particle  $B_{2D}^i$  from the search module corresponds to a particle  $B_{3D}^i$  in the Pose Estimation module. The particles in the Pose Estimation module now set the lowest level of their environment map to the one passed from the search module  $\forall x \forall y : M_{B_{3D}^i}(x, y, 0) = M_{B_{2D}^i}(x, y)$  and all other levels are filled with *candidate* values. Therefore, the initial belief represents a rough belief of the object's pose and the three-dimensional environment in general.

In order to estimate a more precise pose, we use point clouds  $PC_{s_t, a_t}$  instead of images  $I_{s_t, a_t}$  and change the observation to be three-dimensional  $o_t \in V^{w \times h \times d}$ , where  $d$  is now representing the observed cells of the environment along the  $z$ -axis. The update process for the environment map  $M_{s_t}$  after an observation  $o_t$  is received, is the same as in the case of the search module, but now with respect to three-dimensions, as can be seen in Fig. 4.

Using point clouds introduces a different type of noise as images do. In order to handle noisy observations we introduce

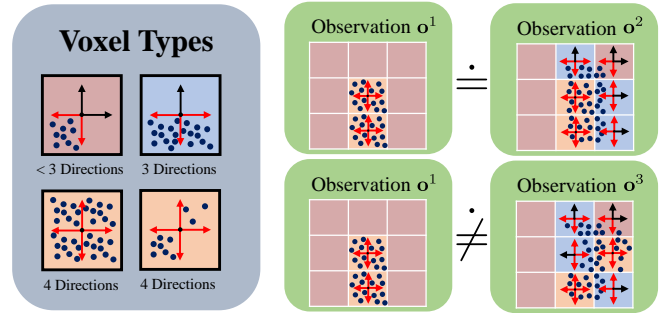


Fig. 6: Soft equality for point cloud observations. Cells with points in all quadrants (orange) are denoted *core* cells.

a soft equality ( $\doteq$ ) relationship. We classify every cell in a level of the environment map  $M_{s_t}$  as one specific type according to the positioning of the points it contains, before assigning an observed value to it. In more detail, only cells classified as *core* cells (see Fig. 6), are used for testing Soft Equality of two observations. This means only observations with equal core cells are considered as equal.

## V. EXPERIMENTS

A new black box simulator  $\mathcal{G}_{PC}(s_t, a_t)$  is introduced, which follows the same intuition of the simulator  $\mathcal{G}_I$ , but now generates point clouds according to the belief in the object's position and so generates three-dimensional observations.



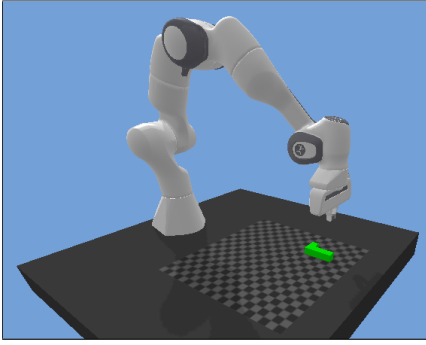


Fig. 7: Simulation environment with the Franka Emika Panda robot arm in PyBullet. A virtual camera is attached to the end-effector to capture RGB images.

This procedure can take more time than when using an RGB image, but also gives a more precise pose estimation of the object.

Our experiments were designed to answer the following questions:

- 1) Does our mapping from images to observations improve over a binary observation decision in the context of AVS?
- 2) Can the combination of search with a pose estimation module estimate object poses in a POMDP?
- 3) How is our approach affected by hyperparameters and different types of objects?

Our environment contains an agent depicted as a Franka Emika Panda robot arm that has to search for objects laying on a table as in Fig. 7. The implementation uses PandaGym [19] and Pybullet [20]. The search space is defined by a  $20 \times 20$  grid on the table with 5 levels in the  $z$ -dimension. The considered objects are letters represented by multiple connected cubes placed in the grid, as depicted in Fig. 8. The action space is  $\mathcal{A} = \{NORTH, EAST, SOUTH, WEST\}$ , which corresponds to moving the end-effector in discrete steps along the  $x$ - and  $y$ -directions while maintaining a fixed height  $z = 5$ . Observations are generated using an RGBD camera mounted at the end-effector, pointing towards the table. The black box simulator  $\mathcal{G}$  uses Open3D [21] to render RGBD images according to the believed object positions.

