

Air-Ground Robot Team Surveillance of Complex 3D Environments

Christopher Reardon and Jonathan Fink

Abstract—Real-world surveillance of complex 3D environments is an extremely challenging problem, especially in the presence of unknown dangers. In this work we create a novel ground and aerial autonomous robotic system to surveil a human-robot team’s surroundings for targets of interest, which could be e.g., disaster victims, infrastructure inspection, data to support research or public safety, or threats to the human team members’ safety. To represent this general case, our system identifies threats to human safety by surveilling the team’s surroundings, identifying threats, and notifying the human team member. We provide an interface that visualizes threat targets and allows the human operator to create and modify the surveillance plan. We note that this 3D surveillance task resembles the Art Gallery Problem (AGP) with a time-sensitive route planning aspect similar to the Traveling Salesman Problem (TSP), both of which are NP-hard. We incorporate a human operator into the decision making process of a surveillance system to address the viewpoint selection and route minimization problems, and to extract semantic information from the scene to increase search effectiveness. We construct a system for this collaborative, human-robot team surveillance task using a low-cost Unmanned Aerial Vehicle (UAV) and a more-capable Unmanned Ground Vehicle (UGV). We evaluate the resulting system with a large experiment set (120 trials) conducted in a real-world, 3D, cluttered, urban setting and examine the difference a scenario-specific plan makes to the detection of threat targets compared to a baseline algorithmically-generated plan.

I. INTRODUCTION

In scenarios where people operate in dangerous environments in real-world settings, those dangers, coupled with other essential tasks, may exceed a person’s ability to observe their surroundings. Furthermore, the task of monitoring the surroundings may be inherently dangerous. There is therefore a significant opportunity for robots working with humans in these scenarios to function as autonomous observation agents to enhance human safety. The observations could be in search of disaster victims [1], [2], bridge and infrastructure damage [3], or data to support scientific research or public safety [4], [5]. We choose to represent this general case by identifying threats to human operators in urban environments, and cast the problem as one of providing situational awareness to protect humans in dangerous scenarios.

Heterogeneous robot teams are challenging in that they require the use of diverse, resource-limited platforms to solve a common task. They also allow for systems that collaborate

to amplify the abilities of each individual agent. Collection of observations in a cluttered 3D urban scene may require a small unmanned aerial system. While the capabilities of these systems have advanced recently [6], [7], [8], they are not completely reliable and can be expensive. Furthermore, in many disaster and search scenarios, ground and aerial robots will often both be present and have the opportunity to work collaboratively. In dangerous environments where resources such as GPS may not be available and speed is critical, we believe that a simple, inexpensive, and expendable COTS UAV, when coupled with a more sophisticated ground robot, could have significant advantages over current approaches in terms of speed of operation and reliability. Using such an UAV in combination with a ground robot capable of autonomous mapping and navigation, we create a team capable of surveilling this challenging environment. By designing such a system and conducting extensive tests in a practical environment we are able to explore a number of important issues surrounding heterogeneous robot teams performing a 3D surveillance task.

In this work, we seek to examine the relationship between automated and scenario-specific solutions to a complicated, time-constrained task. First, we observe that solving the 3D surveillance task is NP-hard and requires complete knowledge of the environment geometry, and conducting autonomous 3D surveillance in an unknown or partially-unknown, cluttered, feature-rich environment is extremely challenging to solve in a real-world setting with time restrictions, e.g., identifying a threat before it impacts human safety. Second, we believe that information from the environment can be used to set search locations and priorities to provide for more efficient (high value, low time cost) observation, which would increase performance given time restrictions.

We believe that by building this system we can examine how a human operator addresses these issues and contributes to the decision making process of an automated surveillance system by 1) reasoning about the viewpoint selection and route minimization aspects of the surveillance task to provide a successful route for the robots’ patrol, and 2) understanding scene semantics and perceiving aspects of the environment to better assign surveillance priorities. Furthermore, we note that in human-safety systems, human involvement in the decision making process also enhances trust [9].

Using our autonomous surveillance team, we conduct extensive experiments in a 3D, cluttered, urban setting. We evaluate the performance of a scenario-specific surveillance plan alongside a naïve, algorithmically-generated plan. Through examining the performance of these approaches in

This research was supported in part by an appointment to the Postgraduate Research Participation Program at the U.S. Army Research Laboratory administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and USARL.

