On Performance Impacts of Coordination via Submodular Maximization for Multi-Robot Perception Planning and the Dynamics of Target Coverage and Cinematography

Micah Corah and Sebastian Scherer
The Robotics Institute, Carnegie Mellon University
{micahc, basti}@andrew.cmu.edu

Abstract*—A largely unstated problem in work on multi-robot perception planning—particularly the line of work focusing on receding-horizon planning for multi-robot teams—is that *identifying problems and applications where multi-robot coordination has significant impacts on performance metrics can be difficult*. So, even though the community studying perception planning for robot teams—especially via methods for greedy planning and submodular maximization—has developed approaches that are both simple and general, incentives for applying these approaches in practice are limited. This work argues that emphasis on certain applications could reap more significant impacts on system performance and greater incentives for development of these approaches and their applications. Specifically, we hypothesize that problems such as target coverage and applications including cinematography are more amenable to realizing significant impacts on performance than problems such as target tracking and localization, target search, or robotic exploration. We further investigate impacts on system performance in terms of broad differences in the dynamics of these different systems, and our initial simulation results highlight dramatic differences in system behavior. Based on this line of reasoning, we suggest that groups developing infrastructure for multi-robot collaboration should consider relevance to applications including target coverage, multi-robot cinematography, and motion capture with mobile robots as part of the design process.

I. Introduction and Related Work

Automated sensing and perception planning systems have seen wide application such as for cinematography [1], robotic exploration [2] (especially in subterranean environments [3, 4]), and for agriculture [5]—all of these works demonstrate real robots solving perception tasks in environments that are often harsh and complex. Further, the significance of perception planning in solving these tasks is frequently self-evident, and robots often behave intuitively whether following and filming a person while avoiding occlusions or navigating around an object to view its occluded back side.

All of these perception planning problems have multi-robot counterparts. In fact, the literature on multi-robot perception planning is quite general and modular, and many single-robot perception planners for any given application can be readily augmented with methods for multi-robot coordination [6–10]. This is due, in part, to the simple algorithms and strong guarantees that have come from communities studying submodular optimization [8, 11–14]—in particular, this paper will focus on the body of work applying this theory for receding-horizon perception planning for multi-robot teams [7, 10, 14, 15]. Further, theory and design for multi-robot sensor and perception planners is also an active research area with recent advances in resilient optimization [16, 17] and parallelizable and distributed planning [18–20].

Despite robust theory and an active community of researchers, impacts of multi-robot perception planning on task performance metrics are mixed and often limited [7, 15, 21] In some cases, coordination may demonstrably improve *optimization performance* despite limited improvement (if at all) in *task performance metrics* [7, 21]. Alternatively, Hollinger et al. [15] report somewhat more significant results for a search scenario, improving in time to capture (search for) a moving target by about 30% in one scenario and listing improvement of 20% for another.[1]

This work *investigates the relationship between explicit coordination in perception planning and task performance for multi-robot teams*. Sometimes, coordination can happen implicitly when robots share observations or beliefs. Implicit coordination may occur when robots spread out as they navigate toward unmapped parts of a building [22] or when navigating away from other robots toward distant targets whose positions are more uncertain [7]. However, *we demonstrate that certain problems do not exhibit this implicit coordination behavior. In those cases, multi-robot coordination is vital to task performance and even successful task execution*.

Consider robots seeking simply to observe (or cover) targets or other objects with known positions in what we call *target coverage* (this differs from *target tracking* where the focus is on reducing uncertainty).[2] We show that such problems lack this implicit coordination due, effectively, to a lack of a notion of uncertainty. This target coverage problem is also relevant to visual tracking and cinematography tasks, and recently Bucker et al. [25] applied greedy planning for coordination in this domain. We suggest that these problems could be important to the future of the study and application of multi-robot coordination for perception tasks. Further, we suggest

[1] For the purpose of this work, we are interested in works that compare task performance for systems that incorporate some explicit coordination in planning to those same systems without explicit coordination but still sharing prior states and observations if relevant to the problem. In the case of [15] cases of "implicit" or "explicit" coordination would both constitute explicit coordination in our framework, and only "no coordination" would constitute no coordination or implicit coordination (due to the sharing of information) in this work.

