Active Multi-Target Search Using Distributed Thompson Sampling

Jun Chen and Philip Dames

Abstract— **Distributed search and track is a canonical task for multi-robot system, encompassing applications from environmental monitoring to disaster response to surveillance. In many situations, the distribution of objects in a search area is irregular, with some areas having high object densities while other areas have low densities. In this paper, we propose a novel distributed control strategy that combines Bernoulli Thompson sampling and Lloyd's algorithm to allow robots to actively search for and track targets, respectively. Compared to previous approaches that use only Lloyd's algorithm, we show that the proposed algorithm significantly accelerates the speed at which robots find targets, particularly when large portions of targets congregate in some fixed areas in the environment.**

I. Introduction

Multi-robot coordination has drawn increasing attention over the past decades as robots become more powerful and low-cost. Compared to single robots, multi-robot systems (MRSs) have improved detection range, mobility, versatility, and robustness to failure. Distributed search and track is one of the most common applications of MRSs. There are two key components to any such system: an estimation system to model and track objects as they are detected and a control system to drive the motion of individual robots in the team towards areas that contain useful information.

A. Multi-Target State Estimation

We focus on the set of problems where robots must detect and track a large number of discrete objects (e.g., people, vehicles, landmarks), which is often modeled as a multi-target tracking (MTT) problem. Different from single target tracking, the main challenge of MTT is matching detections to target tracks, especially in the precense of false negative and false positive detections, a process known as data association. A number of standard MTT algorithms, each of which solve data association in a different way, have been developed, including global nearest neighbor (GNN) [1], joint probabilistic data association (JPDA) [2], multiple hypothesis tracking (MHT) [3], and particle filters [4]. Each of these trackers propagates the posterior of target states over time and solves the data association problem prior to tracking. Another class of MTT techniques is derived from random finite set (RFS) statistics [5] and do not require explicitly solving data association. We use the probability hypothesis density (PHD) filter [6], which tracks the spatial density of targets, making it best suited to situations where each target is not required to have a unique identity, *e.g.*, a rescue robot only needs to know where all of the people are located. We recently developed a distributed PHD filter that is provably equivalent to the centralized solution [7].

B. Distributed Control of MRSs

Lloyd's algorithm [8] is one of the best-known control algorithms for distributed target tracking, the idea of which is to represent target states by a weighting function over the task space and to drive each robot to the weighted centroid of its Voronoi cell [9]. In our prior work, we use the PHD as the weighting function, driving robots to actively track targets [7]. However, when no target is within a robot's Voronoi cell, the robots move erratically, reacting to any false positive detections as well as the dynamically changing shape of their Voronoi cells. As a result, robots often stay within empty sub-regions instead of purposefully seeking out untracked targets, slowing down the rate at which they find targets. This problem is further exacerbated when a majority of targets gather within some small subsets of the environment, as is often the case in real life, *e.g.,* animals cluster around water sources within large nature reserves.

One way to improve this is to have idle robots (*i.e.,* those not tracking targets) sample the task space in a way that balances between searching low-density areas for undetected targets (exploration) and high-density areas to increase the probability of finding a target (exploitation). This coincides with the multi-armed bandit (MAB) problem [10], in which a gambler must decide which arm of K nonidentical slot machines to play to maximize the reward. MABs have been applied to multi-robot task allocation [11] and sensing [12]–[14] problems, though their use is not widespread. One important MAB solution is Thompson sampling (TS) [15], which has recently proven successful in solving MAB problem in a stochastic manner [16]. In fact, Chapelle et al. [17] show that TS is among the most effective and easy-implemented MAB solvers algorithms. TS also allows for delayed feedback after sampling, which best fits distributed MRS scenarios since robots do not receive rewards until they reach their goal. In this paper, we choose to use a dynamic variation of TS [18] for active target search, which handles the temporal variations of the target distribution.

