Teaming Heterogeneous Ground and Micro-Aerial Robots for Following of Non-Cooperative Agents

Ori A. Miller[1], Jason M. Gregory[2], and Christopher Reardon[1]

***Abstract*— From nursing homes to war zones, keeping a close eye on someone's whereabouts is invaluable and often necessary. In most cases, the use of security cameras or even ground robots would be ideal. However, a lack of control of the environment may prohibit cameras. Further, in situations where the environment is unknown or unpredictable, knowledge of the layout may be not available nor accurate. This poses a challenge for a single ground robot to simultaneously navigate and track an agent. To address these challenging constraints, we explore preliminary work in developing a multi-robot system composed of a slow-moving ground robot and an agile aerial robot. The described system utilizes the object recognition capabilities of the ground robot to locate and begin tracking an agent – in our case, a human – and spatially task the aerial robot to get within close proximity of the human to then begin cooperatively tracking. To implement this solution, we address multiple challenges regarding cooperatively tracking a human from both the ground robot's perspective and the aerial robot's perspective.**

I. INTRODUCTION

The field of unmanned mobile robotics has experienced substantial growth in the last few decades. Researchers have tackled more complex problems involving complicated constraints, like restricted environments and differently motivated agents [1]. A recent and important problem is human-aware robot navigation [2]. Human-aware robot navigation involves perceiving and navigating safely among people. Delivering food on city streets or following a nurse through a hectic hospital, both are under the umbrella of human-aware navigation with the latter being person-following. Person-following is a branch of human-aware navigation where the objective is to navigate within a desired proximity of a moving person [3]. As person-following tasks become more complex and involve more constrained environments, sub-tasks can become too difficult to optimally run concurrently. Heterogeneity can be implemented in to systems to be used for human-robot cooperative tasks. A nurse in a field hospital leading an x-ray machine carrying robot to an operating room, or a first responder leading a robot carrying a natural disaster survivor to an exit are both examples where heterogeneity can improve task allocation by splitting the mission into sub tasks, with a fast aerial robot tasked with maintaining localization of the person in the environment while a large slow-moving ground robot is navigating towards the person..

Fig. 1. Our Unity-based simulation environment containing an autonomous ground vehicle, circled in red, an autonomous micro-aerial vehicle, circled in blue, and a person (e.g., soldier model) circled in yellow.

Many solutions to person following, some involving robot teaming and/or heterogeneous systems, have been proposed. There is no one solution that optimally fits all person following applications because each application has its own unique set of objectives, context, and team dynamics. Existing solutions from literature include general use single ground robots [4], general use single aerial robots [5], [6], social ground single-robots [7], [8], social ground homogeneous multi-robot teams [9], and outdoor heterogeneous multi-robot teams [10].

When tasked with following an agent, in our case a person, we cannot assume their behaviors will be compliant with a robotic system's perception and navigation limitations. To this end, we assume there is no a-priori information available to precisely predict the person's motion, and that the person's trajectory is non-ergotic, i.e. even under identical circumstances their behavior may not be deterministic.

A robotic system will have a specific set of constraints, such as payload, sensing, maneuverability, and speed limitations. Micro-aerial robots are extremely payload-limited, which in turn limits sensing and computational power and can prevent fully independent autonomous operation. For this reason, our system pairs a micro-aerial robot (UAV) with a ground robot (UGV) with greater payload capacity for more powerful sensing and compute capabilities such as mapping, navigation, and person detection. This compliments our simple UAV with the ability to localize itself and the person in the same space as the UGV.

*Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-22-2-0078. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

[1]Ori A. Miller and Christopher Reardon are with the Department of Computer Science at the University of Denver. email: {ori.miller, christopher.reardon}@du.edu

[2]Jason M. Gregory is with DEVCOM Army Research Laboratory. email: jason.m.gregory1.civ@army.mil

Finally, the environment may present its own limitations. In our application, we assume the UAV maintains a fixed height above the ground. Where in some outdoor environments a UAV may have the opportunity to gain altitude and direct its sensors downwards to get a birds-eye view of the person in the environment, we seek to address the case where this might not be possible for a variety of reasons, such as indoor settings, intermittent height obstacles such as tree canopies, or airspace restrictions. This space of aerial and ground robot teams in height restricted space is where our research sits.

