MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Bandit-based multi-agent search under noisy observations

Thaker, Parth; Di Cairano, Stefano; Vinod, Abraham P.

TR2023-085 July 11, 2023

Abstract

Autonomous search using teams of multiple agents need tractable coordination strategies between the search agents. The strategy must lower the time to identify interesting areas in the search environment, lower the costs/energy usage by the search agents during movement and sensing, and be resilient to the noise present in the sensed data due to the use of low-cost and low-weight sensors. We propose a data-driven, multi-agent search algorithm to achieve these goals using the framework of thresholding multi-armed bandits. For our algorithm, we also provide finite upper bounds on the time taken to complete the search, on the time taken to label all interesting cells, and on the economic costs incurred during the search.

World Congress of the International Federation of Automatic Control (IFAC) 2023

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

Bandit-based multi-agent search under noisy observations ⋆

Parth Thaker * **Stefano Di Cairano** ** **Abraham P. Vinod** **

* *Arizona State University, Tempe, AZ, USA 85281.*
** *Mitsubishi Electric Research Laboratories, Cambridge, MA, USA 02139.*

Abstract: Autonomous search using teams of multiple agents need tractable coordination strategies between the search agents. The strategy must lower the time to identify interesting areas in the search environment, lower the costs/energy usage by the search agents during movement and sensing, and be resilient to the noise present in the sensed data due to the use of low-cost and low-weight sensors. We propose a data-driven, multi-agent search algorithm to achieve these goals using the framework of thresholding multi-armed bandits. For our algorithm, we also provide finite upper bounds on the time taken to complete the search, on the time taken to label all interesting cells, and on the economic costs incurred during the search.

Keywords: Adaptive sensing, Monitoring, Multi-armed bandits, Environmental applications

1. INTRODUCTION

Autonomous multi-agent search for objects/phenomena of interest over large areas are crucial in several applications, including environmental monitoring, agriculture, search-and-rescue, and wildlife monitoring. Given a grid environment to search, we study the problem of identifying *all* interesting cells (cells that contain an object/phenomenon of interest) using multiple search agents, each equipped with a noisy sensor. We require the search agents to satisfy multiple requirements. First, the search agents must coordinate and quickly identify interesting cells, which is essential in time-sensitive applications like search-and-rescue. Additionally, they must minimize economic costs associated with the search, which could include the energy used by the search agents due to movement and sensing. Finally, they must make decisions on locations to sense based on noisy observations obtained online from low-cost and low-weight sensors typical of such search systems. In this paper, we propose *a data-driven, multi-agent search algorithm that addresses these requirements using the framework of thresholding bandits.*

A solution to the search problem is the *label-then-move search* (see (Rolf et al., 2021) for a variant of this search). In this search strategy, we partition the search space into disjoint sets of grid cells that are assigned to the search agents. Each search agent starts at some cell within its assigned set of cells, collects enough data at a grid cell until it is confident enough to label the grid cell as interesting or uninteresting, and then moves on to another grid cell within its assigned set. The label-then-move search strategy ignores the data collected online to decide on the next location to sense. Consequently, it can spend a significant amount of time in labeling uninteresting cells and may not be well suited for time-sensitive applications.

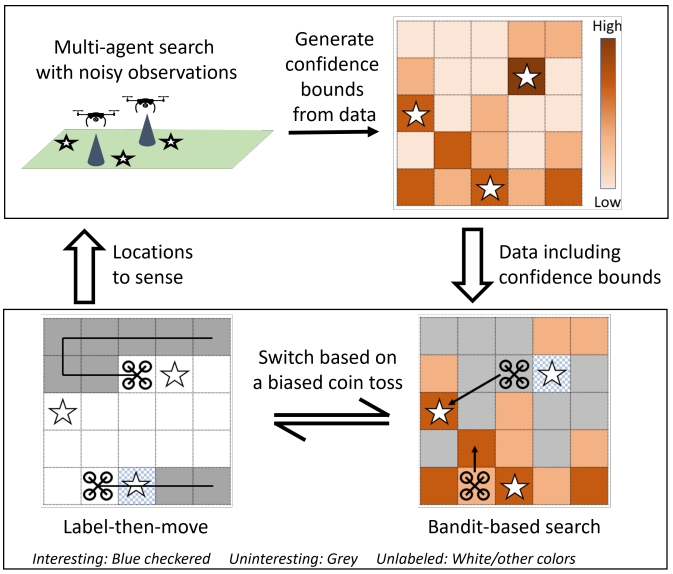

Fig. 1. Data-driven multi-agent search under noisy observations. The proposed approach switches between a bandit-based search and a *label-then-move search* with a user-specified probability. The bandit-based search optimizes a surrogate function constructed using noisy observations for making decisions on locations to sense. The *label-then-move search* makes the agents follow a fixed, pre-determined pattern independent of the data collected online.

Recently, multi-armed bandit (MAB) have also been proposed for the multi-agent search problem. Recall that MAB is a special class of reinforcement learning algorithms where the current actions *do not* impact future reward (Lattimore and Szepesvári, 2020). MAB-based algorithms typically enjoy non-asymptotic guarantees of performance with minimal assumptions, unlike general reinforcement learning algorithms (Lattimore and Szepesvári,

⋆ Corresponding author: vinod@merl.com

2020). (Rolf et al., 2021; Du et al., 2021) propose MAB-based search strategies that identify the maximal or top-k interesting cells in a grid, and require prior knowledge of the number of interesting cells. Instead, our work focuses on identifying *all* interesting grid cells without the prior knowledge of the total number of interesting cells.