[2] However, some works have studied target tracking problems with robots planning to maximize coverage [23, 24]. From that perspective, we would suggest that sharing observations and planning to maximize information gain could be a viable alternative to coverage-based coordination. Likewise, our discussion of problem dynamics may help identify which tracking problems coverage-based approaches would be most relevant to.

that future infrastructure for multi-robot collaboration should likewise consider these problems and applications.

II. Background on submodular maximization and greedy planning

Objectives for sensing and perception tasks frequently have useful monotonicity properties. For example, in many cases, the utility of increasingly large sets of views may always increase, or marginal gains for an observation may decrease conditional on increasingly large sets of prior observations. More formally, we may consider a function $g : 2^{\mathscr{U}} \to \mathbb{R}$ that maps any set of observations (subsets of $\mathscr{U}$, where $2^{\mathscr{U}}$ is the power set) to a scalar utility. We say that g is *monotonic* if $g(A) \geq g(B)$ for $B \subseteq A \subseteq \mathscr{U}$. Likewise, a function is *submodular* (or marginal gains are monotonically decreasing) if $g(A \cup \{x\}) - g(A) \leq g(B \cup \{x\}) - g(B)$ where $x \in \mathscr{U} \setminus A$.

In the problems we are interested in, a team of robots $\mathcal{R} = \{1, \ldots, n_\mathrm{r}\}$ may seek to maximize a perception objective g by each selecting an action from a local set $\mathscr{B}_i$ for $i \in \mathcal{R}$. This optimization problem can be solved near-optimally (at worst within 1/2 of optimal) via a sequential greedy maximization process [12]. Specifically, the greedy algorithm produces the solution $X^\mathrm{g}_{n_\mathrm{r}}$ by greedily selecting actions for each robot:

$$x^\mathrm{g}_i = \operatorname*{arg\,max}_{x \in \mathscr{B}_i} g(\{x\} \cup X^\mathrm{g}_{i-1}), \quad \forall i \in \mathcal{R} \tag{1}$$

whereas $X^\mathrm{g}_i = \{x^\mathrm{g}_i\} \cup X^\mathrm{g}_{i-1}$. In many cases, the maximization step in (1) consists of solving a single-robot perception planning problem, exactly or approximately [7, 8]. As such, this algorithm can enable designers to augment perception planners for individual robots with capacity of multi-robot coordination by writing a simple "for" loop.

III. Problems: Target tracking and coverage and aerial cinematography

Again, let us consider a team of robots $\mathcal{R} = \{1, \ldots, n_\mathrm{r}\}$ and now a set of targets $\mathcal{T} = \{1, \ldots, n_\mathrm{t}\}$. Robot and target states each evolve according to their respective dynamics $\mathbf{x}^\mathrm{r}_{t+1} = f^\mathrm{r}(\mathbf{x}^\mathrm{r}_t, u)$ and $\mathbf{x}^\mathrm{t}_{t+1} = f^\mathrm{r}(\mathbf{x}^\mathrm{t}_t, \epsilon)$ where u is a control input to each robot and ϵ represents some process noise that drives the targets, and for the purpose of this work, ϵ should be independent of the current and prior robot states. Then, depending on the task, robots will also receive some observations $\mathbf{y}$ of the targets at each time-step, and we will assume centralized computation and shared observations.

At each time t, robots will select actions via receding horizon optimization with a horizon of length l. Referring to Sec. II, robots will seek to maximize some perception objective, the local action set $\mathscr{B}_i$ will represent assignment of finite-horizon control actions to robot $i \in \mathcal{R}$, and robots will collectively select sets of actions $X \subseteq \mathscr{U} = \bigcup_{i \in \mathcal{R}} \mathscr{B}_i$.

