2023 IEEE International Conference on Robotics and Automation (ICRA)
Workshop on Active Methods in Autonomous Navigation
29 May – 2 June, 2023, London, UK

Learning Simultaneous Motion Planning and Active Gaze Control for Persistent Monitoring of Dynamic Targets

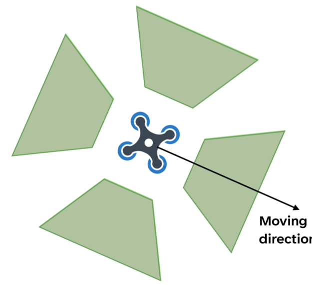
Yizhuo Wang[1], Guillaume Sartoretti[1]

Abstract— **We consider the persistent monitoring problem of a given set of targets moving based on an unknown model by an autonomous mobile robot equipped with a directional sensor (e.g., camera). The robot needs to actively plan both its path and its sensor's gaze/heading direction to detect and constantly re-locate all targets by collecting measurements along its path, to keep the estimated position of each target as accurate as possible at all times. Our recent work discretize the monitoring domain into a graph where the deep-reinforcement-learning-based agent sequentially decide which neighboring node to visit next. In this work, we extend it by duplicating neighboring features multiple times and then combining with its unique gaze features to output a joint decision of "where to go" and "where to look". Our simulation experiments show that active gaze control enhances monitoring performance, particularly in terms of minimum number of re-observation per target, compared to agent with a fixed forward-gaze sensor or greedy gaze selection.**

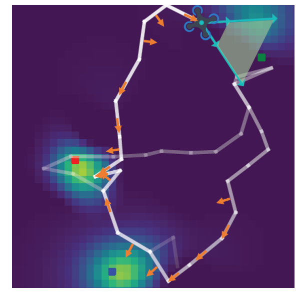
I. Introduction

Persistent monitoring refers to problems where autonomous robots equipped with onboard sensors are tasked with collecting data to maintain an accurate awareness (i.e., belief) over a given domain. There, the use of mobile robots provides greater flexibility in deployment and can handle a broader spectrum of tasks compared to fixed sensor network [1], [2]. In our recent work [3], we introduce a neural approach based on deep reinforcement learning (DRL) to tackle the single-agent, multi-mobile-target persistent monitoring problem. We use cascaded Transformer blocks [4] to let the agent reason about which target, time, and location to attend to across multiple scales, which we show also helps relax the usual limitations of a finite target set. We notice that due to limited payload capacity, lightweight unmanned aerial vehicles (UAVs) are often incapable of carrying heavy, omnidirectional LiDAR sensors, but only lighter, directional sensors like cameras. Therefore, in this paper, we specifically investigate the deployment of a lightweight UAV with a directional sensor to persistently monitor a set of mobile targets moving over a given domain, where the robot must simultaneously plan its motion as well as actively control the gaze of its directional sensor. Specifically, we consider a quadrotor equipped with a binary directional sensor with limited Field-of-View (FoV) [5] (e.g., a camera with a visual classifier for target detection), with this sensor fixed in a certain position relative to the UAV's body frame (e.g., front-facing). Since our quadrotor is omnidirectional, we allow it to fly to its next waypoint into each of its own four

[1]Authors are with the Department of Mechanical Engineering, College of Design and Engineering, National University of Singapore wy98@u.nus.edu, mpegas@nus.edu.sg



(a)



(b)

Fig. 1: **Discrete UAV gazes.** (a) The four possible gaze directions of the UAV (green trapezoidal shades), parallel or perpendicular to the moving direction. (b) Agent belief of all targets and executed trajectory so far (white lines with growing opacity) with corresponding gaze direction (orange arrows); agent position; agent FoV (green shade); and targets (colorful squares) unknown to the agent (except within agent's FoV).

cardinal directions, i.e., the robot can fly forward, sideways to the left/right, or even backwards, allowing it to control the gaze/heading of its onboard directional sensor, as illustrated in Figure 1.