In this paper, we combine the label-then-move search and thresholding MAB-based search strategies, as illustrated in Figure 1. Similarly to (Locatelli et al., 2016; Mason et al., 2020), we decide on the next locations to sense by maximizing a surrogate function that is constructed using the data collected online. While a vanilla application of the thresholding bandit-based approach is near-optimal in terms of search completion time (Locatelli et al., 2016), it may have high economic costs. A key feature of our approach is the ability to trade-off the desire for efficient search with the need to lower the economic costs associated with the search.

Several other strategies have also been proposed for multi-agent search (Drew, 2021; Queralta et al., 2020). Popular approaches include algorithms based on submodular maximization (Krause and Guestrin, 2007), algorithms combining Voronoi-based search (Bullo et al., 2009) with function approximation (Schwager et al., 2015; Luo et al., 2019), active sensing/perception algorithms (Bajcsy et al., 2018), graph-based search algorithms (Kapoutsis et al., 2017; Best et al., 2018), and algorithms based on statistical learning (Marchant and Ramos, 2012; Ghods et al., 2021; Banerjee et al., 2022). However, these works may require perfect sensing, may not have finite-time guarantees on the search performance, and/or may have high economic costs of search associated with movement and sensing.

The main contributions of this paper are: 1) to propose a multi-agent search algorithm that accommodates noisy observation data using a combination of thresholding MAB and label-then-move search, and 2) to characterize the performance of the algorithm by determining finite upper bounds on the time taken to complete the search, time taken to label all interesting cells, and the economic costs incurred during the search by our algorithm. We propose two metrics, *priority labeling time* and *economic cost*, to study the performance of the proposed algorithm. Additionally, with respect to the existing literature in MAB, we integrate coordination requirements and the physical limitations of switching actions directly into the algorithm. Finally, we demonstrate the efficacy of our approach in numerical simulations.

2. PROBLEM FORMULATION

Let $\mathcal{G}$ denote the set of grid cells defining the search environment. Let the autonomous search team have d agents, each equipped with a noisy sensor. In the autonomous multi-agent search problem, we must identify the sequence of grid cells to visit and sense based on the noisy sensor data collected online, and return a set of interesting grid cells. We refer to a cell as *interesting* if it contains an object/phenomenon of interest. In this section, we formalize the autonomous multi-agent search problem as a variant of the thresholding MAB, characterize suitable metrics to analyze the performance of a search algorithm, and state the problems addressed in this paper.

MAB formulation of the multi-agent search problem: We cast the search problem as a $|\mathcal{G}|$-armed bandit problem where the arms are grid cells with $\mathcal{G} \triangleq \{1, 2, 3 \ldots, |\mathcal{G}|\}$, and $|\mathcal{G}|$ denotes the cardinality of the set $\mathcal{G}$. Let $\mathcal{S}^{\complement} = \mathcal{G} \setminus \mathcal{S}$ be the complement of any set $\mathcal{S} \subseteq \mathcal{G}$.

At each time step, the search team of d agents selects a set of d *distinct* grid cells to visit simultaneously. Each visit to a grid cell returns a binary indication of whether the cell is interesting. However, the observation data may be corrupted by noise, arising from sensing limitations and perception errors. Formally, a visit to grid cell $i \in \mathcal{G}$ results in a draw of a sample from a corresponding Bernoulli random variable ν_i with mean μ_i. The mean μ_i is influenced by the underlying spatial distribution of the interesting cells, and the characteristics of the noisy sensors and the perception algorithms used by the agents. We assume that the Bernoulli random variables for any two cells in $\mathcal{G}$ are mutually independent.

Remark 1. We do not assume prior knowledge of μ_i for any grid cell $i \in \mathcal{G}$ or the total number of interesting cells.

Desired outcome of the search: For a user-specified *threshold* $\theta \in (0, 1)$, upon the completion of the search, we seek to identify the set of grid cells,

$$\mathcal{S}_\theta = \{i \in \mathcal{G} | \mu_i \geq \theta\} \subset \mathcal{G}. \tag{1}$$

The set $\mathcal{S}_\theta$ is the set of grid cells that may be sensed as interesting with a probability of at least θ.

We make the following assumption to obtain finite-sample guarantees despite the noisy sensors on the search agents.

Assumption 1. (Labeling error tolerance). The labeling error for grid cells $i \in \mathcal{G}$ with $\mu_i \in (\theta - \epsilon, \theta + \epsilon)$ may be ignored for some (small) tolerance $\epsilon > 0$.

Assumption 1 is motivated by the observation that deciding if $|\mu_i - \theta| > \epsilon$ for any grid cell $i \in \mathcal{G}$ using a finite number of samples becomes harder as μ_i approaches θ (Jun et al., 2016a; Lattimore and Szepesvári, 2020). Under Assumption 1, any set $\mathcal{K} \subseteq \mathcal{G}$ that satisfies

$$\mathcal{S}_{\theta+\epsilon} \subseteq \mathcal{K} \subseteq \mathcal{S}_{\theta-\epsilon}, \tag{2}$$

is an acceptable approximation of $\mathcal{S}_\theta$.

We use the notion of a *search policy* to characterize a multi-agent search.

