

# Mixed Observable RRT: Multi-Agent Mission-Planning in Partially Observable Environments

Kasper Johansson,<sup>1</sup> Ugo Rosolia,<sup>2</sup> Wyatt Ubellacker,<sup>3</sup> Andrew Singletary,<sup>3</sup> and Aaron D. Ames<sup>3</sup>

**Abstract**—This paper considers centralized mission-planning for a heterogeneous multi-agent system with the aim of locating a hidden target. We propose a mixed observable setting, consisting of a fully observable state-space and a partially observable environment, using a hidden Markov model. First, we construct rapidly exploring random trees (RRTs) to introduce the mixed observable RRT for finding plausible mission plans giving way-points for each agent. Leveraging this construction, we present a path-selection strategy based on a dynamic programming approach, which accounts for the uncertainty from partial observations and minimizes the expected cost. Finally, we combine the high-level plan with model predictive control algorithms to evaluate the approach on an experimental setup consisting of a quadruped robot and a drone. It is shown that agents are able to make intelligent decisions to explore the area efficiently and locate the target through collaborative actions.

## I. INTRODUCTION

Planning under uncertainty is important for autonomous systems. Similar to human decision-making robotic systems must be able to act based on observations from the environment. Exploration tasks are one type of problem where robots have partial knowledge about the environment, complemented with observations along the mission. There are several important instances of exploration tasks. One motivating example is the Mars explorations task, where the objective is to allow robots to act autonomously and gather information on Mars [1]. However, there seems to exist no satisfying solution to such exploration when multiple robots are to cooperate and act simultaneously to locate a hidden target in a partially known environment.

We introduce a novel approach to path-planning for multi-agent systems in partially observable environments. Agents act simultaneously to locate a target, as illustrated in Figure 1. We leverage rapidly exploring random trees (RRTs) to generate a sample of event-based plans, where a plan consists of a sequence of way-points that changes as a function of the realized observations agents make. The best plan is found by minimizing an expected cost, formulated as a summation over nodes in trajectory trees, as proposed in [2]. Lastly, to allow a multi-agent system to follow the high-level plan we leverage local model predictive control (MPC) algorithms.

### A. Contribution

Our contribution is threefold. First, we model a cooperative multi-agent exploration mission using a mixed ob-

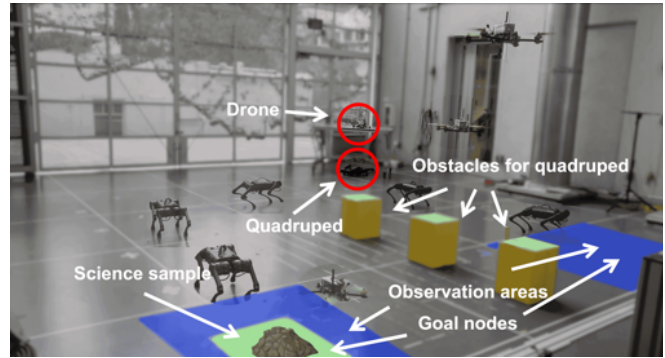


Fig. 1: A quadruped and a drone are initiated inside the red circles. They locate a science sample by sharing observations from the environment.

servable setting. We modify the standard RRT algorithm to generate a sample of plans in the discrete partially observable environment, defining the *mixed observable RRT (MORRT)*. Secondly, we illustrate the advantage of the MORRT over standard RRTs in uncertain environments and apply the algorithm to a system of five agents. Finally, we demonstrate our approach experimentally on hardware by integrating the MORRT plan, consisting of way-points for a heterogeneous multi-agent system, with local MPC algorithms.

The paper is outlined as follows. Section I reviews previous works. In Section II we formulate the problem and describe the mixed observable setting. Section III details our main contribution, the sample-based search algorithm for mixed observable settings, denoted the MORRT. In Section IV we report results from simulations and a hardware experiment with a quadruped robot and a drone, as well as illustrate the advantage over standard RRTs in uncertain environments. Finally, we summarize our findings in Section V.

### B. Related works

**Planning.** The literature on environment mappings for navigation and planning is rich [1], [3]–[5]. Exploration can be done using a policy that maps the system’s state to a control action [2]. The computational burden associated with planning over policies has been extensively studied [6]–[13], and can be reduced when a system’s state can be perfectly measured [2], [6], [11], [14]. Tube MPC strategies involve another type of feedback policies [7]–[10]. Strategies where only partial state knowledge is available have been studied in [15]–[18]. In [2] the authors consider multi-modal measurement noise by planning over a tree of trajectories where each branch is associated with a unique set of observations. We propose a novel way to tackle planning problems under partial observations, based on RRTs.

<sup>1</sup>Stanford University, Stanford, USA. kasperjo@stanford.edu. This work was done while Kasper was at KTH Royal Institute of Technology, Stockholm, Sweden.

<sup>2</sup>Amazon, Luxembourg, Luxembourg. ugo.rosolia@gmail.com

<sup>3</sup>California Institute of Technology, Pasadena, USA. {wubellac, asinglet, ames}@caltech.edu.

**Rapidly exploring random trees.** A popular tool for path-planning is the RRT, which is a sampling-based search algorithm [19]. RRTs have been leveraged for path-planning in various problems, like navigating among moving obstacles [20], kinodynamic motion-planning [21], and probabilistically robust path-planning [22], among many others. The RRT\* algorithm is an extension of the RRT [23], and it has been shown that RRT\* converges to the optimal path [24]. Although RRTs have been shown to perform well in certain tasks, they lack in that they do not allow for information gathering in the form of observations. In this work, we extend the RRT to allow for drastic changes in the path when observations are made and the target belief changes.

**Coordinated multi-robot exploration.** Coordinated exploration for multi-robot systems has been extensively studied, e.g., [25]–[32]. These works tackle problems concerning efficient exploration of an environment, like sweeping [30], exploration with marsupial teams [27], exploring using segmentation [26], etc. More recent works include [33] where the authors propose a hybrid stochastic exploration method based on the Grey Wolf optimization algorithm and coordinated multi-robot exploration algorithms; [34] proposes a framework that combines deterministic Coordinated Multi-Robot Exploration with metaheuristic frequency modified Whale Optimization Algorithm; and [35] proposes another hybrid approach for multi-robot exploration. All aforementioned papers deal with unknown environments. Our work differs in that we assume a partially known environment, where the only unknown is the location of the target.