Christopher Reardon and Jonathan Fink with the U.S. Army Research Laboratory, Adelphi, MD 20783, USA {christopher.m.reardon3.civ@mail.mil, jonathan.r.fink3.civ@mail.mil}

a large-scale experimental study, we can observe and identify important trade-offs in this complex interaction.

The rest of the paper is organized as follows. We review work related to the task in Section II. In Section III we formally define our problem. Section IV details the methods we used to address the surveillance problem. Experimental setup and results are discussed in Section V. Finally, we conclude our paper in Section VI.

II. RELATED WORK

Automated solutions to the path-generation component of the 3D surveillance task require both a solution to the Art Gallery Problem (AGP), where a minimum number of observation points to cover the area is found, and the Traveling Salesman Problem (TSP), where an optimal tour of the points is calculated.

Approximate solutions exist to these NP-hard problems [10]. A method for generating trajectories for UAVs for complete coverage of the exterior of buildings in urban environments was presented in [11]. An iterative method for creating efficient paths that find viewpoints for full coverage of a 3D surveillance problem was shown in [3], where real-world validation was made simpler by inspecting known structures with mostly convex surfaces at a constant height. We note that our effort is to not only cover the exterior of objects but also the interior of highly concave surfaces (i.e., room interiors).

Human control of multi-robot teams has been studied in a variety of applications, including search and rescue (SAR) type tasks. In [12], human control of multiple ground and aerial robots in a search task was demonstrated, with the goal of providing integrated control for minimal human intervention. Degrees of multi-human shared control of multi-robot teams' navigation and victim localization for a search and rescue task were explored in [13]. Recently, [14] used hierarchical reinforcement learning to create a control architecture to allow a robot team to learn to explore and identify victims in an urban SAR task. In our work, we adapt the model of using a human operator to help solve the planning problem in a highly challenging environment.

Previous work has also identified the difficulties in understanding the behavior of targets operating in the environment [15] when solving the surveillance coverage problem. Semantic understanding of scenes is known to be challenging [16], [17], and a human can easily perceive contextual aspects of the environment that a robot may not. Further, human involvement in decision making has been shown to engender trust [9]. Considering this, for this work we have elected to use a human-in-the-loop approach to generate and modify paths. This design will allow us to examine algorithmically generated solutions to the 3D surveillance task coverage problem in combination with a human in future work.

III. PROBLEM

We consider an arbitrary three-dimensional environment represented as the open subset $\mathcal{E} \subset \mathbb{R}^3$. The environment is

populated with an unknown number of targets $\mathbf{Y} = \{\mathbf{y} \in \mathcal{E}\}$ representing threats that must be detected with a sensor positioned by our robotic system. We define the achievable sensor placements to be an open subset in the special Euclidean group $\mathcal{W} \subset \mathcal{E} \times SO(3)$. The boundary of \mathcal{E} , $\partial\mathcal{E}$, is induced by objects that limit sensor visibility. Furthermore, we assume a camera-like sensor with the following definition of visibility:

Definition 1: (*Visibility for a camera-like sensor*) A target $\mathbf{y} \in \mathcal{E}$ is visible from a sensor placement $\mathbf{w} \in \mathcal{W}$ if the following conditions are true:

- 1) Line of sight constraint: The line segment $S(\mathbf{w}, \mathbf{y})$ from the sensor to the target does not intersect $\partial\mathcal{E}$.
- 2) Range constraint: $d_{\min} \leq d(\mathbf{w}, \mathbf{y}) \leq d_{\max}$ where $d(\mathbf{w}, \mathbf{y})$ is the Euclidean distance between \mathbf{w} and \mathbf{y} .
- 3) Sensor field-of-view constraint: the segment $S(\mathbf{w}, \mathbf{y})$ must lie within the cone defined by the sensor's field-of-view.

We note that our problem formulation up to this point mirrors that of the *extended art-gallery problem* [18]. However, rather than require that we find a set of viewpoints to observe the entire environment, we would like to find a smaller set of viewpoints that maximize our likelihood of detecting the targets \mathbf{Y} within a maximum time limit τ . This better aligns our problem statement with that of surveillance where there is a timeliness objective that must be satisfied.