Definition 1. (Search policy). Let $\mathcal{H}(\tau) = \{\mathcal{H}_i(\tau)\}_{i \in \mathcal{G}}$, where $\mathcal{H}_i(\tau)$ is the history of observations at grid cell $i \in \mathcal{G}$ collected by all agents until time τ. Let $\pi_\tau : \mathcal{H}(\tau) \to \mathcal{G}^d$ be a function that maps $\mathcal{H}(\tau)$ to d distinct grid cells in $\mathcal{G}$. We define a search policy $\boldsymbol{\pi}_t$ as a sequence of functions $\boldsymbol{\pi}_t = \{\pi_\tau\}_{0 \leq \tau \leq t}$. We will drop the subscript on $\boldsymbol{\pi}_t$ when time is not relevant.

Performance metrics: Let the multi-agent search using the search policy $\boldsymbol{\pi}$ terminate at time step $T_{\boldsymbol{\pi}} \in \mathbb{N}$, and return a sequence of sets $\{\mathcal{K}(t)\}_{t=0}^{T_{\boldsymbol{\pi}}}$ with $\mathcal{K}(t) \subseteq \mathcal{G}$, $\forall t$. A successful search policy $\boldsymbol{\pi}$ has a low *labeling error* upon termination, i.e., it satisfies

$$\mathbb{P}\left[(\mathcal{S}_{\theta+\epsilon} \setminus \mathcal{K}(T_{\boldsymbol{\pi}})) \cup (\mathcal{K}(T_{\boldsymbol{\pi}}) \setminus \mathcal{S}_{\theta-\epsilon}) = \emptyset\right] \geq 1 - \delta, \tag{3}$$

for some user-specified *labeling error probability* $\delta \in (0, 1)$. Here, (3) enforces (2) by requiring that $\mathcal{K}(T_{\boldsymbol{\pi}})$ includes (almost) every one of the interesting cells and excludes

(almost) every one of the uninteresting cells upon termination, with probability $1-\delta$.

Additionally, the search must have:

(1) low *priority labeling time* $L(\boldsymbol{\pi})$,
$$L(\boldsymbol{\pi}) = \inf\{t \leq T_{\boldsymbol{\pi}} : \mathbb{P}[\mathcal{S}_{\theta+\epsilon} \setminus \mathcal{K}(t) \neq \emptyset] \leq \delta]\}, \quad (4)$$
i.e., the search identify (almost) every one of the interesting cells quickly. By definition, $L(\boldsymbol{\pi}) \leq T_{\boldsymbol{\pi}}$.

(2) low *economic cost* upon termination, i.e., the search has low costs associated with movement and sensing,
$$E(\boldsymbol{\pi}) = \sum_{t=1}^{T_{\boldsymbol{\pi}}} \left(\underbrace{\ell(\mathbf{a}_t, \mathbf{a}_{t-1})}_{\text{movement cost}} + \underbrace{\beta d}_{\text{sensing cost}} \right) \quad (5)$$
where $\mathbf{a}_t$ is the set of d grid cells being sampled at time t according to search policy $\boldsymbol{\pi}_{t-1}$, $\ell : \mathcal{G}^d \times \mathcal{G}^d \to \mathbb{R}$ is a metric on $\mathcal{G}^d$, and $\beta > 0$ is a known constant sensing cost for each agent. Consequently, βd is the sensing cost for the team at each time step.

We pursue probabilistic performance metrics in (3) and (4) due to the uncertainty in sensing and perception.

We now state the two problems tackled by this paper:

Problem 1. Design a multi-agent search algorithm on $\mathcal{G}$ that simultaneously satisfies the criteria (3), (4), and (5).

Problem 2. Determine upper bounds on the time to terminate the search (3), the priority labeling time (4), and the economic cost (5) for the proposed solution to Problem 1.

3. PROPOSED SOLUTION

Algorithm 1 describes the proposed solution for Problem 1. It augments label-then-move search with a thresholding bandit-based search (inspired from (Locatelli et al., 2016)) to satisfy the criteria in (3), (4), and (5). In Algorithm 1, the keep set $\mathcal{K}(t) \subseteq \mathcal{G}$ and the reject set $\mathcal{R}(t) \subseteq \mathcal{G}$ are the sets of grid cells labeled as *interesting* and *uninteresting* respectively, at the time instant t.

Algorithm 1 runs in a loop until all grid cells in $\mathcal{G}$ are assigned to $\mathcal{K}(t)$ or $\mathcal{R}(t)$ (or both). Each loop starts with a toss of a biased coin with the bias set to the aggressiveness parameter, $\alpha \in (0,1)$.

When the current toss of the biased coin returns heads, we use *upper confidence bounds* typical of bandit-based algorithms (Locatelli et al., 2016; Mason et al., 2020; Lattimore and Szepesvári, 2020) to sample the unlabeled cells "most likely" to be interesting. Specifically, we choose d distinct cells that achieve the highest values of acquisition function $J : \mathcal{G} \times \mathbb{N} \to \mathbb{R} \cup \{\infty\}$ at time t,
$$J_{\boldsymbol{\pi}}(i,t) = \hat{\mu}_{i,\boldsymbol{\pi}}(t) + U_{i,\boldsymbol{\pi}}(t,\delta), \quad (6a)$$
$$\hat{\mu}_{i,\boldsymbol{\pi}}(t) = \frac{\sum_{h \in \mathcal{H}_i(t)} h}{|\mathcal{H}_i(t)|}, \quad (6b)$$
$$U_{i,\boldsymbol{\pi}}(t,\delta) = 2\sqrt{\frac{2\log(\log_2(2|\mathcal{H}_i(t)|)) + \log(12|\mathcal{G}|/\delta)}{2|\mathcal{H}_i(t)|}}, \quad (6c)$$
with $\hat{\mu}_{i,\boldsymbol{\pi}}(t) = U_{i,\boldsymbol{\pi}}(t,\delta) = \infty$, whenever $\mathcal{H}_i(t) = \emptyset$.