A. Target tracking

Regarding target tracking, we refer to our prior work on this topic [7]. We define the target tracking task as consisting of minimizing the average entropy of the targets at each time-step.[3] Robots planning for this task will then maximize mutual information

$$g_\mathrm{tracking}(X) = \sum_{j \in \mathcal{T}} \mathbb{I}(\mathbf{X}^\mathrm{t}_{j,t+1:t+l}; \mathbf{Y}_{j,t+1:t+l}(X) | \mathbf{Y}_{j,0:t}, \mathbf{X}^\mathrm{r}_{0:t}) \tag{2}$$

whereas capital letters represent collections of states or observations. That is, the robots seek to maximize information gain with respect to the targets over the planning horizon or, equivalently, they minimize the uncertainty (entropy) of the target states.

B. Target coverage

For the coverage task, target states are known, and robots simply seek to observe (or cover) the targets. At every time-step, each robot observes a set of targets $F^\mathrm{cover}(\mathbf{x}^\mathrm{r}, \mathbf{X}^\mathrm{t}) \subseteq \mathcal{T}$. The coverage task consists of maximizing the average number of covered targets at each time t, that is $\left|\bigcup_{i \in \mathcal{R}} F^\mathrm{cover}(\mathbf{x}^\mathrm{r}_{i,t}, \mathbf{X}^\mathrm{t}_t)\right|$. The finite-horizon coverage objective is then

$$g_\mathrm{coverage}(X) = \mathbb{E}\left[\sum_{k=1}^{l} \left|\bigcup_{i \in \mathcal{R}} F^\mathrm{cover}(\mathbf{x}^\mathrm{r}_{i,t+k}, \mathbf{X}^\mathrm{t}_{t+k})\right|\right]. \tag{3}$$

So, the robots are rewarded for planning to observe greater numbers of targets.

C. Cinematography as weighted coverage

Now, we may fit this idea of target coverage to a more general class such as problems involving cinematography or visual coverage. Consider a group of robots filming one or more human actors. Most directly, we may think of coverage in (3) as representing whether each person is in view of one of the cameras. Alternatively, we may assume a finer discretization of the actors and the scene being filmed with $\mathcal{T}$ representing the elements of that scene and obtain weighted version of the coverage objective. Thus, we argue that target coverage problems we discuss are also broadly representative of visual coverage and cinematography problems and applications.

IV. Problem dynamics and relevance to performance

For the purpose of this work, we will analyze the dynamics of tracking and coverage problems informally. As noted, sharing observations or access to a shared estimate (or filter) in information gathering problems such as target tracking is often sufficient to produce incentives to distribute robots effectively over the environment *without explicit coordination*. Robots seeking to maximize information gain tend to navigate toward parts of the environments associated with high uncertainty whether an unmapped part of a building or a target whose position is not known exactly, and navigating toward regions of high uncertainty often involves navigating away from other

[3] Readers may refer to Cover and Thomas [26] for an introduction to information theory including discussion of entropy and mutual information.

robots (whose observations would reduce uncertainty) thereby distributing the robots across the environment. However, characterizing this intuition formally in a way generally applicable to relevant sensor and perception planning problems may prove difficult.

Instead, we investigate a problem that does not exhibit such implicit coordination: *target coverage*. Rewards for actions in target coverage are instantaneously a function of just the robot states and target states and *not a function of the history of states and observations (as in mapping or tracking)*. As such, prior coverage due to a high density of robots near a location would not reduce rewards for remaining near that same location.

a) Equilibrium behavior for target coverage: Consider a target coverage problem and a team of robots that select actions according to some relevant coverage planner without coordination (myopically). Now, consider the trajectory of an individual robot following any given state-time pair, as it plans to maximize reward on its own. Hypothetically, some other robot may meet the first robot at some point over the course of execution of the coverage process (that is that robot would have some identical state-time pairs in their trajectories). Assuming deterministic planning, (and recalling that there is no coordination between robots) those robots would continue to execute the same trajectory forever, after meeting. As such, joint states where multiple robots share the same positions form a sort of equilibrium (or invariance) in these systems. Then, for target coverage systems without coordination, many robots may congregate at single states associated with high rewards, leading to poor coverage performance for the joint system.[4]

b) Target coverage as information gathering: The target coverage problems we are interested in can also be written as information gathering problems (e.g. like the target tracking problems we discuss) *with a performance criterion based on entropy that is equivalent to original criterion for coverage*.[5] For example, imagine that each of the targets (or other units) being covered is associated with a random value (say a light that may be red or green) that is independent across time-steps. By covering and observing a target a robot would then reduce uncertainty by one bit so that the mutual information (2) and coverage objectives (3) we discuss would be equivalent. A planner maximizing this mutual information for target coverage would then behave identically to the corresponding planner maximizing coverage. As such, the target coverage problem discussed in Sec. III-B can be seen as a special case of an information gathering problem as in Sec. III-A