C. Contributions

In this work, we demonstrate in a series of simulated experiments that a team of robots using the combined TS and Lloyd's algorithms more effectively finds and tracks targets than a team that uses only Lloyd's algorithm. The methods we validate include: 1) a distributed active search algorithm based on dynamic TS, 2) combining the TS-based

J. Chen and P. Dames are with the Department of Mechanical Engineering, Temple University, Philadelphia, PA 19122, USA {jchen,pdames}@temple.edu

This work was supported by NSF Grant IIS-1830419.

search with Lloyd's algorithm for active tracking, 3) a goal swapping algorithm to more effectively assign goals to each robot..

II. Results

A. Simulation Environment

We test our proposed algorithms via MATLAB simulations. The task space is a $100\,\mathrm{m} \times 100\,\mathrm{m}$ square. Targets may either be static or moving within their sub-regions at a maximum speed of $3\,\mathrm{m/s}$. Existing targets may disappear (*i.e.,* leave the environment) and new targets may appear (*i.e.,* enter the environment) so that the number of targets may change over time.

All robots begin each trial at randomized locations within a $20\,\mathrm{m} \times 10\,\mathrm{m}$ box at the bottom center of the environment. Robots have a maximum speed of $10\,\mathrm{m/s}$ and are equipped with isotropic sensors with a sensing radius $\rho_f = 5\,\mathrm{m}$.

We use probability hypothesis density (PHD) filter for target state estimation and assume the robots do not have any prior knowledge of the targets. Thus, the robots use a Gaussian random walk (with $\sigma = 0.35\,\mathrm{m/s}$) for the motion models f, set the survival probability to 1, and the birth PHD to 0. We use the same measurement model for sensors as [7], with the exception of assuming that sensors are all homogeneous and produce no missed or false detections.

B. Qualitative Comparison

We first show how active search using TS qualitatively improves multi-target tracking using a single trial. There are 40 robots searching for 40 targets, where 30 targets are located in a $33 \times 33\,\mathrm{m}$ square sub-region at the lower-left corner of E, and another 10 targets in a $33 \times 33\,\mathrm{m}$ squared sub-region at the top-right corner. Targets locations are drawn uniformly at random within each sub-region. For simplicity, the targets are stationary and the number of targets is constant over time (note: the robots still use a Gaussian motion model within the PHD filter).

Figure 1 shows the locations of robots and targets at various points during exploration using both our previous [7] and new methods. When using our previous method, which only used Lloyd's algorithm, a large portion of robots are idle even when a fair amount of targets are not tracked after 40 s, as the centroids of these orange diamonds are not located in any of the green circles in Fig. 1a. After 60 s, as shown in Fig. 1b, robots tend to move towards the two corners with targets but a large portion of them have still not found any targets. We also see that robots get stuck on the boundary of target clusters, never reaching the interior targets. The result demonstrates the weakness of pure Lloyd's algorithm that idle robots do not actively search for targets, causing an inefficient use of the total sensing capability of the team while searching for untracked targets.

On the other hand, the team using distributed Thompson sampling is able to quickly learn the target distribution and cluster in regions likely to contain targets. As Fig. 1c shows, after 40 s a large number of the robots have already gathered at the large cluster of targets while a handful of other robots continue to search unexplored areas. After 60 s, most of the robots have found a target while a few continue to maintain coverage of these unlikely regions, as Fig. 1d shows. The resulting emergent robot distribution is consistent with the target distribution, as the lower-left corner contains a higher number of robots than the upper-right corner (28 vs. 9, which closely match the number of targets in each region), while 3 robots monitor the rest of the area. This helps the team to track the targets more quickly as the more individuals are needed in an area, the more likely it is that an individual will be assigned to that region. Figure 1e reflects the value of $\boldsymbol{\alpha}$ for each sampling candidates after 60 s, showing that the team has received more reward in areas with higher target concentrations.