Otherwise, we minimize the movement cost ℓ in (5) to decide on the next location to sample. Since ℓ is a metric, a search agent continues to sample its current cell in the next iteration, if the current cell is unlabeled.

Algorithm 1 Multi-agent search under noisy observation

Input: Set of grid cells $\mathcal{G}$, number of agents $d \in \mathbb{N}$, threshold $\theta \in (0,1)$, tolerance $\epsilon > 0$, labeling error probability $\delta \in (0,1)$, aggressiveness param. $\alpha \in (0,1)$
Output: $\{\mathcal{K}(t)\}_{t\geq 1}$, a sequence of (keep) sets of grid cells

Initialize time counter $t \leftarrow 1$
while $\mathcal{R}(t) \cup \mathcal{K}(t) \neq \mathcal{G}$ **do**
  if current toss of α-biased coin returns heads **then**
    Define $\mathbf{a}_t$ by selecting d distinct grid cells that score the highest values in $J_{\boldsymbol{\pi}}$ (6)
  else
    Define $\mathbf{a}_t$ by assigning each agent to a distinct unlabeled cell that minimizes ℓ in (5)
  end if
  Deploy the agents to grid cells $\mathbf{a}_t$ and update the history $\mathcal{H}(t)$ based on collected noisy sensors
  Update sets of labeled grid cells,
  $$\mathcal{K}(t) \leftarrow \{i \in \mathcal{G} | \hat{\mu}_{i,\boldsymbol{\pi}}(t) - U_{i,\boldsymbol{\pi}}(t,\delta) \geq \theta - \epsilon\}, \quad (7a)$$
  $$\mathcal{R}(t) \leftarrow \{i \in \mathcal{G} | \hat{\mu}_{i,\boldsymbol{\pi}}(t) + U_{i,\boldsymbol{\pi}}(t,\delta) \leq \theta + \epsilon\}. \quad (7b)$$
  Increment time counter $t \leftarrow t+1$
end while
return $\{\mathcal{K}(t)\}_{t\geq 1}$

Finally, we complete the loop by updating the sets $\mathcal{K}(t+1)$ and $\mathcal{R}(t+1)$ using (7) based on the data collected in the iteration t. Since $U_{i,\boldsymbol{\pi}}(t,\delta)$ is a non-increasing function of $|\mathcal{H}_i(t)|$, the sets $\mathcal{K}(t)$ and $\mathcal{R}(t)$ are monotonic in t. The definitions used in (7) are motivated by the desire to obtain *anytime guarantees* for Algorithm 1.

Proposition 1. (Anytime algorithm). The following holds for Algorithm 1 at any time $t \geq 1$ with probability of at least $1-\delta$: $\mathcal{K}(t) \subseteq \mathcal{S}_{\theta-\epsilon}$ and $\mathcal{R}(t) \subseteq \mathcal{S}^{\complement}_{\theta+\epsilon}$.

We provide the proof of Proposition 1 in Appendix A. By Proposition 1, Algorithm 1 yields a correct-by-construction (albeit incomplete) labeling of the grid cells, even when it is terminated prematurely.

We conclude this section by noting that Algorithm 1 simplifies to a label-then-move search when $\alpha = 0$, and a thresholding bandit-based search when $\alpha = 1$.

4. PERFORMANCE ANALYSIS

We now focus on Problem 2, and study the performance of Algorithm 1. We show that Algorithm 1 has finite time termination guarantees, and admits high likelihood upper bounds to the incurred economic costs and priority labeling time. These bounds are a natural consequence of the bandit framework which yield non-asymptotic performance guarantees under minimal modeling assumptions.

We will use the following problem-specific parameters for each cell $i \in \mathcal{G}$,
$$\Delta_i = |\mu_i - \theta| + \epsilon, \text{ and } \Omega_i = \min_{j \in \mathcal{S}_{\theta+\epsilon}} |\mu_j - \mu_i|. \quad (8)$$
Informally, Δ_i signifies the separation of the mean μ_i from the threshold, while Ω_i signifies the separation of the mean μ_i from the set $\mathcal{S}_{\theta+\epsilon}$. We will state our results using parameters ϕ_i and γ_i for each cell $i \in \mathcal{G}$,

$$\phi_i = \frac{1}{\Delta_i^2} \log\left(\frac{|\mathcal{G}|}{\delta} \log\left(\frac{|\mathcal{G}|}{\Delta_i^4 \delta}\right)\right), \tag{9a}$$

$$\gamma_i = \frac{1}{\Omega_i^2} \log\left(\frac{|\mathcal{G}|}{\delta} \log\left(\frac{|\mathcal{G}|}{\Omega_i^4 \delta}\right)\right). \tag{9b}$$

Similar to the bandit literature (Lattimore and Szepesvári, 2020), we will show that Δ_i and Ω_i together characterize the difficulty of the search problem in Theorem 1.

Remark 2. For any two scalar functions $f, g : \mathbb{R} \to \mathbb{R}$, we write $f = O(g)$ if there exists a constant $C > 0$ and a scalar $x_0 \in \mathbb{R}$ such that $f(x) \leq Cg(x)$ for every $x \geq x_0$.