Going further, we can think of target coverage as an extreme case of information gathering with a very high entropy rate [26] due to uncertain variables (target colors) being independent across steps. Given this realization, one line of future work may be to characterize different kinds of information gathering problems in terms of entropy rates of the targets, whereas a low entropy rate may be conducive to implicit coordination, and high entropy rates may lead to scenarios that benefit more significantly from explicit coordination or sequential planning.

[4]In fact, steady-state performance (coverage) per robot may approach zero for large numbers of robots if all congregate at the same location.

[5]Specifically, the number of bits of entropy would be equal to the number of uncovered targets.

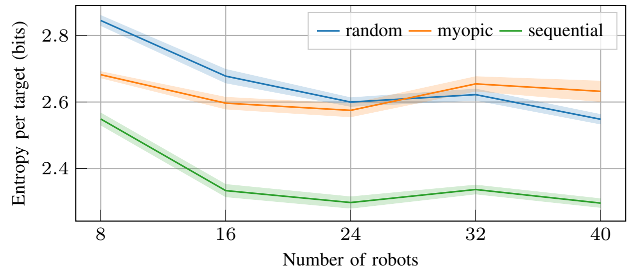


Fig. 1: Representative results for target tracking adapted from [7]. These results depict average uncertainty in target positions (entropy) for time-steps 20–100 of a target tracking task. Shaded regions depict standard error.Although robots either move randomly or perform some form of receding-horizon planning based on shared estimates of target positions. While explicit, *sequential*, coordination does improve performance, the overall reduction in uncertainty is small.

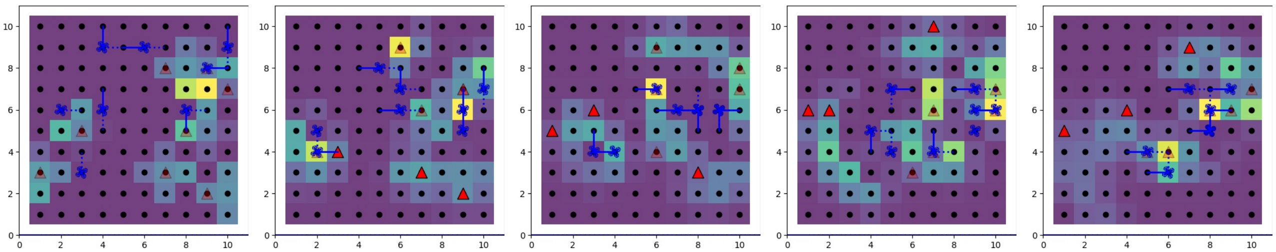
V. Initial results

For the purpose of this work, we present initial results contrasting behavior of target coverage and target tracking systems with and without explicit coordination (*myopic* and *sequential* planning (1) respectively). First, we highlight results for target tracking with and without coordination adapted from [7] in Fig. 1. In this problem, robots and targets are distributed over a two-dimensional grid, and robots seek to maximize information gain (via receding horizon planning with Monte-Carlo tree search) from noisy range observations while the targets move about randomly (robots and targets each move to adjacent cells in four cardinal directions or stay in place). Although coordination via sequential planning does improve (reduce) uncertainty in target positions, that reduction is small (about 0.25 bits).[6]