Aligned with our prior work [3], the agent may or may not know the number and initial positions of the targets *a priori*, while the underlying motion model/dynamics of the targets is always assumed unknown. Based on measurements obtained along its path, the agent must build and update a time-dependent *belief* of each individual target location, to reason about/predict their possible location to frequently re-visit/-locate each of them. To achieve this, our agent aims to maximize *information gain* in the vicinity of the targets (i.e., minimizes the uncertainty over the true target positions) [6], [7]. To manage growing uncertainty for targets outside its FoV, the agent must temporarily stop tracking some of the targets to focus on others, and then return to the untracked ones as soon as possible to re-locate them. We note that the measurements obtained by the agent can significantly depend on the drone's gaze, highlighting the importance of selecting not only the correct UAV location, but also its correct gaze direction. In doing so, an optimal planner must balance exploration and exploitation to *jointly* plan the shortest path and gaze direction in order to cover all targets' predicted areas.

We rely on our spatio-temporal attention network [3], in which we propose to copy each of the agent's current neighboring nodes four times, each with a different gaze direction, to finally output the next neighboring node to visit and the associated gaze to adopt along the way. In our experiment, we discover that compared to fixed forward gaze and greedy gaze selection (i.e., gaze with the highest belief value), allowing active gaze control decreases the target position uncertainty and significantly increase the chance of observing targets in the agent's limited sensor footprint.

II. Problem Setup and Representation

We consider a given set of N mobile targets ($i \in \{1, \ldots, N\}$) moving in a bounded 2D environment. To interpolates the binary measurements and provides richer information about the process, we adopt N Gaussian Processes (GPs) to model the agent belief (see [3] for details).

The agent is equipped with an accurate *binary* sensor that produces measurements $z_{i,t} \in \{0, 1\}$ indicating the presence/absence of target i in the sensor's FoV/footprint $S(x_t)$. Different from the circular FoV used in [3], we now assume a directional sensor (here, sector footprint with 60 degrees horizontal FoV), resulting in a trapezoidal-shaped projected FoV on the 2D search domain. If target i is within the sensor's FoV, a 1 measurement will be obtained at that target's current position, otherwise a 0 measurement is obtained at the agent's current position. The collected timestamped observed locations and their corresponding measurements are then utilized to compute the GP posterior for each target. Note that the number of data entries is truncated according to their recency in time to maintain computational tractability.

III. Persistent Monitoring as a DRL Problem

We use the same approach of casting persistent monitoring into the DRL framework as [3]. In short, we firstly discretize the search domain into a graph with k nearest neighboring nodes connected ($k = 10$ in practice), formulating it as a sequential decision-making problem. Next, we augment the graph with the agent's belief modelled as Gaussian Processes with additional future prediction and agent past trajectory. Upon reaching the previously selected node, the agent chooses which neighboring node to visit next and moves toward it along a straight line, sequentially constructing the agent's trajectory. The agent decisions are reinforced by a reward signal based on the uncertainty reduction in the true vicinity areas of all targets. Specifically, based on our selected Matérn kernel, our reward is a non-negative concave function with time, which naturally incentivizes the agent to visit all mobile targets quickly and frequently.

Our proposed spatio-temporal attention neural network consists of multiple cascaded Transformer blocks [4], sequenced into target, temporal, and spatial encoders, and finally a decoder, to capture dependencies across targets, time, and space. Compared to these dependency features that encode the global *context* of the agent's belief, the gaze direction is a shorter-term information which mainly affects the agent's decision-making in the short term. That is, the robot does not require gaze information for distant nodes, but only for its immediate neighboring nodes. Hence, we only add gaze features to the agent's neighboring nodes in our final decoder, which yields the agent's policy.