Theorem 1. (Upper bounds for Algorithm 1). Each one of the following statements hold for Algorithm 1 with probability $1 - \delta$:

(1) Algorithm 1 terminates at $T_{\boldsymbol{\pi}}$ and satisfies the low labeling error criterion (3) with

$$T_{\boldsymbol{\pi}} \leq \max_{i \in \mathcal{D}_\Delta} O(\phi_i) + \frac{1}{d} \sum\nolimits_{i \in \mathcal{D}_\Delta^{\complement}} O(\phi_i), \tag{10}$$

where $\mathcal{D}_\Delta$ is the union of a grid cell with the smallest Δ_i with a set of $d-1$ grid cells with the largest Δ_i among all cells $i \in \mathcal{G}$.

(2) The priority labeling time (4) for Algorithm 1 is bounded from above as follows,

$$L(\boldsymbol{\pi}) \leq \max_{i \in \mathcal{D}_\Omega} O(\phi_i) + \frac{1}{d} \sum\nolimits_{i \in \mathcal{S}_{\theta+\epsilon} \setminus \mathcal{D}_\Omega} O(\phi_i) + \frac{1}{d} \sum_{i \in \mathcal{S}_{\theta+\epsilon}^{\complement} \setminus \mathcal{D}_\Omega} \min\left\{O(\gamma_i) + \frac{4(1-\alpha)|\mathcal{G}|^2}{\alpha\delta}, O(\phi_i)\right\}, \tag{11}$$

where $\mathcal{D}_\Omega$ is a set of d grid cells characterized by Δ_i, Ω_i, and α.

(3) The economic cost (5) incurred by Algorithm 1 is bounded from above (with $M = \max_{\mathbf{a}, \mathbf{a}' \in \mathcal{G}^d} \ell(\mathbf{a}, \mathbf{a}')$),

$$E(\boldsymbol{\pi}) \leq O(|\mathcal{G}| - 1) + d \max_{i \in \mathcal{D}_\Delta} O((M\alpha + \beta)\phi_i) + \sum\nolimits_{i \in \mathcal{D}_\Delta^{\complement}} O((M\alpha + \beta)\phi_i). \tag{12}$$

See Appendix B for a sketch of the proof of Theorem 1. In Theorem 1, *big-O* notation hides constants factors which are independent of system parameters.

From (10), Algorithm 1 may take more time to terminate when Δ_i is small for at least one grid cell, i.e., μ_i is close to θ for some $i \in \mathcal{G}$. Additionally, the upper bound on the termination time $T_{\boldsymbol{\pi}}$ does not have a purely inverse-linear relationship the number of agents d, i.e., $T_{\boldsymbol{\pi}}$ is not upper bounded by an expression containing only $\frac{1}{d}\sum_{i \in \mathcal{G}} O(\phi_i)$. Instead, the upper bound in (10) has an additional term $\max_{i \in \mathcal{D}_\Delta} O(\phi_i)$ independent of d, which corresponds to the diminishing benefit of significantly increasing the number of agents.

We now analyze the role played by the aggressiveness parameter α in the performance of Algorithm 1. We observe that the upper bound on $T_{\boldsymbol{\pi}}$ is independent of α, consistent with the intuition that Algorithm 1 with $\alpha > 0$ and label-then-move search (Algorithm 1 with $\alpha = 0$) takes the same number of iterations for a search problem with identical Δ_i. This is because all grid cells must be labeled at the end, and both approaches rely on similar concentration inequalities to label a cell.

Recall that Algorithm 1 simplifies to label-then-move search for $\alpha = 0$, where the deployments of the agents are decided solely based on the associated movement costs. Consequently, as seen from the upper bounds, such an approach incurs a low economic cost $E(\boldsymbol{\pi})$, but may incur a high priority labeling time $L(\boldsymbol{\pi})$. On the other hand, setting $\alpha = 1$ simplifies Algorithm 1 to a pure bandit-based search that samples grid cells based on the maxima of the acquisition function $J_{\boldsymbol{\pi}}$ (6). Consequently, as also seen from the upper bounds, such an approach will result in low priority labeling time $L(\boldsymbol{\pi})$, but high economic cost $E(\boldsymbol{\pi})$. Thus, by varying $\alpha \in (0, 1)$, the method can achieve the desired trade-off between the priority labeling time and the economic costs of search.

5. EXPERIMENTS

We use a numerical simulation to compare Algorithm 1 to three baselines, `AdaSearch` (Rolf et al., 2021), a pure bandit-based search (Algorithm 1 with $\alpha = 1$), and label-then-move search (Algorithm 1 with $\alpha = 0$).

We setup the multi-agent search problem as follows: Consider a search environment of 10×10 grid cells with mean $\mu_i = 0.85$ for interesting cells $i \in \mathcal{G}$ and $\mu_j = 0.15$ for uninteresting cells $j \in \mathcal{G}$. We set the team size $d = 5$ with randomly chosen starting locations. We also set 10 randomly chosen grid cells as interesting. We set the tolerance $\epsilon = 10^{-3}$, labeling error probability $\delta = 10^{-3}$, and the threshold parameter $\theta = 0.5$. We set sensing costs $\beta = 0.01$, and define ℓ as the sum of the Manhattan distance between the agents' current and next locations.