## II. PROBLEM FORMULATION

We describe the discrete partially observable environment, the multi-agent system, and the control objectives.

### A. Environment model

We consider a multi-agent system of  $M$  agents with the aim of locating a target that is hidden in  $\mathbb{R}^n$ . Agents have some prior knowledge of where the target is located; they know that it is at one of a finite number of known *goal nodes*, points in  $\mathbb{R}^n$ . To find the hidden target, agents can make observations in *observation areas*, certain regions of state-space, and observations are shared between agents. Agents have a map of the environment, including goal nodes, observation areas, and potential obstacles. However, agents do not know at which of the goal nodes the target is located. As an example, see Figure 1 where a quadruped and a drone are locating a science sample. In this example the goal nodes lie inside the blue observation areas, so agents can make observations near potential target locations. However, it is important to mention that observation areas can in general be placed anywhere, not only near goal nodes.

We assume a finite number,  $G$ , of goal nodes where the target can be located, corresponding to  $G$  different environment states  $e$ . The environment state may change at each time step and it can only be inferred through noisy observations. The evolution of the environment state  $e$ , representing the

goal location, is modeled using a *hidden Markov model (HMM)* [36] given by the tuple  $\mathcal{H} = (\mathcal{E}, \mathcal{O}, T, Z)$ :

- $\mathcal{E} = \{0, 1, \dots, G - 1\}$  is a set of partially observable environment states.
- $\mathcal{O} = \{0, 1, \dots, G - 1\}$  is the set of observations for the partially observable state  $e \in \mathcal{E}$ .
- The function  $T : \mathcal{E} \times \mathcal{E} \times \mathbb{R}^n \rightarrow [0, 1]$  describes the probability of transitioning to a state  $e'$  given the current environment state  $e$  and system's state  $x$ , i.e.,  $T(e', e, x) := \mathbb{P}(e'|e, x)$ .
- The function  $Z : \mathcal{E} \times \mathcal{O} \times \mathbb{R}^n \rightarrow [0, 1]$  describes the probability of observing  $o$ , given the current environment state  $e$  and the system's state  $x$ , i.e.,  $Z(e, o, x) := \mathbb{P}(o|e, x)$ .

Let  $b_k \in \mathcal{B} = \{b \in \mathbb{R}_{0+}^{|\mathcal{E}|} : \sum_{e=0}^{|\mathcal{E}|-1} b[e] = 1\}$  (where entry  $b_k[e]$  is the posterior probability, at time  $k$ , that the environment's state  $e_k$  equals to  $e \in \mathcal{E}$ ) be the belief vector.

### B. Control design objectives

At time  $k$  each agent  $a$  tries to reach a way-point  $x_{k+1}^a$ , computed as

$$x_{k+1}^a = x_k^a + \Delta x_k^a, \quad (1)$$

where  $\Delta x_k^a \in \mathbb{R}^n$  is the way-point change that must satisfy  $x_k^a + \Delta x_k^a \in \mathcal{X}_{\text{free}}^a := \mathcal{X}^a \setminus \mathcal{X}_{\text{obs}}^a$ , where  $\mathcal{X}^a \in \mathbb{R}^n$  is the state-space and  $\mathcal{X}_{\text{obs}}^a \in \mathbb{R}^n$  the obstacle region, of agent  $a$ .

Let  $j(k)$  be the number of observations made until time  $k$  and  $\mathbf{o}_{j(k)} = [o_1, \dots, o_{j(k)}]$  the set of observations. As the environment is not perfectly known, and noisy observations are collected during execution, it is not ideal to compute a sequence of way-points offline. Instead, we compute a function  $\pi_k^a$ , which maps the environment observations up to time  $k$  to the way-point change for agent  $a$ , i.e.,

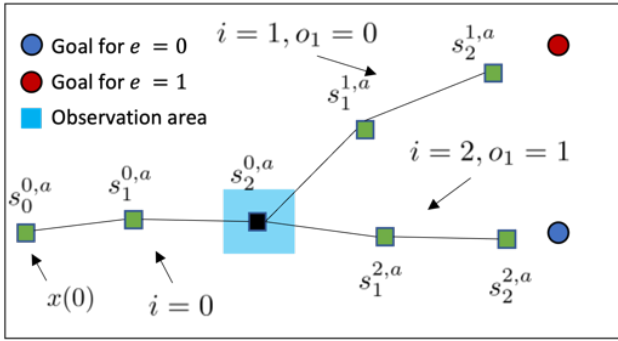
$$\Delta x_k^a = \pi_k^a(\mathbf{o}_{j(k)}). \quad (2)$$

We denote the policy of a mission, ending at some time  $N_p$ , of agent  $a$  by  $\pi^a = [\pi_0^a, \dots, \pi_{N_p-1}^a]$  and a plan  $\pi$  as the collection of all policies:  $\pi = \{\pi^a\}_{a=1}^M$ . A path is a realization of way-points  $p^a = [x_0^a, x_1^a, \dots, x_{N_p}^a]$ , such that  $x_{k+1}^a = x_k^a + \Delta x_k^a, \forall k = 0, 1, \dots, N_p - 1$ .  $p = [p^1, \dots, p^M]$  is a collection of paths. Given a plan  $\pi$ , we define the set of possible collections of paths under this plan by  $\mathcal{P}_\pi$ . Note that, by (2), each  $p \in \mathcal{P}_\pi$  is a collection of paths corresponding to a realized set of observations agents make.

