On-line Human Intervention for Robot Behavior Adaptation

Maggie Wigness[1] and John G. Rogers III[1]

***Abstract*— Unmanned ground vehicles operating in unstructured environments as part of a human-robot team should be expected to learn and adapt on-line during operation just as their human counterparts. Further, this on-line adaptation should require minimal human teammate oversight to ensure each teammate can maintain an appropriate level of awareness. We discuss how human intervention can be used during on-line operation to quickly correct and adapt navigation behaviors learned from inverse reinforcement learning. We discuss the relevance of adaptation with respect to long-term environment changes, domain shifts, and novel information discovery.**

I. INTRODUCTION

When human-robot teams operate together in a dynamic environment, it can be expected that behaviors of each teammate will need to adapt based on the current environment and mission status. For example, severe weather or seasonal shifts could drastically change the terrain, causing areas that were previously traversable to present safety issues. Further, novel information, terrain or objects may be encountered throughout operation that require adaptation.

In our previous work [1], we described how to use inverse reinforcement learning to quickly train reward functions that encode traversal behaviors, e.g., driving near the edge. Our approach was based on the principal of maximum entropy [2], and was shown to learn reliable reward functions with only a few human demonstrations. Although this learning technique required minimal human effort, the training process was performed completely off-line through batch processing a set of demonstrated trajectories.

For human-robot teams to collaborate at high operational tempo, it will be necessary for the robot teammates to adapt on-line when their behaviors deteriorate due to changes in the environment or mission requirements. We discuss how our previous learning from demonstration technique can be modified for learning from intervention. In this case, the human teammate has the ability to interrupt autonomous robot operation at any moment to provide corrected traversal behavior. Human interventions allow the robot to gain additional training exemplars without being pulled from operation, and instead, use the trajectories provided during intervention to directly update the behavior model with onboard processing.

II. RELATED WORK

Learning from intervention has been used for a variety of tasks. Saunders et al. [3] discuss the utility of learning from intervention and its ability to reduce catastrophic failures when intelligent systems are learning. The combination of demonstration and intervention has been explored in simulation for an autonomous perching task [4], and for robot manipulation and gesture tasks [5]. We use similar human intervention styles as these approaches, but focus on adapting traversal behaviors on a mobile robot in a real-world environment.

[1]U.S. Army Research Laboratory, Adelphi, MD, USA {maggie.b.wigness,john.g.rogers59}.civ@mail.mil

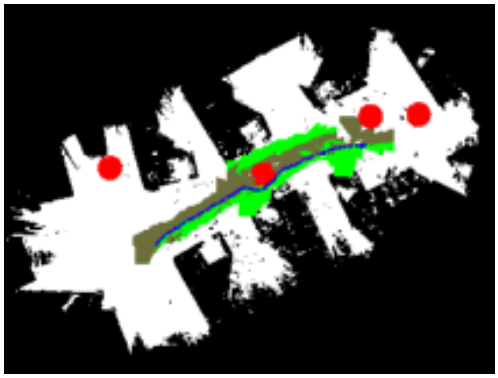

Fig. 1. Illustration of four environment feature maps used for the inverse reinforcement learning of traversal behaviors. Human demonstration of "optimal" traversal behavior (denoted by the blue line) is performed with respect to the features obstacles (black), grass (green), road (gold), and dangerous regions (red).

III. EXPERIMENTS

We use our previous learning from demonstration setup [1] that learns a reward function given 1) a set of demonstrated trajectories that exemplify "optimal" traversal patterns for the desired behavior, and 2) a set of environment feature maps, i.e., binary occupancy grids that encode the presence or absence of specific terrain and objects. Fig. 1 shows a color-coded overlay of four environment feature types used in the behavior learning: obstacles (black), grass (green), road (gold), and dangerous regions (red).

We define an experimental scenario in which the robot has been deployed to execute a task with a traversal behavior reward function that was learned from human demonstration off-line. This reward function encodes the "edge of road" traversal behavior, where the robot should navigate throughout the environment while trying to maintain close proximity to the road's edge. During operation, the robot's behavior needs to be adapted as additional environment information is encountered. Specifically, the human-robot team is given additional information about potentially dangerous regions in the environment. For experimentation, these dangerous areas are encoded as circular regions with a pre-defined radius.

Fig. 2 depicts the traversal behavior planned by the robot in the presence of these dangerous regions given its original