To address the timeliness issue, we start by recognizing that targets will not be located uniformly across the environment but will instead seek to avoid detection. Thus, we model the target locations y as a random variable with a probability distribution $\mathcal{D}_y(\partial\mathcal{E}, \mathcal{T})$. Specifically, the distribution \mathcal{D}_y is a function of features of the environment boundary $\partial\mathcal{E}$ that limit visibility, e.g., the targets are hiding, and *threat-cues*, \mathcal{T} , e.g., visual features by which the target "gives-away" its location.

We finish addressing the timeliness objective by noting that simply computing a list of viewpoints $\{\mathbf{w}_i \in \mathcal{W}\}$ that satisfy some probability of target detection based on Definition 1 is inadequate. Indeed, the time to traverse from one viewpoint to another $\delta t(\mathbf{w}_a, \mathbf{w}_b)$ and the total time to traverse all viewpoints $\sum \delta t(\mathbf{w}_i, \mathbf{w}_{i+1})$ are important when we consider that solutions must execute within the maximum time limit τ . Thus, we must also choose an ordering that visits $\{\mathbf{w}_i \in \mathcal{W}\}$ in an efficient manner, i.e., the traditional *traveling salesman problem*, and a feasible velocity profile for the search-route.

Finally, we can state our problem as:

Problem 1: For a given environment $\mathcal{E} \subset \mathbb{R}^3$, its boundary $\partial\mathcal{E}$, and threat-cues \mathcal{T} , find a sequence of viewpoints $\{\mathbf{w}_i \in \mathcal{W}\}$ and velocity-profile $\{\delta t(\mathbf{w}_i, \mathbf{w}_{i+1})\}$ such that $\sum \delta t(\mathbf{w}_i, \mathbf{w}_{i+1}) < \tau$. The probability that every target in \mathbf{Y} is visible according to Definition 1, per the threat distribution \mathcal{D}_y , should be maximized.

IV. APPROACH

We address the problem in an experimentally-validated manner using a multi-robot team. At a low-level, this team

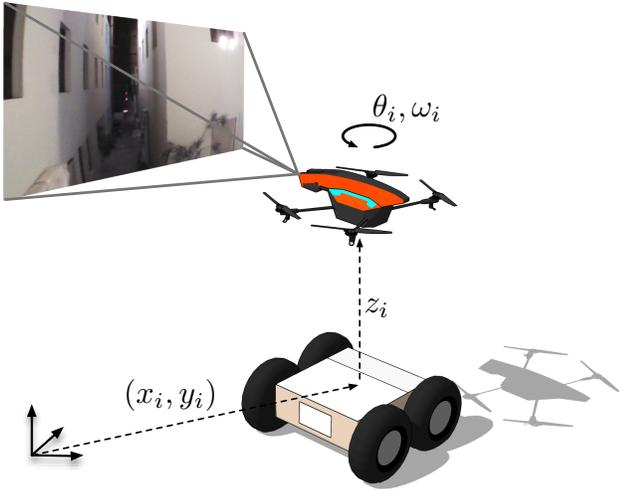


Fig. 1. A viewpoint in the route, $\mathbf{w}_i = (x_i, y_i, \theta_i)$ specifies the location of the ground robot (x_i, y_i) , the height of the aerial robot z_i , and the orientation of the aerial robot θ_i .

should be capable of creating a high-fidelity 3D map of the environment, localizing camera observations in the environment, and identifying or detecting targets from those observations. At a high level, the team must be capable of generating routes that maximize the probability of target visibility per Problem 1, achieving good performance in a highly complex 3D environment. This work presents a framework for experimental study of this challenging problem.

We propose that a human operator working with a collaborative air-ground team of robots is well-suited to the observation task described above in the complex three-dimensional environments that are relevant to our application. An autonomous ground robot offers the computational and sensing capabilities to quickly provide a spatial model of the environment, e.g., a colored point-cloud or texture-mapped 3D mesh. The autonomous aerial robot, while limited in computation and sensing, is able to provide vantage points inaccessible to the ground robot.

We propose a method for collaborative air-ground navigation that allows these robots to work together. In order to solve Problem 1, we then define an observation route for the collaborative air-ground team that consists of a sequence of viewpoints and velocity profile that the system must execute (Fig. 1). Finally, the human operator interacts with the team to create and modify the observation route according to higher-level semantic understanding he or she has about the task.