Given the initial state  $x(t) = [x^1(t), \dots, x^M(t)]$  and environment belief  $b(t)$ , as well as a plan  $\pi$ , we define the cost of the plan for agent  $a$ , at time  $t$ , as follows:

$$\begin{aligned} J^a(x(t), b(t)) &= \mathbb{E}_{\mathcal{P}_\pi} \left[ \sum_{k=0}^{N_p-1} h(x_k^a, \Delta x_k^a, e_k) + h_{N_p}(x_{N_p}^a, e_{N_p}) \middle| b(t) \right], \end{aligned} \quad (3)$$

where the stage cost  $h : \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{E} \rightarrow \mathbb{R}$  and terminal cost  $h_N : \mathbb{R}^n \times \mathcal{E} \rightarrow \mathbb{R}$  are functions of the partially observable environment state  $e \in \mathcal{E}$ , and the expectation is over the set of possible realizations of agent paths  $\mathcal{P}_\pi$ .



**Fig. 2:** A trajectory tree for agent  $a$ , representing a plan  $\pi$ . Nodes and edges correspond to way-points and way-point changes, respectively, and each branch  $i$  is associated with a unique observation vector. Here,  $\mathcal{I}_\pi = \{0, 1, 2\}$ ;  $\mathbb{P}(1) = \mathbb{P}(2) = 0$ ;  $\circ(0) = 1$ , and  $\circ(1) = \circ(2) = 0$ . Branch 0 corresponds to the initial belief  $b(0)$ , and branches 1 and 2 correspond to the observation vectors  $\mathbf{o}_1 = [0]$  and  $\mathbf{o}_1 = [1]$ , respectively.

### III. THE MIXED OBSERVABLE RRT

This section describes the main contribution of our work. We introduce the *mixed observable RRT (MORRT)* and explain how it can be leveraged to compute a plan offline, given an environment map.

#### A. Cost reformulation

The policy of agent  $a$  can be illustrated by a trajectory tree where nodes and edges represent way-points and way-point changes, respectively. Each branch in the tree is a path that the agent commits to, given a specific observation and environment belief update. Basically, the controller plans different trajectories, and future decisions are a function of the realized environment observations. A plan, which is the collection of  $M$  policies, can be illustrated as a collection of  $M$  trajectory trees. The example in Figure 2 consists of three trajectory branches; since observations are shared between agents, each agent in this example has a trajectory tree with three branches, corresponding to the branches in the figure.

Denote the set of trajectory branches, for a given plan  $\pi$ , by  $\mathcal{I}_\pi$ . We introduce, for each branch  $i \in \mathcal{I}_\pi$  and each agent  $a \in \{1, 2, \dots, M\}$ , the vector  $s^{i,a} = [s_0^{i,a}, \dots, s_{N_i}^{i,a}]$ , which given a plan  $\pi$ , denotes the way-points of agent  $a$  in the branch  $i$ , where  $N_i + 1$  is the number of way-points in this branch. It is important to note that, for each  $s_\ell^{i,a}$ , there is a one-to-one mapping to  $x_k^a$  for some  $k$ , where  $k$  is the time index; introducing the  $s$ -notation makes it easy to distinguish between different branches without keeping track of the time index  $k$  in each branch. Moreover, let the function  $\mathbb{P}(i)$  return the parent branch of branch  $i$ , and let  $\circ(i) = 1$  if branch  $i$  ends at an *observation node*—a way-point inside of an observation area—and  $\circ(i) = 0$  otherwise (see Figure 2).

Next, as the environment states are discrete the expected cost (3) can be rewritten as a summation over way-points. We use Bayesian belief updates, i.e., after observing  $o$  at time  $k$  the belief is updated as

$$b_k[e] = \frac{Z(e, o, x_k)}{\mathbb{P}(o|x_k, b_{k-1}[e])} \sum_{e' \in \mathcal{E}} T(e, e', x_k) b_{k-1}[e']$$

for the functions  $Z$  and  $T$  defined in Section II-A.

**Proposition 1.** Denote the state of the multi-agent system at time  $k$  as  $x_k = [x_k^1, \dots, x_k^M]$ . Let  $o_i$  denote the observation corresponding to the branch  $i \in \mathcal{I}_\pi$ , and  $v^i[e] = \Theta(o_i, s_0^i) v^{\mathbb{P}(i)}[e]$  the unnormalized belief, where

$$\Theta(o_i, s_0^i) = \text{diag}(Z(0, o_i, s_0^i) \dots Z(|\mathcal{E}| - 1, o_i, s_0^i)).$$

The expected cost (3) can be expressed equivalently as

$$J^a(\mathcal{I}_\pi, x(t), b(t)) = \sum_{i \in \mathcal{I}_\pi} \left\{ \sum_{\ell=0}^{N_i-1} \sum_{e \in \mathcal{E}} v^i[e] h(s_\ell^{i,a}, \Delta s_\ell^{i,a}, e) + \mathbb{1}[\circ(i) = 0] \sum_{e \in \mathcal{E}} v^i[e] h_N(s_{N_i}^{i,a}, e) \right\}, \quad (4)$$

where  $s_0^0 = x(t)$ .

*Proof.* This is a direct extension of Proposition 1 in [2]: instead of making observations at each time step  $k = 1, 2, \dots$ , agents make observations at the beginning of each branch  $i \in \mathcal{I} \setminus \{0\}$ , which amounts to a summation over the nodes in each branch, rather than each time step.  $\square$

#### B. Building the MORRT

We now show how to construct an MORRT for the single-agent case. The multi-agent extension is discussed in Section III-C. First, we introduce a modified RRT that stores observations collected in observation areas. We leverage this RRT algorithm to construct the MORRT. Algorithm 1 constructs the RRT while storing observations. It takes as input a tree  $\mathcal{T}$  and three parameters  $\Delta x$ ,  $N$ , and  $K$ . Each tree  $\mathcal{T}$  has three features:  $\mathcal{T}.\text{x\_obs}$ , a set of all observation nodes in  $\mathcal{T}$ ;  $\mathcal{T}.\text{areas}$ , a set of observation areas; and  $\mathcal{T}.\text{child}$ , the child RRTs in a tree of RRTs. Moreover, each node  $x \in \mathcal{T}$  has three features:  $x.\text{area}$ , an observation area, if  $x$  is inside of one;  $x.\text{count}$ , the path length from the root node (henceforth denoted  $x_{\text{init}}$ ) to  $x$  in  $\mathcal{T}$ ; and  $x.\text{parent}$ , the parent node in  $\mathcal{T}$ . In line 2 of Algorithm 1, a node  $x_{\text{rand}}$  is sampled in state-space and the nearest node to  $x_{\text{rand}}$  in  $\mathcal{T}$  is denoted  $x_{\text{near}}$ . The `new_node` function (line 3) generates a new node  $x_{\text{new}}$  between  $x_{\text{rand}}$  and  $x_{\text{near}}$ , a distance  $\Delta x$  from  $x_{\text{near}}$  (considering obstacles); `add` adds this node to the tree. In line 4 the `count` and `parent` of  $x_{\text{new}}$  are defined. If  $x_{\text{new}}$  is inside an observation area that is in  $\mathcal{T}.\text{areas}$ , the node is added to the set of observation nodes and the area is stored as  $x_{\text{new}}.\text{area}$  (lines 5–6). Lines 1–7 repeat until