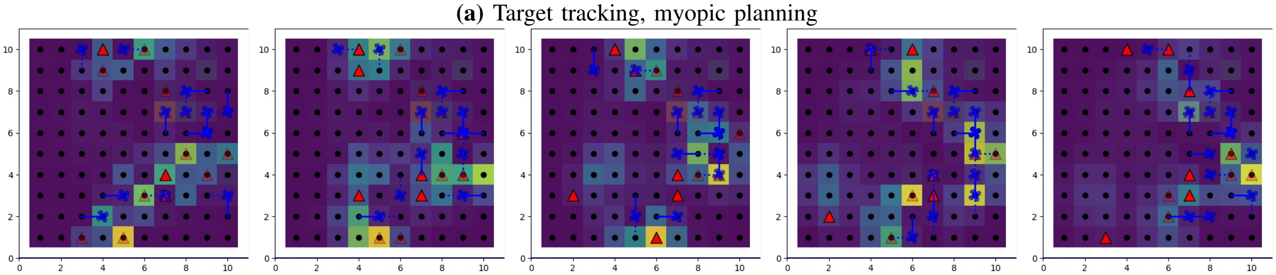
We also consider a similar scenario that adapts the prior problem to target coverage. Robots now are given access to target positions and seek only to cover targets. Specifically, robots cover targets within a radius of 2 cell widths. Fig. 2 then illustrates the different behaviors of tracking and coverage systems, each with 8 robots and 8 targets. While the target tracking systems behave similarly,[7] *differences in behavior of target coverage with and without coordination are immediately apparent and drastic*. By the 9^{th} (and last) time-step, the robots planning myopically are all approaching the same location while covering only half of the targets. For contrast, with coordination, the robots cover no less than 6 of 8 targets.

[6]To interpret these results, average entropy between 2 and 3 corresponds to a uniform distribution over 4–8 possible target locations. A difference in uncertainty of 0.25 bits would correspond to a 16% reduction in possible target locations or (given this is a two-dimensional problem) an 8% reduction in a radius representing uncertainty in target position.

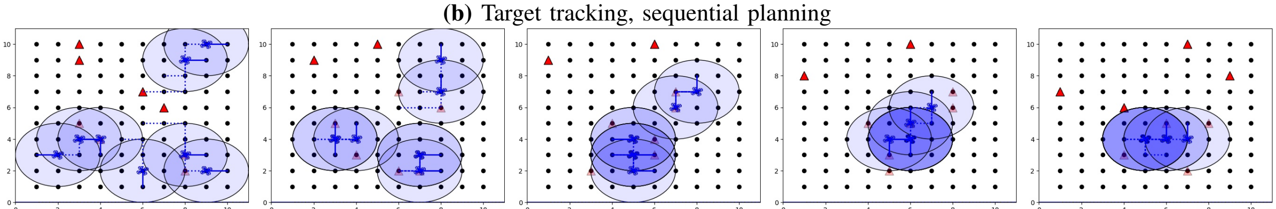
[7]Robots without coordination are perhaps more bunched up, but that may be due to different target positions. Figure 1 is a more accurate reference for target tracking performance.



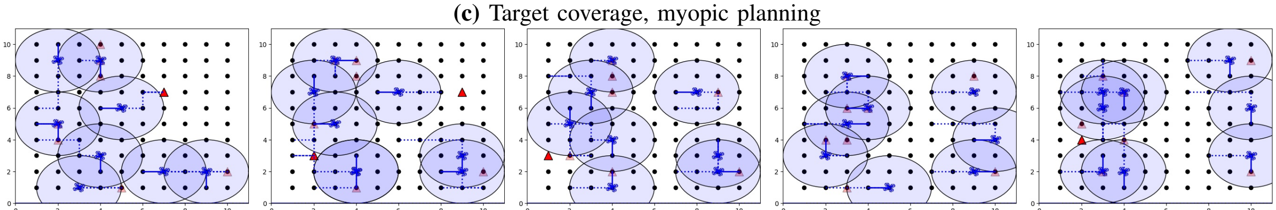
(a) Target tracking, myopic planning



(b) Target tracking, sequential planning



(c) Target coverage, myopic planning



(d) Target coverage, sequential planning

Fig. 2: Qualitative comparison of target tracking and target coverage. Each sequence shows 8 robots tracking or covering 8 targets at every other time-step from 1–9. Targets (red) are highlighted when uncovered or if entropy is above 2.8 bits. Target tracking behavior appears similar regardless of whether (a) planning without coordination or (b) via sequential planning as both benefit from shared estimates of target states. On the other hand, target coverage (c) without coordination **performs poorly** and *quickly converges toward an equilibrium with all robots at the same location* while (d) the corresponding sequential planner **performs well** with most targets covered and robots distributed well over the targets.