We adapt `AdaSearch` (Rolf et al., 2021) to solve the multi-agent search problem. Recall that `AdaSearch` adjusts the number of samples collected at each cell based on any valid data-driven confidence bounds. In our implementation of `AdaSearch`, we utilized the confidence bounds defined in (6). We also assumed that the agents follow identical raster paths and recompute the sample visitation counts upon completing a loop around the environment. Unlike Algorithm 1, `AdaSearch` additionally requires the total number of interesting cells to label the cells.

We analyze how the different search strategies label interesting cells to the keep set $\mathcal{K}(t)$ as time progresses in the algorithm. Figure 2 shows the performance of the algorithms on 100 randomly generated search problems. Based on our experiments, we recommend the choice of $\alpha = 0.2$ for the given choice of ℓ and β.

Priority labeling time (Fig. 2, top): As expected, the proposed solution (Algorithm 1 with $\alpha = 0.2$) and a pure bandit-based search (Algorithm 1 with $\alpha = 1$) detects $\mathcal{S}_{\theta+\epsilon}$ with a smaller number of samples as compared to `AdaSearch` and label-then-move search (Algorithm 1 with $\alpha = 0$). The latter search strategies require a large amount of samples, possibly due to the pre-determined search pattern used by the agents.

Economic cost (Fig. 2, bottom): The proposed solution and label-then-move search incur lower economic costs when compared to AdaSearch and pure bandit-based search. From (5), the economic cost is a linear combination of sampling cost and movement cost. Since β is small, the

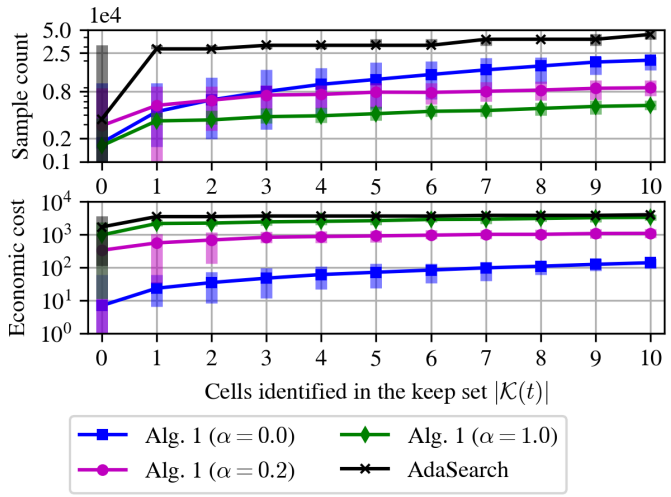

Fig. 2. Priority labeling time for various search strategies with number of samples collected by the team over $\mathcal{G}$ (top), and incurred economic cost (bottom). The proposed solution (Algorithm 1 with $\alpha = 0.2$) achieves a good compromise as compared to other strategies — label-then-move search (Algorithm 1 with $\alpha = 0$), a pure bandit-based search (Algorithm 1 with $\alpha = 1$), and `AdaSearch` (Rolf et al., 2021).

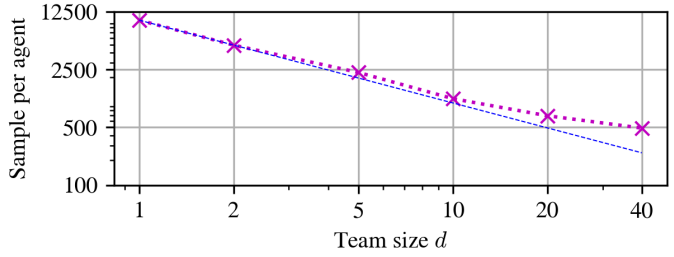

Fig. 3. The median number of samples needed per agent to characterize the keep set $\mathcal{K}(T_{\boldsymbol{\pi}})$ (magenta line with crosses) in 100 randomly chosen search problems using the proposed solution (Algorithm 1 with $\alpha = 0.2$) decreases with increasing team size d. The blue line shows the trend needed to achieve an inversely proportional relationship between the samples needed per agent and the team size.

incurred costs are primarily driven by the movement costs, and the proposed solution and label-then-move search make the agents move relatively less when compared to other approaches. We expect the performance of the search strategies to be similar to Fig. 2, top, in applications where sensing is expensive and movement is cheap (higher β).

Impact of the team size d on the search (Fig. 3): As expected, the number of samples needed per agent to characterize the keep set $\mathcal{K}(T_{\boldsymbol{\pi}})$ decreases with increasing team size d. However, the reduction in the samples needed per agent does not exhibit an inversely proportional relationship with the team size, as indicated by Theorem 1.

We conclude with a note that Algorithm 1 has minimal computational overhead. A non-optimized Python code took ≈ 0.3 milliseconds per iteration on a standard laptop.

6. CONCLUSION & FUTURE WORK

We propose a data-driven, multi-agent search algorithm that accommodates noisy observations when searching for *all* interesting grid cells. We combined recent results from thresholding MABs with a standard label-then-move search to lower the time to identify interesting areas in the search environment and lower the costs incurred by the search agents during movement and sensing, while accommodating noisy observations.

The multi-agent search strategy proposed in this work has two drawbacks. First, it does not enforce the physical limitations on the mobile sensors are enforced as hard constraints during exploration. Second, it does not consider the effect of temporal changes in the search environment. Our future work will extend the proposed solution to address these drawbacks.

REFERENCES

Bajcsy, R., Aloimonos, Y., and Tsotsos, J.K. (2018). Revisiting active perception. *A. Robots*, 42(2), 177–196.

Banerjee, A., Ghods, R., and Schneider, J. (2022). Multi-agent active search using detection and location uncertainty. *arXiv preprint arXiv:2203.04524*.