---

**Algorithm 1** RRT  $\leftarrow$  RRT( $\mathcal{T}, \Delta x, N, K$ )

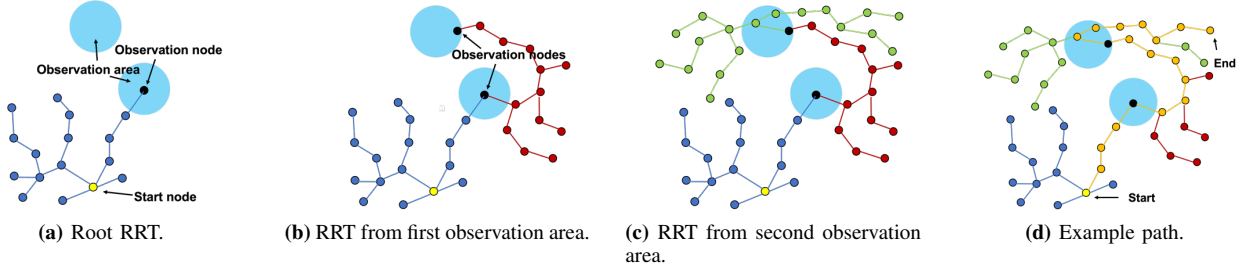
---

```

1: repeat
2:    $x_{\text{rand}} \leftarrow \text{random\_state}()$ ;  $x_{\text{near}} \leftarrow \text{nearest}(x_{\text{rand}}, \mathcal{T})$ 
3:    $x_{\text{new}} \leftarrow \text{new\_node}(x_{\text{near}}, x_{\text{rand}}, \Delta x)$ ;  $\mathcal{T}.\text{add}(x_{\text{new}})$ 
4:    $x_{\text{new}}.\text{count} \leftarrow x_{\text{near}}.\text{count} + 1$ ;  $x_{\text{new}}.\text{parent} \leftarrow x_{\text{near}}$ 
5:   if  $x_{\text{new}}$  in an area in  $\mathcal{T}.\text{areas}$  then
6:      $\mathcal{T}.\text{x\_obs}.\text{append}(x_{\text{new}})$ ;  $x_{\text{new}}.\text{area} \leftarrow \text{area}$ 
7: until  $N$  observations or  $K$  nodes

```

---



**Fig. 3:** The MORRT expansion for  $N = 1$ . The MORRT expands from left to right. Blue, red, and green nodes belong to three separate RRTs. The yellow path to the right shows one possible path in the MORRT.

the tree  $\mathcal{T}$  has  $N$  observation nodes—nodes in observation areas—or  $K$  total nodes, whichever happens first.<sup>1</sup>

The MORRT is illustrated in Algorithm 2, which constructs a tree of RRTs where the nodes are observation nodes inside of observation areas, and the edges are standard RRTs that extend between these observation nodes. In line 1 the count of  $x_{\text{init}}$  is set to zero. In line 2 a “root RRT” is initialized at  $x_{\text{init}}$ , with the set of observation nodes initially empty, and the set of (available observation) areas initialized as  $\text{areas}$ . In line 3 the root RRT is extended with Algorithm 1. Once the root RRT has been expanded it is placed in a set of  $\text{parents}$  (line 4). From each observation node in the parent RRT (line 6), a “child RRT” is initialized with an empty set of observation nodes (line 7). The child RRT’s set of available observation areas is the parent’s  $\text{areas}$  minus the observation area of  $x_{\text{obs}}$  (line 8); with this construction, the further up in the tree of RRTs we get, the fewer observation areas become available to explore, until eventually no more areas will be explored following line 5 in Algorithm 1. The child RRT is expanded with the RRT function (line 9). Finally, the child RRT is appended to the set of parents (and considered as a parent in a following iteration) and defined as a child of the parent (line 10). Lines 5–10 repeat until there is no parent RRT with observation nodes, i.e., until  $\mathcal{T}_{\text{parent}}.x_{\text{obs}}$  in line 6 is empty for all remaining parents. Note that the algorithm will terminate since the set  $\mathcal{T}.\text{areas}$  shrinks as the depth in the tree of RRTs increases, and there is only a finite number of observation areas. Eventually  $\mathcal{T}.x_{\text{obs}}$  will be empty for all  $\mathcal{T}$  deep enough in the tree of RRTs, following lines 7–8 of Algorithm 2 and lines 5–6 of Algorithm 1.

Figure 3 illustrates Algorithm 2 with  $N = 1$ . An RRT is expanded until it reaches an observation area (Figure 3a). Since  $N = 1$  a new RRT is immediately initialized from this observation node and expanded until it reaches the second observation area (Figure 3b). Once all observation areas have been visited, a final RRT of fixed size is expanded from the most recently visited observation area (Figure 3c), which follows from the fact that line 5 of Algorithm 1 will never hold for this RRT, so the tree will expand to  $K$  nodes. Figure 3d shows a possible path through the MORRT; however, all paths ending at a childless node are possible.

**Selecting the best plan from the MORRT.** The MORRT

<sup>1</sup>RRT\* [37] can be leveraged by introducing a *cost heuristic* for a path, like the cost of committing to this path regardless of observations.