VI. Relevance for Systems and Infrastructure

We argue that applications such as cinematography and motion capture are (or should be) exceedingly important to the community studying multi-robot perception planning and informative path planning. Likewise, we suggest that infrastructure for evaluating multi-robot collaboration should also consider these applications. Toward this end, designers of such infrastructure may wish to consider some of the following:

- General support for view planning and visual coverage
- Support for multi-camera systems and high-bandwidth communication
- Deployment of multi-robot systems alongside human subjects or actors
- Integration with marker-based or fixed-camera [27] motion capture systems for validation
- Development of photo-realistic simulation scenarios relevant to cinematography such as involving AirSim [28]

VII. Conclusions and Future Work

This work has studied different kinds of perception planning problems, target tracking and target coverage, and highlighted significant differences between the two for robots planning with and without coordination. As such, explicit coordination, such as via greedy methods for submodular maximization, is particularly important to problems that are similar to target coverage, such as cinematography applications. In the future, we would like to develop more formal and general analysis of the equilibrium dynamics for target coverage and other perception planning problems. We are also interested in continuing to develop systems for multi-robot cinematography and related applications [29], and we hope that this line of work will spur interest in improving optimization performance in these systems such as by applying advanced algorithms for submodular maximization such as the continuous greedy algorithm [30, 31].

References

[1] R. Bonatti, Y. Zhang, S. Choudhury, W. Wang, and S. Scherer, “Autonomous drone cinematographer: Using artistic principles to create smooth, safe, occlusion-free trajectories for aerial filming,” in *Proc. of the Intl. Sym. on Exp. Robot.*, 2018.

[2] K. Goel, M. Corah, C. Boirum, and N. Michael, “Fast exploration using multirotors: Analysis, planning, and experimentation,” in *Field and Service Robotics*, Tokyo, Japan, Aug. 2019.

[3] W. Tabib, K. Goel, J. Yao, C. Boirum, and N. Michael, “Autonomous cave surveying with an aerial robot,” *IEEE Trans. Robotics*, 2021.

[4] T. Dang, M. Tranzatto, S. Khattak, F. Mascarich, K. Alexis, and M. Hutter, “Graph-based subterranean exploration path planning using aerial and legged robots,” *J. Field Robot.*, vol. 37, no. 8, pp. 1363–1388, 2020.

[5] F. Sukkar, G. Best, C. Yoo, and R. Fitch, “Multi-robot region-of-interest reconstruction with Dec-MCTS,” in *Proc. of the IEEE Intl. Conf. on Robot. and Autom.* IEEE, 2019, pp. 9101–9107.

[6] M. Corah, “Sensor planning for large numbers of robots,” Ph.D. dissertation, Carnegie Mellon University, 2020.

[7] M. Corah and N. Michael, “Scalable distributed planning for multi-robot, multi-target tracking,” in *Proc. of the IEEE/RSJ Intl. Conf. on Intell. Robots and Syst.*, Prague, Czech Republic, Sep. 2021.

[8] A. Singh, A. Krause, C. Guestrin, and W. J. Kaiser, “Efficient informative sensing using multiple robots,” *J. Artif. Intell. Res.*, vol. 34, pp. 707–755, 2009.

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

[10] B. Schlotfeldt, V. Tzoumas, and G. J. Pappas, “Resilient active information acquisition with teams of robots,” *IEEE Trans. Robotics*, 2021.

[11] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, “An analysis of approximations for maximizing submodular set functions-I,” *Math. Program.*, vol. 14, no. 1, pp. 265–294, 1978.

[12] M. L. Fisher, G. L. Nemhauser, and L. A. Wolsey, “An analysis of approximations for maximizing submodular set functions-II,” *Polyhedral Combinatorics*, vol. 8, pp. 73–87, 1978.

[13] A. Krause and C. E. Guestrin, “Near-optimal nonmyopic value of information in graphical models,” in *Proc. of the Conf. on Uncertainty in Artif. Intell.*, Edinburgh, Scotland, 2005.