Robot agents can be constrained by the nature of their desired tasks and environments they are working in. Other agents related to the robots mission can be motivated in a way to require the robots to ignore any assumptions they might have made. To address these specific velocity, height, and agent related constraints, we developed a heterogeneous multi-robot system capable of following a non-cooperative agent through a height restricted environment.

II. PROBLEM STATEMENT

We first define a team of one ground robot, r_g, and one aerial robot, r_a, a dynamically moving human, h, and a sequence S. S, shown in Fig. 2, is a sequence of n points visited by h over time. $S = \{(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)\}$, where (x_t, y_t) represents the $x-$ and $y-$coordinates of h in Cartesian space at time $t \in [0, n]$. Our objective is to minimize the distance between r_g and h while maximizing r_a's view of h over the sequence S.

It is assumed that r_g is non-holonomic, and the maximum velocity of r_g is less than that of h, i.e., $|r_{g\_v}(t)| < |h_v(t)|$ for all t. It is also assumed that r_a and h are holonomic, and the maximum velocity of h is less than that of r_a, i.e., $|h_v(t)| < |r_{a\_v(t)}|$ for all t. Both r_g and r_a have no prior knowledge of the environment, and trivial mobility challenges in the environment.

View-shed is defined as the set of points, (x, y), the robot has direct line-of-sight from their current position constrained to the field of view of the on-board camera. A robots view-shed is represented as $r_a\_vs(t)$ for the UAV and $r_g\_vs(t)$ for the UGV. h is initially located in the view-shed of r_g, $h_{x,y} \in r_g\_vs(t)$ for $t = 0$. It is assumed when h is in either the view-shed of r_g or r_a, h is localized in the frame of the team's shared map of the environment.

At each time step t, two sets of candidate goal locations, C_g, navigable by the UGV, and C_a, navigable by the UAV, are computed and a set of possible future location of h, $F_{predicted}$, is computed over a static time-horizon. C_g is composed of set of (x, y) coordinates that r_g can navigate to, and C_a has coordinates that r_a can navigate to. $F_{predicted}(t) = \{(xpred_0, ypred_0), \ldots, (xpred_{th}, ypred_{th})\}$, where th is a specified time-horizon, is computed by r_g as a list of predicted locations, (x, y), of h th seconds in the future.

r_g selects a goal location, r_g_G, from C_g that minimizes the distance between r_g and h at time t, while r_g selects a goal location , r_a_G, from C_a for r_a that maximizes coverage of r_a_vs over h's predicted motion, $F_{predicted}(t)$. Fig. 3 displays the optimal goal location selection.

The problem can be formalized as a multi-objective optimization problem, Eq. (1), where we minimize objective f_1, the distance between r_g and h over sequence S, and maximize objective f_2, the coverage of r_a_vs over h's predicted motion, $F_{predicted}(t)$, over sequence S.

$$F = (min(f_1), max(f_2)) \tag{1}$$

$$f_1 = \sum_{t=0}^{n} ||(r_g\_goal, S[t])||_2 \tag{2}$$

$$f_2 = \sum_{t=0}^{n} |F_{predicted}(t) \in r_a\_vs(r_a\_G)| \tag{3}$$

To evaluate the performance of the team we will collect the total distance traveled by r_g, time it took r_g to reach h, average distance between r_g and h, and percentage of time h is in r_a_vs.

To address this problem we present preliminary work on the development of a real-time and context-aware person following system able to handle unpredictable indoor environments using a heterogeneous robot team composed of an unmanned aerial vehicle and unmanned ground vehicle.

III. APPROACH

To solve the problem, we formulate a solution for goal location selection for both the ground and aerial robot. The approach as a whole involves motion prediction, candidate goal location generation, goal location scoring, and goal location selection. Person detection, localization, and tracking are outside the scope of this work.

In our formulation, time-horizon refers to the number of seconds a motion model of a person is predicted into the future. Position history refers to the number of seconds in the past over which position data is collected of a person and used to calculate a motion model. Motion prediction refers to the process of estimating the future motion or trajectory of an object or agent based on its current state and past observations.