---

**Algorithm 2**  $\mathcal{T}_{\text{root}} \leftarrow \text{MORRT}(x_{\text{init}}, \Delta x, N, K, \text{areas})$

---

```

1:  $x_{\text{init}}.\text{count} \leftarrow 0$ 
2:  $\mathcal{T}_{\text{root}}.\text{init}(x_{\text{init}})$ ;  $\mathcal{T}_{\text{root}}.x_{\text{obs}} \leftarrow \{\}$ ;  $\mathcal{T}_{\text{root}}.\text{areas} \leftarrow \text{areas}$ 
3:  $\mathcal{T}_{\text{root}} \leftarrow \text{RRT}(\mathcal{T}_{\text{root}}, \Delta x, N, K)$  //ALGO 1
4:  $\text{parents} = \{\mathcal{T}_{\text{root}}\}$ 
5: for each  $\mathcal{T}_{\text{parent}}$  in  $\text{parents}$  do
6:   for each  $x_{\text{obs}}$  in  $\mathcal{T}_{\text{parent}}.x_{\text{obs}}$  do
7:      $\mathcal{T}_{\text{child}}.\text{init}(x_{\text{obs}})$ ;  $\mathcal{T}_{\text{child}}.x_{\text{obs}} \leftarrow \{\}$ 
8:      $\mathcal{T}_{\text{child}}.\text{areas} \leftarrow \mathcal{T}_{\text{parent}}.\text{areas} \setminus \{x_{\text{obs}}.\text{area}\}$ 
9:      $\mathcal{T}_{\text{child}} \leftarrow \text{RRT}(\mathcal{T}_{\text{child}}, \Delta x, N, K)$  //ALGO 1
10:     $\text{parents}.\text{append}(\mathcal{T}_{\text{child}})$ ;  $\mathcal{T}_{\text{parent}}.\text{child} \leftarrow \mathcal{T}_{\text{child}}$ 

```

---

(Algorithm 2) generates a tree of RRTs as shown in Figure 3. A mission plan is a sub-tree of RRTs, where each RRT originates from an observation node and has at most  $|\mathcal{O}|$  branches, as shown in Figure 2. The objective now is to find the sub-tree  $\mathcal{I}_{\pi}$  of RRTs that minimizes the cost (4), which is done with dynamic programming on the tree of RRTs. We start from the latest trees in the tree of RRTs and evaluate the cost (4). Then, moving backward we compute the cost in a dynamic programming manner through the tree of RRTs, while storing the best plan from each observation node.

### C. Multi-agent extension

The only modification in the multi-agent case is how the RRT is expanded. In the multi-agent extension, the RRT function, lines 3 and 9 of Algorithm 2, is substituted with Algorithm 3—a multi-agent extension of the RRT. In lines 1–2 of Algorithm 3  $M$  separate RRTs are generated with the RRT function (Algorithm 1). The collection of the  $M$  agent’s RRTs defines the *multi-agent RRT (MA-RRT)* (line 3). All combinations of the agent’s observation nodes are stored in an  $\text{observations}$  set (line 4). This set is modified in lines 5–9 to make all agent’s paths to the observation the same length. Each combination  $x_{\text{obs}}$  of observation nodes (line 5) consists of  $M$  single-agent observation nodes. We define

---

**Algorithm 3**  $\text{MA-RRT} \leftarrow \text{MA-RRT}(M, x_{\text{init}}, \Delta x, N, K)$

---

```

1: for  $a = 1, 2, \dots, M$  do
2:    $\mathcal{T}_a \leftarrow \text{RRT}(x_{\text{init}}[a], \Delta x, N, K)$  //ALGO 1
3:  $\text{MA-RRT} \leftarrow \{\mathcal{T}_1, \dots, \mathcal{T}_M\}$ 
4:  $\text{observations} \leftarrow \mathcal{T}_1.x_{\text{obs}} \times \mathcal{T}_2.x_{\text{obs}} \times \dots \times \mathcal{T}_M.x_{\text{obs}}$ 
5: for each  $x_{\text{obs}}$  in  $\text{observations}$  do
6:    $L \leftarrow \min\{x.\text{count} \text{ for } x \in x_{\text{obs}}\}$ 
7:   for each  $x$  in  $x_{\text{obs}}$  do
8:     while  $x.\text{count} > L$  do
9:        $x \leftarrow x.\text{parent}$ 
10:  $\text{MA-RRT}.x_{\text{obs}} \leftarrow \text{observations}$ 

```

---

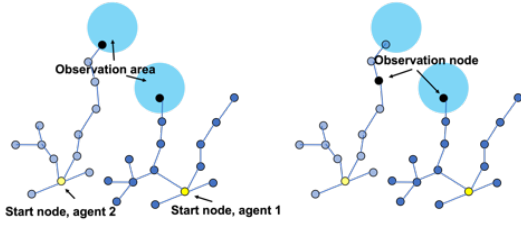


Fig. 4: Illustration of lines 5–9 of Algorithm 3

$L$  as the minimum `count` variable of these nodes (line 6) (recall that, for an agent  $a$ ,  $x.count$  denotes the distance from  $x_{init}[a]$  to  $x$  in  $\mathcal{T}_a$ ). Line 7 loops through all observation nodes  $x \in x\_obs$ , and lines 8–9 replace  $x$  with its parent node until  $x.count$  equals  $L$ . The MA-RRT observation nodes is the set of modified observations (line 10). Now, all agents have the same path length to the observations.

Figure 4 illustrates lines 5–10 of Algorithm 3 for  $M = 2$  (two agents) and  $N = 1$ . The observation node of agent 2 is moved two nodes upwards in the path, to make both agent’s paths from start to the observation node the same length.

**Complexity.** The MORRT complexity can be expressed in terms of the number of RRTs that are built. Let  $A$  denote the number of observation areas. The MORRT builds a tree of RRTs, and the number of branches originating from each node is at most  $N^M$ , where  $N$  is a parameter of Algorithm 2. Hence, the number of RRTs needed for an MORRT is at most  $(N^M)^A = N^{M \times A}$ , which is a function of the number of agents  $M$ , the number of observation areas  $A$ , and the parameter  $N$ . The computational time can be reduced by, for each agent, only keeping one observation node in each observation area before taking the Cartesian product in line 4 of Algorithm 3; the complexity would then be  $A^{M \times A}$ , which reduces the computational time when  $A < N$ . To reduce complexity in large systems we can assign different observation areas to different agents. This can be done randomly over several rounds, before selecting the best assignment, or deterministically if a good assignment is known beforehand. We illustrate this in Section IV-B.