Agent needs to be aware of its neighboring node features with information within each gaze direction to decide which neighboring node and which sensor's heading to orient. We extract gaze-specific features $\mathbf{h}_j$ at agent's neighboring node j from the agent belief. To facilitate learning, the feature $\mathbf{h}_{j,g}$ of gaze index g consists of (1) the maximum predicted value inside that footprint, (2) relative angle of that maximum value, (3) gaze direction in world frame, (4) distance to that maximum value, and (5) the gaze index g. Since the encoder output does not contains gaze information, we duplicate each neighboring node feature ($\tilde{\mathbf{h}}_j^{STG}$ in [3]) four times and then concatenate with their unique gaze features $\mathbf{h}_j$, obtain $\tilde{\mathbf{h}}_{j,g}^{STG}$. By doing so, $\tilde{\mathbf{h}}_{j,g}^{STG}$ now both contains spatio-temporal context and the information within the gaze directions of each neighboring nodes. In the end, the decoder output a policy distribution, from which we jointly determine/sample which neighboring node to visit and which gaze direction to orient. Note that we also attempt to output the next neighboring node first and then output its associated gaze direction autoregressively. However, it turns out to perform worse than simple duplication due to the order constraint it imposes. This implies that the decision of where to go and where to look needs to be inferred together, other than in sequence, which potentially limits the selection of a perfect gaze but at a less informative area.

We adopt the PPO algorithm to train our neural network [8], details are similar to [3].

IV. Experiments

In our experiment, we adopt the overall average uncertainty in all target areas (lower is better) **Unc** and the minimum number of target re-observations obtained among all targets (higher is better) **MinOb** as metrics.

We report the comparison results of our active gaze control with two baseline methods, one employing *fixed forward-gaze* and the other using *greedy gaze* towards the highest belief value. The results are shown in Table I, with variations in the number of targets N and the speed ratio between the targets and the agent r_v (all instances with $|V| = 200$ nodes and $T = 100$ history steps). As we see, except for highly dynamic environments, active gaze/heading selection improves the performance by a large margin, especially in terms of **MinOb**. We also observe that the greedy gaze selection works particularly effective in the search domain where there are fewer targets in the search domain. This can be attributed to the spatial sparsity of the targets, which contributes to their increased ease and predictability of being located. Overall, the comparison implies that our network effectively facilitates the learning of an active gaze for the robot to observe possible target positions, likely by further relaxing the constraints imposed by our graph discretization. That being said, the robot's active gaze control allows it to sense areas that were previously unobservable or poorly discretized, expanding its effective monitoring domain. Furthermore, it is important to note that the gaze direction with

TABLE I: **Comparison between our learned active gaze control and an agent with fixed-forward sensor (100 instances each).** We report the overall average uncertainty at each target area (standard deviation of uncertainty between targets in parentheses) and the minimum number of re-observation among all targets.

Metric	**Unc**				**MinOb**			
Speed Ratio r_v	1/30	1/20	1/10	1/7	1/30	1/20	1/10	1/7
Sensor	Target number $N = 2$							
Fixed gaze	0.664(0.059)	0.687(0.070)	0.743(0.092)	0.795(0.094)	25.01	25.00	16.49	12.39
Greedy gaze	0.611(0.062)	0.633(0.083)	**0.718**(0.133)	**0.763**(0.130)	**50.34**	47.83	**32.12**	**26.74**
Active gaze	**0.610**(0.063)	**0.623**(0.072)	0.737(0.108)	0.801(0.103)	49.03	**49.74**	30.13	20.74
Sensor	Target number $N = 4$							
Fixed gaze	0.721(0.102)	0.736(0.112)	0.783(0.126)	**0.814**(0.120)	9.67	9.01	7.47	6.63
Greedy gaze	0.703(0.126)	0.724(0.141)	0.785(0.152)	0.819(0.140)	15.22	13.34	10.08	**8.14**
Active gaze	**0.696**(0.121)	**0.711**(0.130)	**0.774**(0.139)	0.826(0.125)	**15.66**	**15.12**	**11.98**	8.13
Sensor	Target number $N = 6$							
Fixed gaze	0.750(0.112)	0.764(0.119)	0.795(0.127)	**0.816**(0.126)	6.28	5.71	5.41	4.76
Greedy gaze	0.746(0.139)	0.758(0.147)	0.810(0.147)	0.830(0.143)	6.80	6.81	6.00	5.83
Active gaze	**0.732**(0.132)	**0.746**(0.138)	**0.794**(0.139)	0.830(0.133)	**9.10**	**8.33**	**7.62**	**6.59**
Sensor	Target number $N = 8$							
Fixed gaze	0.761(0.116)	0.775(0.124)	**0.803**(0.128)	**0.821**(0.129)	4.52	4.28	4.52	4.17
Greedy gaze	0.763(0.141)	0.778(0.146)	0.815(0.147)	0.837(0.141)	4.72	4.67	4.85	4.47
Active gaze	**0.753**(0.131)	**0.765**(0.140)	0.808(0.141)	0.829(0.136)	**6.47**	**6.07**	**6.32**	**4.98**