The goal of motion prediction, in our case, is to anticipate the future behavior of a person in order to make informed decisions. This approach uses a simple second degree polynomial to model the person's motion, $P(t)$. Using position data over the position history, a polynomial fitting algorithm is used to calculate the coefficients. The input is the time in

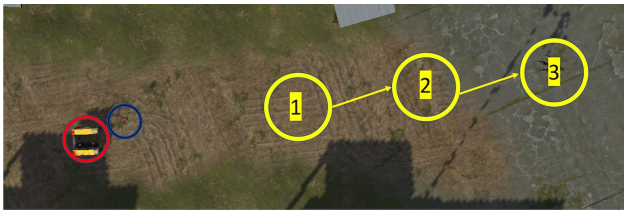

Fig. 2. Positions of the yellow circles represent the progression of the person's position over time t, a visualization of S, represented by the numbers.

seconds the position was observed and the output is a set of coordinates, (x, y). Points are then generated with this polynomial by getting coordinates between current time and the time-horizon in one second intervals. Each point is then used as the input for a normal Gaussian distribution with a σ of 0.9 generating 10 points each. These points, $F_{predicted}$, are meant to characterize the potential motion of the person over the time-horizon.

Candidate goal location generation in this instance uses a naïve implementation where goal locations are selected at three different distances in a radius around the person. A vector is calculated between the current position of the person and the person's predicted coordinate at a time-horizon: this is called the time-horizon vector.

(1) The closest set of candidate goal locations, one meter from the current location of the person, selects four goal locations forward, backward, left, and right of the time-horizon vector.

(2) The next set of candidate goal locations are calculated with a radius equal to the length of the time-horizon vector. Six goal locations, equally spaced out in a circle, are generated with this radius.

(3) The final set of candidate goal locations are selected similar to the second. Eight goal locations are selected, equally spaced out in a circle around the person's current position with a radius two times the distance of the time-horizon vector.

All candidate goal locations are oriented towards the mean of predicted motion points, $F_{predicted}$. Finally, all of the candidate goal locations in the set are checked to see if they can be navigated toward. If a goal location is in or behind obstacles prohibiting one of the robot's ability to navigate to its goal location, it is not included in that robot's set of candidate goal locations. If r_g can navigate to the goal location, it is included in C_g, and if r_a can navigate to it, it is included in C_a. If the goal location is in unexplored space, it is included by default.

There are two goal location scoring formulas, one for each robot. Eq. (4) addresses Eq. (2), where r_g scores goal locations based on the goal location's distance to the person's predicted location over some time-horizon. Similarly, Eq. (5) addresses Eq. (3), where r_a scores goal locations based on the goal location's view of the person's predicted motion over some time-horizon.

r_g assigns each candidate goal location's score the euclidean distance between the goal location and the value of the predictive motion polynomial at the time-horizon z:

$$c_{score} \rightarrow ||(c, avg(F_{predicted}(z)))||_2 \ \forall \ c \in C_g \tag{4}$$

r_a assigns each candidate goal location's score the number of predicted motion points, $F_{predicted}$, in r_a's view-shed:

$$c_{score} \rightarrow |F_{predicted} \in r_a\_vs| \ \forall \ c \in C_a \tag{5}$$

The final step in the approach is goal location selection. At each time step, r_g selects the goal location in C_g with the lowest score, and r_a select the goal location in C_a with the highest score.

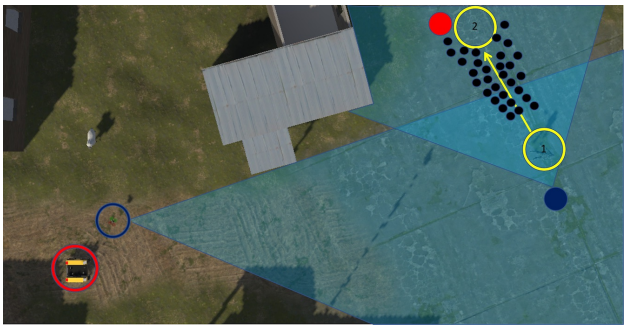

Fig. 3. Transparent light blue area visualizes UAVs current view-shed. Dark blue circle and area visualize goal location and goal view-shed that maximizes view of the person's projected motion. Black dots represent the possible future locations of h, $F_{predicted}$. Red circle visualizes a UGV goal location that minimizes UGV distance to person.