#### IV. RESULTS

This section presents the main results: simulations for single and multiple agents, and an implementation on hardware.<sup>2</sup> We use quadratic costs, i.e.,  $h(x^a, \Delta x^a, e) = \|x - x_{g,e}\|_Q^2$ ,  $h_N(x^a, e_N) = \|x^a - x_{g,e}\|_{Q_N}$ , where  $\|x\|_Q^2 = x^T Q x$ , and  $x_{g,e}$  is the goal node corresponding to state  $e$ .

##### A. Single-agent simulation and benchmark comparison

We first compare the MORRT to a standard RRT benchmark algorithm on a single-agent example. For the RRT benchmark, the agent commits to a single goal (the most likely one) using a standard RRT, and if the agent crosses an observation area, the belief is updated and the agent recommits to the new best target.

The agent is initialized under an obstacle in a square environment (Figure 5). At the bottom there is a hill where the agent can see above the obstacle and receive an 80%

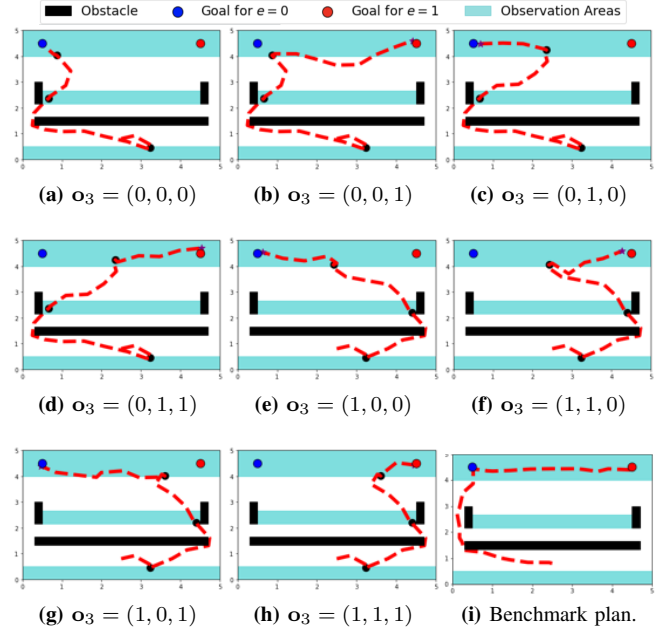


Fig. 5: Plan for an agent initiated below the obstacle. Black dots represent observation nodes.

accurate observation. On the other side of the obstacle there are two more obstacles; in-between these the agent can view both goal nodes and receive a 90% accurate observation. Finally, at the top, once the agent is near the goal nodes, it receives a perfect observation. The initial belief is  $[0.5, 0.5]$ , meaning that the probabilities of the target being in the top left and top right are both 50%.

Figures 5a-5h show the MORRT plan for the eight possible realizations of a mission; observation nodes are illustrated by black dots. The agent first moves down to collect an observation. If the first observation is  $o_1 = 0$  it goes around the obstacle to the left (Figures 5a-5d); otherwise it goes around to the right (Figures 5e-5h). The agent then moves in-between the top obstacles to collect another observation, before moving either left (if  $o_2 = 0$ ) or right (if  $o_2 = 1$ ). Finally, at the top observation area the agent receives a perfect observation and commits left or right.

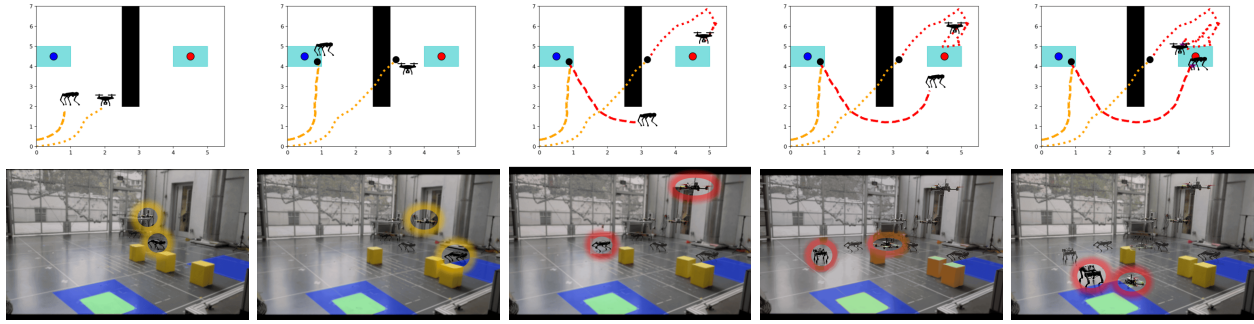
Figure 5i shows the plan generated by the standard RRT benchmark algorithm. The agent first commits to one target (the top left in this case) and misses the bottom and middle observation areas. Once at the top left target, the agent receives a perfect observation; if this observation is  $o = 0$ , it stays at the top left; otherwise it must travel to the top right.

Table I shows a quantitative comparison between the MORRT and the RRT benchmark. Note that here the expected-, best-, and worst-case costs are with respect to the noisy observations in the mission, not the randomness in the RRT and MORRT algorithms; the algorithms only ran once, as the results tend to be consistent.

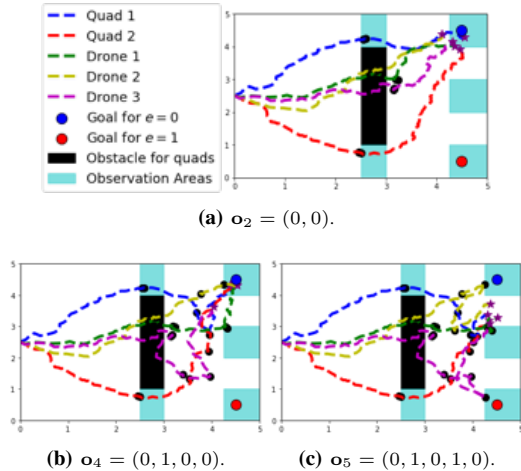
TABLE I: MORRT compared to the standard RRT benchmark.