A. Autonomous robot team

Our collaborative air-ground team of robots consists of a sensing and computation-limited unmanned air vehicle (UAV) and an unmanned ground vehicle (UGV) capable of autonomous mapping and navigation (Fig. 2). These platforms work together in a challenging three-dimensional environment to build a model of the space, navigate to a



Fig. 2. The air-ground robot team operating in the 3D urban environment.

series of viewpoints, and collect geo-referenced images with the purpose to detect and localize threats to a human operator.

The UAV is the *Parrot AR.Drone 2.0*¹ controlled via the *Robotic Operating System (ROS)* [19] *ardrone_autonomy* package², with both forward and downward-facing cameras. It is capable of local flight-stabilization from onboard inertial measurements, optical flow on a downward facing camera, and SONAR-based height measurements. While this platform accepts body-frame velocity inputs, it is not able to detect obstacles in the environment or maintain a good estimate of its global pose. The platform is capable of onboard detection of fiducial markers and can efficiently stream video from the forward or downward-facing camera over its wireless communication link.

The UGV is a custom-built 4-wheel skid-steer ground robot developed at the Army Research Laboratory (ARL). It is equipped with processing payload containing a Quad-Core Intel i7 ICOM express board and a 256 GB solid-state drive (SSD). The UGV collects 3D point cloud data by nodding a Hokuyo UTM-30LX-EW LIDAR³ with a Dynamixel MX-28 servo⁴. The Hokuyo LIDAR has a 270° field of view, 30 m range, and 1 mm resolution. Accurate local state information is achieved using a MicroStrain 3DM-GX3-25 inertial measurement unit (IMU)⁵. We leverage the ROS infrastructure to implement custom algorithms for Simultaneous Localization and Mapping (SLAM) and waypoint-navigation [20].

B. Collaborative air-ground navigation

Because our UAV lacks reliable localization capabilities, a fiducial is mounted to the top of the UGV that can be detected by the UAV. By communicating with the ground

¹<http://ardrone2.parrot.com/>

²https://github.com/AutonomyLab/ardrone_autonomy

³<http://www.hokuyo-aut.jp/02sensor/07scanner/download/products/utm-30lx-ew/>

⁴<http://www.trossenrobotics.com/dynamixel-mx-28-robot-actuator.aspx>

⁵<http://www.microstrain.com/inertial/3DM-GX3-25>

robot, the aerial robot is able to update its pose in a global reference frame and compute controls to stay centered over the ground robot. Solutions to Problem 1 consist of viewpoints $\mathbf{w}_i = (x_i, y_i, z_i, \theta_i)$ that must be achieved by the aerial robot as depicted in 1. In order to navigate to a viewpoint \mathbf{w}_i , we employ the following strategy.

The UAV begins in a localized state over the UGV. The UGV plans a kinematically-feasible obstacle-free path to the (x_i, y_i) coordinates of the viewpoint \mathbf{w}_i , i.e., under the desired viewpoint. The UGV then drives its computed path with the UAV following by moving to maintain position above the UGV. If, at any time the UAV loses sight of the UGV, it communicates with the UGV to stop progress and wait for confirmation to proceed. The UAV begins searching for the UGV in the last observed direction and sends a resume signal after successfully detecting the UGV.

When the (x_i, y_i) coordinates of viewpoint \mathbf{w}_i are achieved by the ground robot, it pauses and communicates with the UAV that \mathbf{w}_i has been reached. The UAV then controls to the specified altitude and orientation (z_i, θ_i) for that viewpoint. When the UAV has captured an image at the desired viewpoint \mathbf{w}_i , it communicates with the UGV to navigate to the next viewpoint \mathbf{w}_{i+1} . During experiments, the operator can manually adjust the UAV’s flight to prevent a safety incident, or in the event that the UAV loses track of the UGV and cannot relocate it.

C. Target detection

Targets for this experiment consist of *AR.Drone* fiducials and are detected onboard the UAV and transmitted via the *AR.Drone* API. Detections are only recorded when the UAV is over the UGV for localization of the target. Detections are filtered and presented to the operator in Rviz ⁶ using text and shape markers. Clustering of target detections is performed at a 1m distance. To address noise, angle, and lighting variations, a moving average filter was applied to the depth (distance to target) component of the detection. To be considered a viable target location, at least 10 consistent readings are required; less than 10 readings are ignored. Confidence in the location is expressed to the operator in Rviz as $n/50$, where n is the number of consistent readings. Figure 3 shows example visualized target detection results.