C. Quantitative Comparison

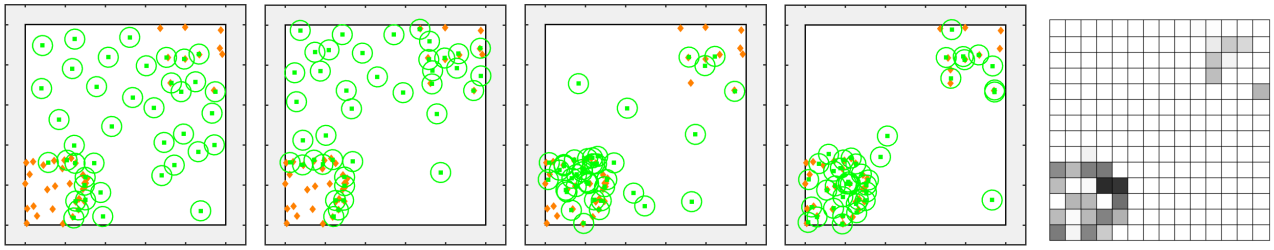
To quantify the improvement in performance, we will use the first order Optimal SubPattern Assignment (OSPA) metric [19], a commonly-used approach in MTT. We use $c = 10\,\mathrm{m}$, $p = 1$, and measure the error between the true and estimated target sets. Note that a lower OSPA value indicates a more accurate tracking of the target set. We report the median OSPA value over the final 150 s of each trial, allowing the team to reach a steady state and smoothing out the effects of spurious measurements that cause the OSPA to fluctuate. We also show the 95% rise time of the OSPA error metric, *i.e.,* the time it takes for the OSPA error to reach a value within 5% of the final value, to measure the speed at which robots reach steady state.

1) Tests Design: We test a range of team sizes (from 10 to 40 robots) and both search strategies (from [7] and the new method). For each configuration (environment, target type, team size) we run 10 trials, with the results aggregated into boxplots showing the steady state OSPA (to measure accuracy) and the 95% rise time (to measure speed).

2) Results: Figure 2 shows the results. As we see in Fig. 2a and Fig. 2b, the median OSPA decreases as the number of robots increases. This agrees with intuition as more robots should be able to better locate targets and there should be diminishing rewards with each added robot, *i.e.,* going from 10 to 15 robots is more significant than 35 to 40 robots. We see that for static targets (Fig. 2a and Fig. 2c), our proposed method shows significantly lower OSPA error and rise time and that the variation of both parameters across trials is smaller, meaning it more accurate, faster, and more reliable. For dynamic targets (Fig. 2b and Fig. 2d), the OSPA error of teams using our proposed method are comparable or slightly higher and exhibit slightly more variation across trials, though neither effect is significant. Like the static case, our new method decreases both the magnitude and variation of the rise time, meaning it faster and more repeatable. We hypothesize that these differences in behavior are due to the ability of robots to actively sample the environment using coarse global information.

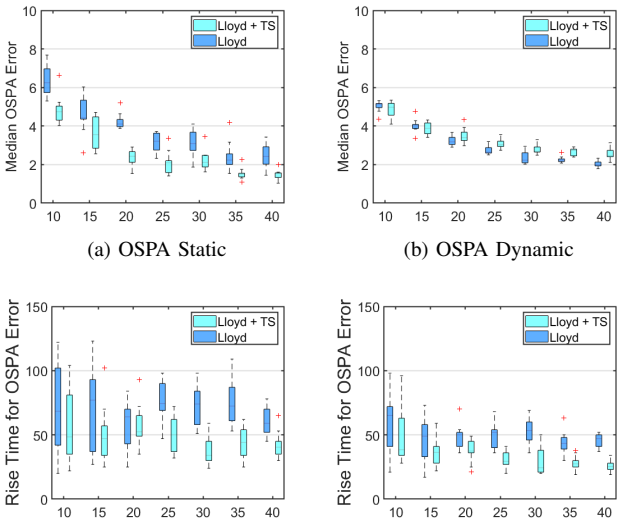
III. Conclusions

We develop a distributed control algorithm combining a novel distributed Thompson sampling algorithm with Lloyd's



(a) Lloyds 40s (b) Lloyds 60s (c) Lloyds+TS 40s (d) Lloyds+TS 60s (e) α Values

Fig. 1. Figures show comparison of applying pure Lloyd's algorithm and a combined Lloyd's algorithm with Thompson sampling after 100 s. In Fig 1a,1b,1c and 1d, green squares and circles show robot locations and sensor footprints, respectively. Orange diamonds show the locations of targets. Fig 1e maps α values in Fig 1d by darkness, with a darker color indicates a higher value.