	Expected cost	Best cost	Worst cost	Time [s]
MORRT	<b>2200</b>	1560	3980	200
Benchmark	2890	<b>1270</b>	<b>3760</b>	<b>30</b>

<sup>2</sup>Code available at <https://github.com/kasperjo/MixedObservableRRT.git>.



**Fig. 6:** Experimental results from the setup in Figure 1. The yellow obstacles in the bottom row move but are always contained inside the static black obstacle region in the top row. Blue and red goal nodes correspond to  $e = 0$  and  $e = 1$ , respectively. The black dots illustrate an observation node.



**Fig. 7:** Mission plan for two quadrupeds and three drones. Purple stars illustrate the end of a mission, and black dots represent observation nodes.

The realized cost of the best and worst case scenarios are lower for the benchmark. This is because the benchmark only makes a single observation, and takes the shortest path there, while the MORRT visits all observation areas to reduce uncertainty. The expected cost is much lower for the MORRT than the benchmark. The computational time for the MORRT is higher than for the benchmark, as expected, since the benchmark is a standard RRT, while the MORRT finds a tree of RRTs, where the chosen path depends on the observation.

### B. Multi-agent simulation

Figure 7 illustrates the system. Two quadrupeds and three drones start at  $(x, y) = (0, 2.5)$ . The quadrupeds must avoid the obstacle. There are five observation areas, and the initial belief is  $[0.5, 0.5]$ . We assign the quadrupeds to the areas close to the obstacle, and the drones to the remaining three areas; this reduces the computational time as mentioned in Section III-C. Observation areas around the goal nodes yield 90% accurate observations, and other observations are 80% accurate. The plan is illustrated in Figure 7. There are six outcomes, but due to symmetry we show the results for when the agents commit to the top goal node. Observation nodes, where agents make and communicate observations, are illustrated by black dots. The quadrupeds go around the obstacle to collect one observation each. If the observations agree, all robots go to the target (Figure 7a). If the observations contradict, two drones make additional observations, one in

the top right and one in the middle. If these observations align, the robots commit to the target (Figure 7b); otherwise the third drone makes an additional observation in the bottom right before the robots commit (Figure 7c).

### C. Hardware experiment

1) *Experimental setup and hardware models:* The setup is illustrated in Figure 1. A quadruped and a drone are initiated in the origin of a  $5\text{m} \times 5\text{m}$  environment, and the mission objective is to locate the science sample. The initial belief is  $[0.5, 0.5]$  and observations are perfect.

We fix the  $z$ -coordinate of the drone and compute a plan in  $\mathbb{R}^2$ . We then model the agents as unicycles [38] and leverage two MPC algorithms to make them follow MORRT way-points. The plan is computed offline, while the MPC algorithms act locally online. We plan for all possible observations, but during the experiment we fix the outcome and feed observations through a centralized information flow. At 60Hz the MPC algorithms solve for velocity and yaw-rate inputs, which are fed to the quadruped and drone.

2) *Experimental results:* The experiment is illustrated in Figure 6. The quadruped observes  $o_1 = 1$  in the second snapshot, and communicates to the drone; recall that we fix the observation beforehand, but a full plan is computed. Since the observation is perfect, both agents move to the top right goal—taking into account the obstacle for the quadruped—without making another observation. The drone arrives first and waits for the quadruped to arrive.

## V. CONCLUSION

We introduced the MORRT, which plans a mission for heterogeneous multi-agent exploration, and showed how autonomy can be enabled when agents move and act simultaneously. The approach was illustrated in simulation and on hardware. We found that agents made intelligent collaborative decisions based on environmental observations. The problem formulated in this paper can be applied to real-world settings, including search missions, like the Mars exploration task or rescue operations, where the objective is to locate a hidden target. There are several directions for future work, including extending the MORRT to account for system dynamics, adding communication constraints to mimic a distributed setting, incorporating collision constraints, and analyzing convergence properties.