The reward function is defined as

$$\begin{aligned}
 R_1 &= P_{\text{action}} \\
 R_2 &= \begin{cases} P_{\text{re-observe}} & \text{if } \mathbf{M}_s = \mathbf{M}_{s'} \\ 0 & \text{otherwise} \end{cases} \\
 R_3 &= \begin{cases} R_{\text{terminal}} & \text{if } s' \text{ is terminal} \\ 0 & \text{otherwise} \end{cases} \\
 R_{1,2,3} &= R_1 + R_2 + R_3 \\
 R_{2D} &= R_{1,2,3} + R_{\text{exploration}} + R_{\text{discovery}} \\
 R_{3D} &= R_{1,2,3} + R_{\text{refinement}},
 \end{aligned}$$

where  $P_{\text{action}}$  is a constant penalty for performing an



Fig. 8: Examples of objects considered in the experiments.

action,  $P_{\text{re-observe}}$  is a penalty depending on the amount of the environment map that is re-observed when performing an action and  $R_{\text{terminal}}$  is the reward for reaching a terminal state. The search module uses  $R_{2D}$  where  $R_{\text{exploration}}$  is a reward for observing cells that are marked as *candidate*, while  $R_{\text{discovery}}$  is a reward for observing objects. The pose estimation module uses  $R_{3D}$ , where  $R_{\text{refinement}}$  is a reward for discarding cells marked as *object* while they are actually *empty*.

#### A. Evaluations

In Fig. 9 the left and middle figures compare the performance of our search module using binary observations  $o \in \{\text{object\_found}, \neg \text{object\_found}\}$  against our proposed image to observation mapping, where  $\mathbf{o} \in V^{w \times h}$ . The results show that our method is able to find the object faster (less average steps) and more reliably (more times the terminal state is reached). Therefore, confirming that it is beneficial to introduce a feature extraction map from images to observations, instead of a simple binary classification to decide whether the object is seen or not.

In Fig. 9 the rightmost plot shows the performance of combining the search and pose estimation modules - AVS + APE. From the plot is clear that larger objects are harder to handle than smaller ones. With more maximum number of setups all objects could be potentially estimated, at the expense of higher computation.

## VI. CONCLUSION & FUTURE WORK

In those situations in which the pose of a certain object is difficult to assess from a single image, actively deciding the actions the agent should do to improve the pose estimation is a mandatory skill for our robots. Nevertheless, solving directly the problem for the objects full pose might be computationally demanding, due to the sensor inputs and the dimension of the possible solutions. To deal with this computational limitation, we propose to split the active pose estimation problem in two stages: A searching process and a pose estimation one. In our work, we propose to model both stage problems as POMDP and solve them with POMCP methods. Framing both stages under the same umbrella allow us to comprehend how both phases are related to each other and allow us to share information between each other such

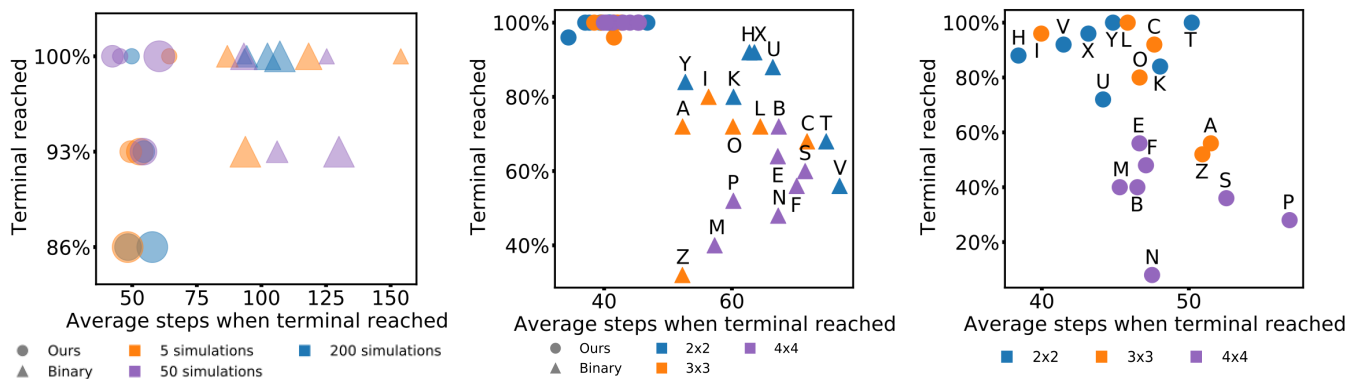


Fig. 9: Comparisons between our mapping from RGB images to observations vs. a binary object detector. The number of simulations and particles in the middle and right figures are fixed to  $N_{\text{sim}} = 200$  and  $K = 200$ , respectively. In the legends, the digits near the marker, e.g.  $2 \times 2$  indicate the area occupied by the object. Left: results for Search Module with object  $L$ . The marker size is proportional to the number of used particles. The colors indicate the number of simulations. Middle: Results for Search Module with all objects and a fixed number of particles and simulations. Right: Results for combining the Search and Pose Estimation Modules with all objects and fixing the number of particles and simulations.

as the computed belief states. Along the paper, we present two proposed algorithms to integrate POMCP into the search a pose estimation problems and evaluate its performance for a high number of different objects. Through an extensive ablation study, we have evaluated our proposed framework and measure the required number of control steps to solve the task under several variables: different shape and size objects or different type of belief updates.