Best, G., Faigl, J., and Fitch, R. (2018). Online planning for multi-robot active perception with self-organising maps. *A. Robots*, 42(4), 715–738.

Bullo, F., Cortés, J., and Martinez, S. (2009). *Distributed control of robotic networks*. Princeton Univ. Press.

Drew, D.S. (2021). Multi-agent systems for search and rescue applications. *Current Robotics Rep.*, 2, 189–200.

Du, B., Qian, K., Iqbal, H., Claudel, C., and Sun, D. (2021). Multi-robot dynamical source seeking in unknown environments. In *Intl Conf. Robotics and Autom.*, 9036–9042.

Ghods, R., Banerjee, A., and Schneider, J. (2021). Decentralized multi-agent active search for sparse signals. In *Proc. Conf. Uncertainty Artif. Intelli.*, volume 161, 696–706.

Jun, K.S., Jamieson, K., Nowak, R., and Zhu, X. (2016a). Top arm identification in multi-armed bandits with batch arm pulls. In *Artificial Intelligence and Statistics*, 139–148. PMLR.

Jun, K.S., Jamieson, K., Nowak, R., and Zhu, X. (2016b). Top arm identification in multi-armed bandits with batch arm pulls. In *Proc. Intl Conf. Artifi. Intelli. Stats.*, volume 51, 139–148.

Kapoutsis, A.C., Chatzichristofis, S.A., and Kosmatopoulos, E.B. (2017). DARP: Divide areas algorithm for optimal multi-robot coverage path planning. *J. Intelli. Robotic Syst.*, 86(3), 663–680.

Krause, A. and Guestrin, C. (2007). Near-optimal observation selection using submodular functions. In *AAAI*, volume 7, 1650–1654.

Lattimore, T. and Szepesvári, C. (2020). *Bandit algorithms*. Cambridge Univ. Press.

Locatelli, A., Gutzeit, M., and Carpentier, A. (2016). An optimal algorithm for the thresholding bandit problem. In *Intl. Conf. Machine Learning*, 1690–1698.

Luo, W., Nam, C., Kantor, G., and Sycara, K. (2019). Distributed environmental modeling and adaptive sampling for multi-robot sensor coverage. In *Proc. Intl. Conf. Auto. Agents Multi-Agent Syst.*, 1488–1496.

Marchant, R. and Ramos, F. (2012). Bayesian optimisation for intelligent environmental monitoring. In *IEEE Int'l Conf. Intelli. Robots Syst.*, 2242–2249.

Mason, B., Jain, L., Tripathy, A., and Nowak, R. (2020). Finding all ϵ-good arms in stochastic bandits. *Adv. Neural Info. Process. Syst.*, 33, 20707–20718.

Queralta, J., Taipalmaa, J., Pullinen, B., Sarker, V., Gia, T., Tenhunen, H., Gabbouj, M., Raitoharju, J., and Westerlund, T. (2020). Collaborative multi-robot search and rescue: Planning, coordination, perception, and active vision. *IEEE Access*, 8, 191617–191643.

Rolf, E., Fridovich-Keil, D., Simchowitz, M., Recht, B., and Tomlin, C. (2021). A successive-elimination approach to adaptive robotic source seeking. *IEEE Tran. Robotics*, 37(1), 34–47.

Schwager, M., Vitus, M., Powers, S., Rus, D., and Tomlin, C.J. (2015). Robust adaptive coverage control for robotic sensor networks. *IEEE Tran. Ctrl. Netw. Syst.*, 4(3), 462–476.

Appendix A. PROOF FOR PROPOSITION 1

Let the undesirable events be $\mathcal{E}_{\mathcal{K}}(t) = \{\mathcal{K}(t) \setminus \mathcal{S}_{\theta-\epsilon} \neq \emptyset\}$ and $\mathcal{E}_{\mathcal{R}}(t) = \left\{\mathcal{R}(t) \setminus \mathcal{S}^{\complement}_{\theta+\epsilon} \neq \emptyset\right\}$. To prove Proposition 1, we want to show $\mathbb{P}\left[\left(\bigcup_{t\geq 1}\mathcal{E}_{\mathcal{K}}(t)\right) \bigcup \left(\bigcup_{t\geq 1}\mathcal{E}_{\mathcal{R}}(t)\right)\right] \leq \delta$. By Boole's inequality, it is suffices to show that $\mathbb{P}\left[\bigcup_{t\geq 1}\mathcal{E}_{\mathcal{K}}(t)\right] \leq \delta/2$ and $\mathbb{P}\left[\bigcup_{t\geq 1}\mathcal{E}_{\mathcal{R}}(t)\right] \leq \delta/2$.

Recall that, by the choice of $U_{i,\boldsymbol{\pi}}(t,\delta)$ in (6c),

$$\mathbb{P}\left[\bigcap_{t\geq 1}\{|\mu_i - \hat{\mu}_i| \leq U_{i,\boldsymbol{\pi}}(t,\delta)\}\right] \geq 1 - \frac{\delta}{2|\mathcal{G}|}, \qquad \text{(A.1)}$$

for any grid cell $i \in \mathcal{G}$ and search policy $\boldsymbol{\pi}$ (see Lemma 1 in (Jun et al., 2016b) with $\omega = \sqrt{\delta/(12|\mathcal{G}|)}$).