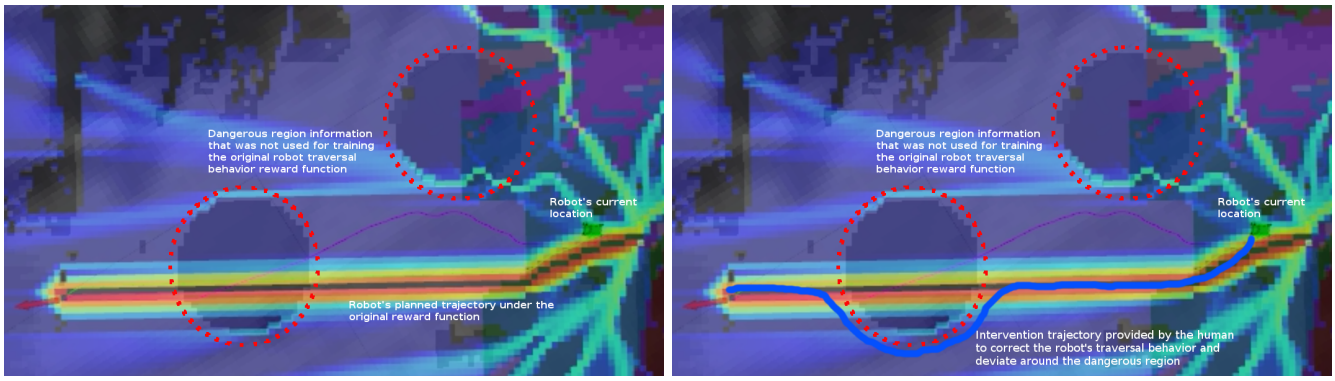

Fig. 2. Visualization of the on-line intervention a human provides to its robot teammate to adapt its traversal behavior reward function. (Left) The path planned by the robot given its current learned behavior, which was not trained to consider knowledge of dangerous regions. (Right) Illustration of the trajectory the human teammate provides during intervention of the robot's execution. The robot uses this trajectory provided during intervention to update its traversal behavior.

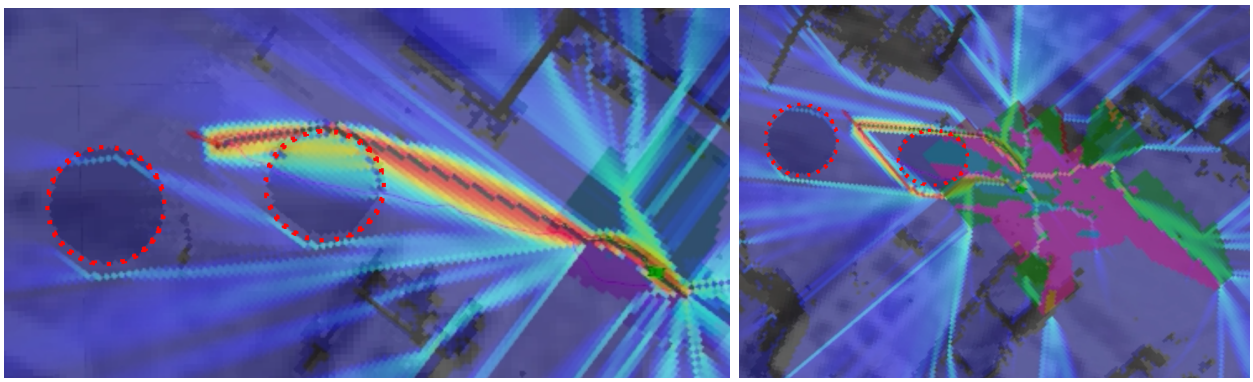

Fig. 3. Visualization of the reward map generated after human intervention to adapt the robot's behavior such that it avoids dangerous regions within the area.

reward function learning (left), and the intervention provided by the human to adapt this behavior (right). The dangerous regions are seen as dark circles in the robot's map and outlined in red for easier visualization. Using its original reward function (illustrated as the heat map in the figure) the robot plans a trajectory that would pass straight through one of the dangerous regions. As the human teammate is overseeing the robot behavior, the human immediately intervenes and redirects the robot's trajectory (shown as the blue line) around this region. This interaction is recorded by the robot and used to update its traversal reward function.

The trajectory and feature maps from the new human intervention is added to the collection of previous examples. In the current implementation, the training routine is executed using the prior reward function weights as an initialization point after each new example is added; however, this optimization is limited to 30 seconds to facilitate fast updates. After only two interventions, the updated robot reward map can be seen in Fig. 3. Here, the reward function clearly exhibits a strong preference to avoid the dangerous region.

IV. CONCLUSIONS AND FUTURE WORK

The use of human intervention allows for on-line adaptation of traversal behaviors for a mobile robot with minimal human oversight. The intervention is similar to off-line demonstration, but the robot is able to take the input and directly process it on-line to update its model. Moving forward we will investigate how best to interleave demonstration and intervention, and identify visualizations and techniques that allow the human teammate to immediately evaluate the learning performance so as not to prematurely release the robot back into operation.

REFERENCES

[1] M. Wigness, J. G. Rogers, and L. E. Navarro-Serment, "Robot navigation from human demonstration: Learning control behaviors," in *Proceedings of the Int. Conf. on Robotics and Automation*. IEEE, 2018.

[2] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning." in *Proceedings of the Conf. on Artificial Intelligence*, vol. 8. AAAI, 2008, pp. 1433–1438.

[3] W. Saunders, G. Sastry, A. Stuhlmueller, and O. Evans, "Trial without error: Towards safe reinforcement learning via human intervention," in *Proceedings of the Int. Conf. on Autonomous Agents and MultiAgent Systems*, 2018, pp. 2067–2069.

[4] V. G. Goecks, G. M. Gremillion, V. J. Lawhern, J. Valasek, and N. R. Waytowich, "Efficiently combining human demonstrations and interventions for safe training of autonomous systems in real-time," in *AAAI Conf. on Artificial Intelligence*, 2019.

[5] B. Akgun, M. Cakmak, J. W. Yoo, and A. L. Thomaz, "Trajectories and keyframes for kinesthetic teaching: A human-robot interaction perspective," in *Proceedings of the Int. Conference on Human-Robot Interaction*. ACM, 2012, pp. 391–398.