TABLE II: **Active gaze distribution over 100 instances.**

Gaze direction	Front	Rear	Left	Right
Percentage	45.0%	15.2%	19.7%	20.1%

the highest value in the agent's belief does not necessarily guarantee the best choice. Greedy gaze selection may suffer from inaccurate belief and the inability to effectively observe all targets, while active gaze control learns to balance the observation of all potential target locations instead of solely focusing on a specific position.

We further examine the gaze distribution throughout the persistent monitoring mission. Table II reports the gaze distribution with $|V| = 200, T = 100, N = 4$ and $r_v = 1/20$. There, we observe that almost half of the path segments were covered by the forward gaze ($g = 1$), while the remaining three gazes had a comparable proportion, accounting for approximately $15\% - 20\%$ each. Although there are no bias towards any particular gaze, it appears that the agent learned to prefer gaze that aligns with its direction of motion. More tests are needed, but we believe that the cause of this phenomenon is mainly twofold. Firstly, the area of the world covered by the agent's sensor is most drastically changed through time along the agent's main moving direction, resulting in a broader sensing area to persistently monitor the search domain. That is, the sensor overlap between successive timesteps is larger when using the left/right/rear gazes, thus offering the agent less new information; this seems to result in the agent learning to sparingly adopt those gaze directions, except when the agent requires repeated confirmation of target positions, in favor of generally more informative ones (front). Secondly, the agent generally learns to actively select its gaze that points to the nearest estimated target position, which is especially evident when there are fewer targets, as depicted in Figure 1(b). Since the agent learns to sequentially visit targets along the shortest path, which often follows a straight line, the next target is most likely to be positioned ahead of the agent. Consequently, selecting the forward gaze emerges to be the optimal choice in such common scenarios.

V. Conclusion and Future Works

In this work, we extend our spatio-temporal attention network to allow an autonomous quadrotor to actively select its gaze in persistent monitoring tasks. To this end, the decoder takes the gaze-specific features of each neighboring node as input, enabling the agent to make joint decisions on where to go and where to look simultaneously. Our simulation experiments demonstrate the effectiveness of active robot gaze selection, showing how our approach significantly improves the performance in almost every environment compared to fixed-front gaze and greedy (next-best) gaze.

Future work will primarily look at more flexible gaze control that allows the robot to adjust its gaze in a continuous space. We also plan to train a model with task-specificity and implement/deploy it on the drones.

References

[1] Y. Sung and P. Tokekar, "Algorithm for searching and tracking an unknown and varying number of mobile targets using a limited fov sensor," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 6246–6252.

[2] Z. Kashino, G. Nejat, and B. Benhabib, "A hybrid strategy for target search using static and mobile sensors," *IEEE Transactions on Cybernetics*, vol. 50, no. 2, pp. 856–868, 2020.

[3] Y. Wang, Y. Wang, Y. Cao, and G. Sartoretti, "Spatio-temporal attention network for persistent monitoring of multiple mobile targets," 2023. [Online]. Available: https://arxiv.org/abs/2303.06350

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[5] H. Coffin, I. Abraham, G. Sartoretti, T. Dillstrom, and H. Choset, "Multi-agent dynamic ergodic search with low-information sensors," in *2022 IEEE International Conference on Robotics and Automation (ICRA)*, 2022, pp. 11 480–11 486.

[6] M. Popović, T. Vidal-Calleja, G. Hitz, J. J. Chung, I. Sa, R. Siegwart, and J. Nieto, "An informative path planning framework for uav-based terrain monitoring," *Autonomous Robots*, vol. 44, pp. 889–911, 2020.

[7] Y. Cao, Y. Wang, A. Vashisth, H. Fan, and G. A. Sartoretti, "CAtNIPP: Context-aware attention-based network for informative path planning," in *6th Annual Conference on Robot Learning*, 2022.

[8] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.