For any $t \geq 1$, $\mathcal{K}(t) \setminus \mathcal{S}_{\theta-\epsilon} \neq \emptyset$ if and only if $\hat{\mu}_{i,\boldsymbol{\pi}}(t) - U_{i,\boldsymbol{\pi}}(t,\delta) \geq \theta - \epsilon > \mu_i$ for some $i \in \mathcal{G}$. Consequently, $\mathcal{K}(t) \setminus \mathcal{S}_{\theta-\epsilon} \neq \emptyset \Rightarrow |\hat{\mu}_{i,\boldsymbol{\pi}}(t) - \mu_i| \geq U_{i,\boldsymbol{\pi}}(t,\delta)$ for some $i \in \mathcal{G}$. By (A.1) and Boole's inequality,

$$\begin{aligned}\mathbb{P}\left[\bigcup_{t\geq 1}\mathcal{E}_{\mathcal{K}}(t)\right] &\leq \mathbb{P}\left[\bigcup_{i\in\mathcal{G}}\bigcup_{t\geq 1}\{|\hat{\mu}_{i,\boldsymbol{\pi}}(t) - \mu_i| \geq U_{i,\boldsymbol{\pi}}(t,\delta)\}\right] \\ &\leq \sum_{i\in\mathcal{G}}\frac{\delta}{2|\mathcal{G}|} \leq \frac{\delta}{2}.\end{aligned}$$

The proof for $\mathbb{P}\left[\bigcup_{t\geq 1}\mathcal{E}_{\mathcal{R}}(t)\right] \leq \delta/2$ follows similarly. ■

Appendix B. PROOF SKETCH FOR THEOREM 1

Bounding $T_{\boldsymbol{\pi}}$ (10): Recall that the number of sufficient number of samples required for successful classification, with high confidence, of a grid cell $i \in \mathcal{G}$ can be tightly upper-bounded by $\mathcal{O}(\phi_i)$ (see (6) in (Jun et al., 2016a) with $\omega = \sqrt{\delta/(2|\mathcal{G}|)}$). If we allowed multiple search agents to visit a grid cell simultaneously, then $\frac{1}{d}\sum_{i\in\mathcal{G}}\mathcal{O}(\phi_i)$ upper bounds $T_{\boldsymbol{\pi}}$, due to the independence assumption between the cells. However, the agents are required to stay in distinct cells at all times. Consequently, some of the search agents are rendered ineffective when less than d cells are left to be labelled. (10) upper bounds $T_{\boldsymbol{\pi}}$ by accounting for the worst-case inefficiency — the last d cells is a set of "easy-to-classify" $d-1$ cells and a "hardest-to-classify" cell.

Bounding $L(\boldsymbol{\pi})$ (11): We split the time taken to classify all interesting cells by Algorithm 1 into three parts:

(i) *Classifying interesting cells* : The number of samples sufficient for classification of interesting cell i is $\mathcal{O}(\phi_i)$,

(ii) *Sampling uninteresting cells when biased coin toss yields heads* : Here, Algorithm 1 samples grid cells while maximizing J, see (6). The number of sufficient samples of the uninteresting cell j after which, it will be sampled by *only* after classifying *all* interesting cells upper-bounded by $\mathcal{O}(\gamma_j)$. Additionally, we add a margin of $\frac{4(1-\alpha)|\mathcal{G}|^2}{\alpha\delta}$ to account for the worst-case low-probability event of revisiting cell j due to the switching to label-then-move.

(iii) *Sampling uninteresting cells when biased coin toss yields tails* : Here, Algorithm 1 samples grid cells based only on the distance metric ℓ. In the worst case, we may sample an uninteresting cell long enough to classify it, which is $O(\phi_i)$.

Combining these parts, we have

$$L(\boldsymbol{\pi}) \leq \underbrace{\sum \mathcal{O}(\phi_i)}_{\text{interesting cells (i)}} + \underbrace{\sum \min\left\{\mathcal{O}(\phi_i), \mathcal{O}(\gamma_i) + \frac{4(1-\alpha)|\mathcal{G}|^2}{\alpha\delta}\right\}}_{\text{uninteresting cells (ii) and (iii)}}.$$

for a team with $d = 1$. Similar to $T_{\boldsymbol{\pi}}$, we obtain (11) by accounting for oversampling due to inefficiency arising from the presence of $d > 1$ agents.

Bounding $E(\boldsymbol{\pi})$ (12): The economic cost (5) consists of the movement cost and the sampling cost. At every time step t, Algorithm 1 performs an α biased coin-toss. For coin tosses corresponding to heads, Algorithm 1 moves to the grid cells which maximizes the activation function (6). In this case, Algorithm 1 incurs a movement cost of at most $M \triangleq \max_{\mathbf{a},\mathbf{a}'} \ell(\mathbf{a},\mathbf{a}')$. For coin tosses corresponding to tails, Algorithm 1 searches for the nearest unlabelled cell in the neighbourhood. The total cost incurred by Algorithm 1 during the *entire* run is no larger than the cost incurred to visit all of the cells in some pre-defined sequence, which we know is $O(|\mathcal{G}| - 1)$.

$$\sum_{1\leq\tau\leq t}\mathbb{E}\left[\ell(\mathbf{a}_\tau, \mathbf{a}_{\tau-1})\right] \leq (1-\alpha)O(|\mathcal{G}|-1) + \alpha M t \qquad \text{(B.1)}$$

The sampling cost accrued at iteration t of Algorithm 1 is $\beta t d$. We complete the proof by adding these bounds, and applying the bound in (10) on t. ■