## REFERENCES

- [1] P. Nilsson, S. Haesaert, R. Thakker, K. Otsu, C. Vasile, A. akbar Aghamohammadi, R. Murray, and A. Ames, "Toward specification-guided active mars exploration for cooperative robot teams," in *Robotics: Science and Systems*, 2018.
- [2] U. Rosolia, Y. Chen, S. Daftry, M. Ono, Y. Yue, and A. D. Ames, "The mixed-observable constrained linear quadratic regulator problem: the exact solution and practical algorithms," 2021.
- [3] G. Lakemeyer and B. Nebel, Eds., *Exploring Artificial Intelligence in the New Millennium*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003.
- [4] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. Cambridge, Mass.: MIT Press.
- [5] C. Stachniss, *Robotic Mapping and Exploration*. Berlin: Springer, 2009.
- [6] P. J. Goulart, E. C. Kerrigan, and J. M. Maciejowski, "Optimization over state feedback policies for robust control with constraints," *Automatica*, vol. 42, no. 4, pp. 523–533, 2006.
- [7] L. Chisci, J. Rossiter, and G. Zappa, "Systems with persistent disturbances: predictive control with restricted constraints," *Automatica*, vol. 37, no. 7, pp. 1019–1028, 2001.
- [8] D. Mayne, M. Seron, and S. Raković, "Robust model predictive control of constrained linear systems with bounded disturbances," *Automatica*, vol. 41, no. 2, pp. 219–224, 2005.
- [9] S. Yu, C. Maier, H. Chen, and F. Allgöwer, "Tube MPC scheme based on robust control invariant set with application to Lipschitz nonlinear systems," *Systems & Control Letters*, vol. 62, no. 2, pp. 194–200, 2013.
- [10] J. Fleming, B. Kouvaritakis, and M. Cannon, "Robust tube MPC for linear systems with multiplicative uncertainty," *IEEE Transactions on Automatic Control*, vol. 60, no. 4, pp. 1087–1092, 2015.
- [11] Y.-S. Wang, N. Matni, and J. C. Doyle, "A system-level approach to controller synthesis," *IEEE Transactions on Automatic Control*, vol. 64, no. 10, pp. 4079–4093, 2019.
- [12] A. Liniger, X. Zhang, P. Aeschbach, A. Georghiou, and J. Lygeros, "Racing miniature cars: Enhancing performance using stochastic MPC and disturbance feedback," in *2017 American Control Conference (ACC)*, 2017, pp. 5642–5647.
- [13] Y. Chen, U. Rosolia, W. Ubellacker, N. Csomay-Shanklin, and A. D. Ames, "Interactive multi-modal motion planning with branch model predictive control," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 5365–5372, 2022.
- [14] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski, "Adjustable robust solutions of uncertain linear programs," *Mathematical Programming*, vol. 99, pp. 351–376, 2004.
- [15] D. Mayne, S. Raković, R. Findeisen, and F. Allgöwer, "Robust output feedback model predictive control of constrained linear systems," *Automatica*, vol. 42, no. 7, pp. 1217–1222, 2006.
- [16] I. Alvarado, D. Limon, T. Alamo, and E. Camacho, "Output feedback robust tube based MPC for tracking of piece-wise constant references," in *2007 46th IEEE Conference on Decision and Control*, 2007, pp. 2175–2180.
- [17] M. Cannon, Q. Cheng, B. Kouvaritakis, and S. V. Raković, "Stochastic tube MPC with state estimation," *Automatica*, vol. 48, no. 3, pp. 536–541, 2012.
- [18] Y. Chen, U. Rosolia, and A. D. Ames, "Decentralized task and path planning for multi-robot systems," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4337–4344, 2021.
- [19] S. LaValle and J. Kuffner, "Randomized kinodynamic planning." *I. J. Robotic Res.*, vol. 20, pp. 378–400, 2001.
- [20] C. Fulgenzi, C. Tay, A. Spalanzani, and C. Laugier, "Probabilistic navigation in dynamic environment using rapidly-exploring random trees and gaussian processes," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008, pp. 1056–1062.
- [21] J. Kim and J. Ostrowski, "Motion planning a aerial robot using rapidly-exploring random trees with dynamic constraints," in *IEEE International Conference on Robotics and Automation*, vol. 2, 2003, pp. 2200–2205.
- [22] M. Kothari and I. Postlethwaite, "A probabilistically robust path planning algorithm for UAVs using rapidly-exploring random trees," *Journal of Intelligent & Robotic Systems*, vol. 71, 2013.
- [23] S. Karaman and E. Frazzoli, "Incremental sampling-based algorithms for optimal motion planning," *CoRR*, vol. abs/1005.0416, 2010. [Online]. Available: <http://arxiv.org/abs/1005.0416>
- [24] S. Karaman, M. R. Walter, A. Perez, E. Frazzoli, and S. Teller, "Anytime motion planning using the RRT\*," in *IEEE International Conference on Robotics and Automation*, 2011, pp. 1478–1483.
- [25] W. Burgard, M. Moors, D. Fox, R. Simmons, and S. Thrun, "Collaborative multi-robot exploration," in *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065)*, vol. 1, 2000, pp. 476–481 vol.1.
- [26] K. M. Wurm, C. Stachniss, and W. Burgard, "Coordinated multi-robot exploration using a segmentation of the environment," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008, pp. 1160–1165.
- [27] K. M. Wurm, C. Dornhege, P. Eyerich, C. Stachniss, B. Nebel, and W. Burgard, "Coordinated exploration with marsupial teams of robots using temporal symbolic planning," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 5014–5019.
- [28] W. Burgard, M. Moors, C. Stachniss, and F. Schneider, "Coordinated multi-robot exploration," *IEEE Transactions on Robotics*, pp. 376–386, 2005.
- [29] M. Jager and B. Nebel, "Dynamic decentralized area partitioning for cooperating cleaning robots," in *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*, 2002, pp. 3577–3582 vol.4.
- [30] D. Kurabayashi, J. Ota, T. Arai, and E. Yoshida, "Cooperative sweeping by multiple mobile robots," in *Proceedings of IEEE International Conference on Robotics and Automation*, 1996, pp. 1744–1749 vol.2.
- [31] K. Singh and K. Fujimura, "Map making by cooperating mobile robots," in *[1993] Proceedings IEEE International Conference on Robotics and Automation*, 1993, pp. 254–259 vol.2.
- [32] A. Stroupe, R. Ravichandran, and T. Balch, "Value-based action selection for exploration and dynamic target observation with robot teams," in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, 2004, pp. 4190–4197 Vol.4.
- [33] K. Albina and S. G. Lee, "Hybrid stochastic exploration using grey wolf optimizer and coordinated multi-robot exploration algorithms," *IEEE Access*, vol. 7, pp. 14 246–14 255, 2019.
- [34] F. Gul, I. Mir, W. Rahiman, and T. U. Islam, "Novel implementation of multi-robot space exploration utilizing coordinated multi-robot exploration and frequency modified whale optimization algorithm," *IEEE Access*, vol. 9, pp. 22 774–22 787, 2021.
- [35] F. Gul, S. Mir, and I. Mir, "Coordinated multi-robot exploration: Hybrid stochastic optimization approach," in *AIAA SCITECH 2022 Forum*, 2022, p. 1414.
- [36] V. Krishnamurthy, *Partially Observed Markov Decision Processes: From Filtering to Controlled Sensing*. Cambridge University Press, 2016.
- [37] S. Karaman and E. Frazzoli, "Sampling-based algorithms for optimal motion planning," *The International Journal of Robotics Research*, vol. 30, no. 7, pp. 846–894, 2011.
- [38] M. Aicardi, G. Casalino, A. Bicchi, and A. Balestrino, "Closed loop steering of unicycle like vehicles via Lyapunov techniques," *IEEE Robotics Automation Magazine*, vol. 2, no. 1, pp. 27–35, 1995.