Our proposed approach allows the use of multiple possible features to represents the belief update metric. Nevertheless, in practise, we limit to the color features. In future works, we aim to explore the performance of the proposed belief update metric for additional features, from handcrafted ones such as edges to learned features. Additionally, we aim to explore the performance of our proposed method in real robot pose estimation tasks and study the required changes to adapt it to a real robot problem.

## REFERENCES

- [1] X. Li, Y. Weng, L. Yi, L. J. Guibas, A. L. Abbott, S. Song, and H. Wang, "Leveraging SE(3) equivariance for self-supervised category-level object pose estimation," *CoRR*, vol. abs/2111.00190, 2021.
- [2] K. Kleeberger, M. Völkl, R. Bormann, and M. F. Huber, "Investigations on output parameterizations of neural networks for single shot 6d object pose estimation," *CoRR*, vol. abs/2104.07528, 2021.
- [3] S. Zhang, W. Zhao, Z. Guan, X. Peng, and J. Peng, "Keypoint-graph-driven learning framework for object pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 1065–1073.
- [4] J. Lin, Z. Wei, Z. Li, S. Xu, K. Jia, and Y. Li, "Dualposenet: Category-level 6d object pose and size estimation using dual pose network with refined learning of pose consistency," *CoRR*, vol. abs/2103.06526, 2021.
- [5] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Sensor fusion IV: control paradigms and data structures*, vol. 1611. Spie, 1992, pp. 586–606.
- [6] D. Silver and J. Veness, "Monte-carlo planning in large pomdps," in *Advances in Neural Information Processing Systems*, J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, Eds., vol. 23. Curran Associates, Inc., 2010.
- [7] Y. Wang, F. Giuliani, R. Berra, A. Castellini, A. D. Bue, A. Farinelli, M. Cristani, and F. Setti, "POMP: pomcp-based online motion planning for active visual search in indoor environments," *CoRR*, vol. abs/2009.08140, 2020. [Online]. Available: <https://arxiv.org/abs/2009.08140>
- [8] F. Giuliani, A. Castellini, R. Berra, A. D. Bue, A. Farinelli, M. Cristani, F. Setti, and Y. Wang, "POMP++: pomcp-based active visual search in unknown indoor environments," *CoRR*, vol. abs/2107.00914, 2021. [Online]. Available: <https://arxiv.org/abs/2107.00914>
- [9] A. Goldhoorn, A. Garrell, R. Alquézar, and A. Sanfeliu, "Continuous real time pomcp to find-and-follow people by a humanoid service robot," in *2014 IEEE-RAS International Conference on Humanoid Robots*, 2014, pp. 741–747.
- [10] M. Saito, H. Chen, K. Okada, M. Inaba, L. Kunze, and M. Beetz, "Semantic object search in large-scale indoor environments," in *IROS 2011*, 2011.
- [11] X. Ye, Z. Lin, J.-Y. Lee, J. Zhang, S. Zheng, and Y. Yang, "Gapple: Generalizable approaching policy learning for robotic object searching in indoor environment," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4003–4010, 2019.
- [12] D. S. Chaplot, D. P. Gandhi, A. Gupta, and R. R. Salakhutdinov, "Object goal navigation using goal-oriented semantic exploration," *Advances in Neural Information Processing Systems*, vol. 33, 2020.
- [13] X. Ye, Z. Lin, H. Li, S. Zheng, and Y. Yang, "Active object perceiver: Recognition-guided policy learning for object searching on mobile robots," *CoRR*, vol. abs/1807.11174, 2018. [Online]. Available: <http://arxiv.org/abs/1807.11174>
- [14] A. C. Hernandez, M. Durner, C. Gomez, I. Grixia, O. Teikmanis, Z.-C. Marton, and R. Barber, "Searching for objects in human living environments based on relevant inferred and mined priors," in *2021 European Conference on Mobile Robots (ECMR)*, 2021, pp. 1–7.
- [15] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial Intelligence*, vol. 101, no. 1, pp. 99–134, 1998. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S000437029800023X>
- [16] R. Coulom, "Efficient selectivity and backup operators in monte-carlo tree search," in *Proceedings of the 5th International Conference on Computers and Games*, ser. CG'06. Berlin, Heidelberg: Springer-Verlag, 2006, p. 72–83.
- [17] L. Kocsis and C. Szepesvári, "Bandit based monte-carlo planning," vol. 2006, 09 2006, pp. 282–293.
- [18] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2, pp. 235–256, 2002.
- [19] Q. Gallouédec, N. Cazin, E. Dellandréa, and L. Chen, "panda-gym: Open-Source Goal-Conditioned Environments for Robotic Learning," *4th Robot Learning Workshop: Self-Supervised and Lifelong Learning at NeurIPS*, 2021.
- [20] E. Coumans and Y. Bai, "Pybullet, a python module for physics simulation for games, robotics and machine learning," <http://pybullet.org>, 2016–2021.
- [21] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," *arXiv:1801.09847*, 2018.