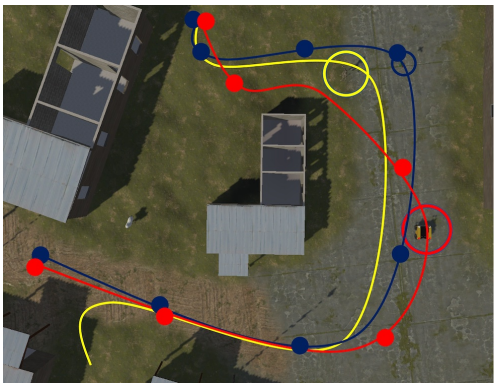

Fig. 4. This figure represents the current naïve approach. Lines are paths the agents took, filled circles are goal locations selected. Blue is the UAV, red is the UGV, and yellow is the person.

IV. PRELIMINARY OUTCOMES

Here we discuss preliminary progress on system implementation and expected outcomes, including motion prediction and a naïve goal location selection algorithm.

A. Experiment Setup

The experiment occurs in an outdoor urban environment rendered in a Unity ROS simulator. Figs. 1 display the agents and environment. The experiment space is located around a 10×13 meter building. Figs. 4 and 7 display the path, in yellow, of the person as it starts close to the starting location of both the ground and aerial robot. The starting positions of the UGV and UAV are marked on Figs. 4 and 7 by the red and blue circles farthest to the left.

B. Current Implementation

Our current implementation uses ROS as a system framework. Ground-truth simulation data is used for person detection, localization, and tracking. Motion prediction is currently implemented as described. The naïve goal location selection algorithm works by obtaining coordinates from the motion prediction polynomial with an input of a time-horizon, which is parameterized, and selecting that coordinate as the goal location. A robot is required to traverse

within two meters of their goal location before a new goal location is selected. This is a simplistic algorithm that results in the robots following the path of the person, as depicted in Fig. 4.

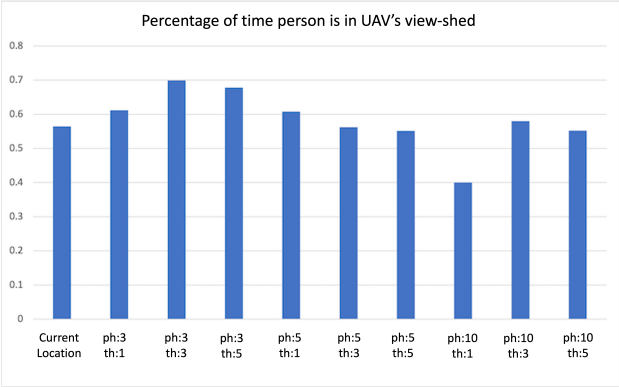

Fig. 5. UAV metric for current implementation. Y-axis is the percentage of the total time of the experiment where the person was in the UAV's view-shed.

C. Preliminary Results

Figs. 6 and 5 show metric results we have collected from a simulation experiment we ran with the current implementation. In these figures, *ph* refers to position history, the number of seconds of previous person position data is used to fit the person motion model, and *th* refers to time-horizon, the number of seconds in the future the motion model is predicting to generate the goal location. The goal of this experiment was to evaluate the current implementation with a variety of parameters. We evaluated the position history parameter with values of 3, 5, and 10 seconds. We also evaluated the time-horizon parameter with values of 0 (no motion prediction), 1, 3, and 5 seconds. 12 total configurations were evaluated, and their metric values averaged over 60 runs per configuration. A run is considered successful if the UGV reaches navigates within 1 meter of the person. Looking at Figs. 6 and 5, we see the results of evaluating the 12 different configurations do not show a clear trend from modifying each parameter. This leads us to postulate that the environmental features, such as characteristic dimension and tortuosity, may have an impact on the results. Configurations with a position history of 10 experienced a disproportionately higher amount of failed runs. Future work could involve testing this in different environments. Values are only computed with successful runs, but position history 10 configurations failed to maintain the person in either the UGV or UAV's view-shed enough to localize the person enough to navigate towards them, thus leading to failed runs.

D. Expected Results

As shown in Fig. 7, both the UGV and UAV should exhibit distinctly different behavior from each other as well as the person. We expect to see the UAV take wide turns and choose goal locations that provide an unobstructed view of the potential motion of the person. This is compared to the

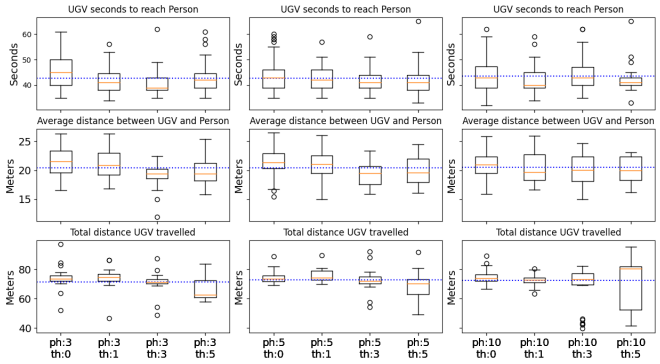

Fig. 6. UGV metrics for current implementation. Dotted blue line in each subplot represents the average value of the subplot.