[14] N. A. Atanasov, J. Le Ny, K. Daniilidis, and G. J. Pappas, “Decentralized active information acquisition: Theory and application to multi-robot SLAM,” in *Proc. of the IEEE Intl. Conf. on Robot. and Autom.*, Seattle, WA, May 2015.

[15] G. Hollinger, S. Singh, J. Djugash, and A. Kehagias, “Efficient multi-robot search for a moving target,” *Intl. Journal of Robotics Research*, vol. 28, no. 2, pp. 201–219, 2009.

[16] V. Tzoumas, A. Jadbabaie, and G. J. Pappas, “Resilient non-submodular maximization over matroid constraints,” *arXiv preprint arXiv:1804.01013*, 2018.

[17] V. Tzoumas, K. Gatsis, A. Jadbabaie, and G. J. Pappas, “Resilient monotone submodular function maximization,” in *Proc. of the IEEE Conf. on Decision and Control*. IEEE, 2017, pp. 1362–1367.

[18] B. Gharesifard and S. L. Smith, “Distributed submodular maximization with limited information,” *IEEE Trans. Control Netw. Syst.*, 2017, to appear.

[19] D. Grimsman, M. S. Ali, J. P. Hespanha, and J. R. Marden, “The impact of information in greedy submodular maximization,” *IEEE Transactions on Control of Network Systems*, 2018.

[20] M. Corah and N. Michael, “Distributed submodular maximization on partition matroids for planning on large sensor networks,” in *Proc. of the IEEE Conf. on Decision and Control*, Miami, FL, Dec. 2018.

[21] ——, “Active estimation of mass properties for safe cooperative lifting,” in *Proc. of the IEEE Intl. Conf. on Robot. and Autom.* Singapore: IEEE, May 2017.

[22] ——, “Volumetric objectives for multi-robot exploration of three-dimensional environments,” in *Proc. of the IEEE Intl. Conf. on Robot. and Autom.*, Xi’an, China, May 2021.

[23] L. Zhou, V. Tzoumas, G. J. Pappas, and P. Tokekar, “Resilient active target tracking with multiple robots,” *IEEE Robot. Autom. Letters*, vol. 4, no. 1, pp. 129–136, 2019.

[24] P. Tokekar, V. Isler, and A. Franchi, “Multi-target visual tracking with aerial robots,” in *Proc. of the IEEE/RSJ Intl. Conf. on Intell. Robots and Syst.*, Chicago, IL, Sep. 2014.

[25] A. Bucker, R. Bonatti, and S. Scherer, “Do you see what i see? coordinating multiple aerial cameras for robot cinematography,” in *Proc. of the IEEE Intl. Conf. on Robot. and Autom.* IEEE, 2021, pp. 7972–7979.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY: John Wiley & Sons, 2012.

[27] H. Joo, H. Liu, L. Tan, L. Gui, B. Nabbe, I. Matthews, T. Kanade, S. Nobuhara, and Y. Sheikh, “Panoptic Studio: A massively multiview system for social motion capture,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2015.

[28] S. Shah, D. Dey, C. Lovett, and A. Kapoor, “AirSim: High-fidelity visual and physical simulation for autonomous vehicles,” in *Field and Service Robotics*, 2017.

[29] C. Ho, A. Jong, H. Freeman, R. Rao, R. Bonatti, and S. Scherer, “3D human reconstruction in the wild with collaborative aerial cameras,” in *Proc. of the IEEE/RSJ Intl. Conf. on Intell. Robots and Syst.* IEEE, 2021, pp. 5263–5269.

[30] G. Calinescu, C. Chekuri, M. Pal, and J. Vondrák, “Maximizing a monotone submodular function subject to a matroid constraint,” *SIAM J. Comput.*, vol. 40, no. 6, pp. 1740–1766, 2011.

[31] A. Robey, A. Adibi, B. Schlotfeldt, G. J. Pappas, and H. Hassani, “Optimal algorithms for submodular maximization with distributed constraints,” *arXiv preprint arXiv:1909.13676*, 2019.