(a) OSPA Static (b) OSPA Dynamic

(c) Rise Time Static (d) Rise Time Dynamic

Fig. 2. Boxplots show median OSPA errors and the 95% rise time of different teams of robots tracking static and dynamic targets using pure Lloyd's algorithm and combination of Lloyd's and TS.

algorithm to allow robots to effectively search for and track multiple moving targets without having any prior knowledge about targets. In particular, we see that the addition of TS allows robots to share coarse global information about recently detected targets in an efficient and scalable manner. As a result, teams using TS are able to more accurately track (stationary) targets, locate targets more quickly, and increase the consistency in performance. These trends are more pronounced in situations where targets are unevenly distributed within the search space.

References

[1] P. Konstantinova, A. Udvarev, and T. Semerdjiev, "A study of a target tracking algorithm using global nearest neighbor approach," in *Proceedings of the International Conference on Computer Systems and Technologies (CompSysTech'03)*, 2003, pp. 290–295.

[2] S. Hamid Rezatofighi, A. Milan, Z. Zhang, Q. Shi, A. Dick, and I. Reid, "Joint probabilistic data association revisited," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3047–3055.

[3] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 5–18, 2004.

[4] A. Doucet, B.-N. Vo, C. Andrieu, and M. Davy, "Particle filtering for multi-target tracking and sensor management," in *Proceedings of the Fifth International Conference on Information Fusion*, vol. 1. IEEE, 2002, pp. 474–481.

[5] R. P. Mahler, *Statistical multisource-multitarget information fusion*. Artech House Norwood, MA, 2007, vol. 685.

[6] ——, "Multitarget bayes filtering via first-order multitarget moments," *IEEE Transactions on Aerospace and Electronic systems*, vol. 39, no. 4, pp. 1152–1178, 2003.

[7] P. M. Dames, "Distributed multi-target search and tracking using the PHD filter," *Autonomous Robots*, vol. 44, no. 3, Mar. 2020. [Online]. Available: https://doi.org/10.1007/s10514-019-09840-9

[8] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[9] J. Cortes, S. Martinez, T. Karatas, and F. Bullo, "Coverage control for mobile sensing networks," *IEEE Transactions on robotics and Automation*, vol. 20, no. 2, pp. 243–255, 2004.

[10] 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.

[11] G. Pini, A. Brutschy, G. Francesca, M. Dorigo, and M. Birattari, "Multi-armed bandit formulation of the task partitioning problem in swarm robotics," in *International Conference on Swarm Intelligence*. Springer, 2012, pp. 109–120.

[12] W. Luo and K. Sycara, "Adaptive sampling and online learning in multi-robot sensor coverage with mixture of Gaussian processes," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 6359–6364.

[13] Y. Shi, N. Wang, J. Zheng, Y. Zhang, S. Yi, W. Luo, and K. Sycara, "Adaptive informative sampling with environment partitioning for heterogeneous multi-robot systems," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2020.

[14] G. Best, O. M. Cliff, T. Patten, R. R. Mettu, and R. Fitch, "Dec-MCTS: Decentralized planning for multi-robot active perception," *The International Journal of Robotics Research*, vol. 38, no. 2-3, pp. 316–337, 2019.

[15] W. R. Thompson, "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, vol. 25, no. 3/4, pp. 285–294, 1933.

[16] D. Russo, B. Van Roy, A. Kazerouni, I. Osband, and Z. Wen, "A tutorial on Thompson sampling," *arXiv preprint arXiv:1707.02038*, 2017.

[17] O. Chapelle and L. Li, "An empirical evaluation of Thompson sampling," *Advances in neural information processing systems*, vol. 24, pp. 2249–2257, 2011.

[18] N. Gupta, O.-C. Granmo, and A. Agrawala, "Thompson sampling for dynamic multi-armed bandits," in *International Conference on Machine Learning and Applications and Workshops*, vol. 1. IEEE, 2011, pp. 484–489.

[19] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3447–3457, 2008.