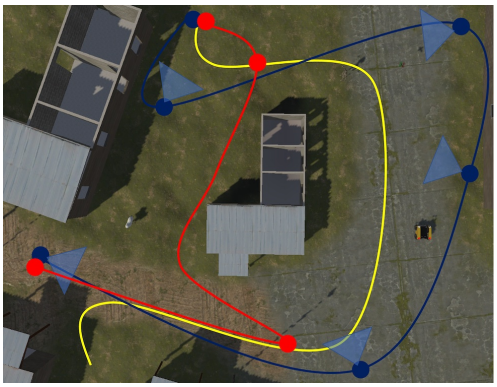

Fig. 7. Anticipated outcome of the proposed approach. The UAV (blue) takes wider path to ensure viewing of the person's movement. The UGV ends up taking a short cut because the motion prediction allowed for a more optimal location selection.

UGV which we expect to see take the shortest possible path. This is due to the UGV having knowledge of the persons position at anytime. In Fig. 7 the UGV starts by directly following the path of the person, but because the UAV has positioned itself to have a full view of the person as it rounds the corner, by sharing this information with the UGV the UGV is able to produce a more accurate motion prediction. This allows the UGV to choose a goal location that intercepts with the person's path.

V. FUTURE WORK

Our proposed approach and current implementation demonstrate a multi-robot system capable of following a non-cooperative agent successfully through a height-restricted environment. Future directions of this work involve implementing the proposed goal location selection and goal location scoring functions, first in simulation, to run extensive testing against comparable state of the art solutions, then real life experiments using a mobile ground robot and a UAV with a moving person as a validation test. Possible future work could also include expanding the robot team to contain multiple UAVs and UGVs, scaling the person tracking to multiple people or agents, and task allocation between robots.

REFERENCES

[1] Rubio, Francisco, Francisco Valero, and Carlos Llopis-Albert. "A review of mobile robots: Concepts, methods, theoretical framework, and applications." International Journal of Advanced Robotic Systems 16.2 (2019): 1729881419839596.

[2] Kruse, Thibault, et al. "Human-aware robot navigation: A survey." Robotics and Autonomous Systems 61.12 (2013): 1726-1743.

[3] Tarmizi, Aine Ilina, et al. "Latest trend in person following robot control algorithm: A review." Journal of Telecommunication, Electronic and Computer Engineering (JTEC) 9.3 (2017): 169-174.

[4] Xing, Guansheng, et al. "People-following system design for mobile robots using kinect sensor." 2013 25th Chinese Control and Decision Conference (CCDC). IEEE, 2013.

[5] Vasconcelos, Francisca, and Nuno Vasconcelos. "Person-following uavs." 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2016.

[6] Minaeian, Sara, Jian Liu, and Young-Jun Son. "Vision-based target detection and localization via a team of cooperative UAV and UGVs." IEEE Transactions on systems, man, and cybernetics: systems 46.7 (2015): 1005-1016.

[7] Gockley, Rachel, Jodi Forlizzi, and Reid Simmons. "Natural person-following behavior for social robots." Proceedings of the ACM/IEEE international conference on Human-robot interaction. 2007.

[8] B.J. Lee, J. Choi, C. Baek and B. -T. Zhang, "Robust Human Following by Deep Bayesian Trajectory Prediction for Home Service Robots," 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018, pp. 7189-7195, doi: 10.1109/ICRA.2018.8462969.

[9] Batista, Murillo Rehder, Douglas Guimarães Macharet, and Roseli Aparecida Francelin Romero. "Socially acceptable navigation of people with multi-robot teams." Journal of Intelligent & Robotic Systems 98 (2020): 481-510.

[10] Owen, Mark, et al. "Moving ground target tracking in urban terrain using air/ground vehicles." 2010 IEEE Globecom Workshops. IEEE, 2010.