D. Observation route

We define an observation route to consist of “viewpoints”, $(\{\mathbf{w}_i \in \mathcal{W}\})$, from Sec. III) and a velocity profile that affects the time it takes to traverse viewpoints $\delta t(\mathbf{w}_i, \mathbf{w}_{i+1})$. While the speed of the ground-air system when traversing the (x, y) component of a viewpoint is fixed to be the maximum reliable speed for stable control, the rotational velocity ω of the aerial platform when achieving the θ_i of a viewpoint is controllable and affects both the reliability of a detection and time to complete the observation route – there is a trade-off that must be considered when choosing this parameter. Thus we append rotational velocity ω to the observation route so that $\mathbf{w}_i = (x_i, y_i, z_i, \theta_i, \omega_i)$.



Fig. 3. Visualized detection results, rotated to view the building from the interior. Small yellow markers indicate target detection readings, large red markers indicate clusters (i.e., target location estimations).

1) *Route Duration*: As observation routes are generated, we calculate the time required to traverse from one viewpoint to the next, $\delta t(\mathbf{w}_i, \mathbf{w}_{i+1})$. While computing $\delta t(\mathbf{w}_i, \mathbf{w}_{i+1})$ for a deterministic system would be straight-forward and based solely on the distance traveled divided by the speed set points, this is clearly not the case for a real-world system. Uncertainty arises due to variability of the path the ground robot will take due to obstacles in the environment, noisy estimates of the relative pose of the ground and aerial robots and stabilization of the aerial robot. All of these factors combined make it challenging to predict traversal times with accuracy. Since a full uncertainty analysis and robust planning framework are beyond the scope of this work, we adopt an empirical approach to estimate the expected value of traversal cost under a decoupled motion model.

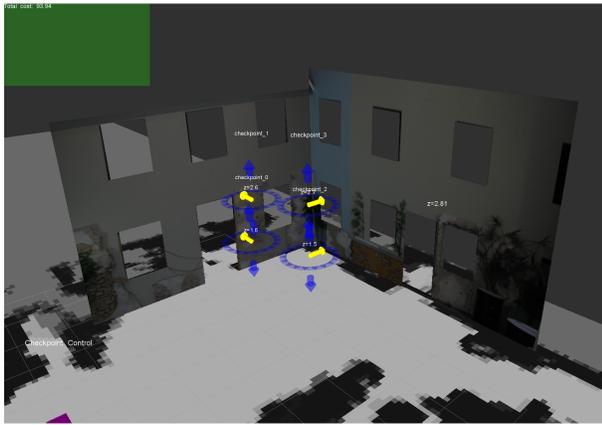
We collected experimental data of the system in several modes – UAV tracking UGV while controlling to the (x, y) component of a viewpoint; the UAV changing altitude to control the z component of a viewpoint; the UAV changing heading to control the θ component of a viewpoint. We found that the UAV rotation performance is dependent upon altitude so we adopt a linear model for expected time to control the heading of the aerial platform.

UAV rotation was measured at both first (1.5m) and second (2.8m) stories, at each speed $(\omega \in \{1, 2, 3\} \text{ rad/sec})$. Times are shown in Table I, where $(z_{max} - z_i)$ scales the rotational cost component by the difference in the maximum operating altitude and the current operating altitude, because rotating at lower heights is more time-consuming as the UAV must make more frequent adjustments to maintain its position.

TABLE I
DURATION MEASUREMENTS PER BEHAVIOR.

Behavior	Duration (s)
UAV Z traversal	$\ z_{i+1} - z_i\ \cdot 6.5$
UAV rotation	$8 \cdot \ \theta_{i+1} - \theta_i\ \cdot (z_{max} - z_i) / \omega_i$
UAV+UGV traversal	$2 \cdot \ (x_i, y_i) - (x_{i+1}, y_{i+1})\ $

⁶<http://wiki.ros.org/rviz>



(a) Control viewpoints



(b) Human-generated checkpoints

Fig. 4. Viewpoint generation environment using RVIZ. (a) Automatically generated uniform coverage viewpoints and (b) human-generated viewpoints.

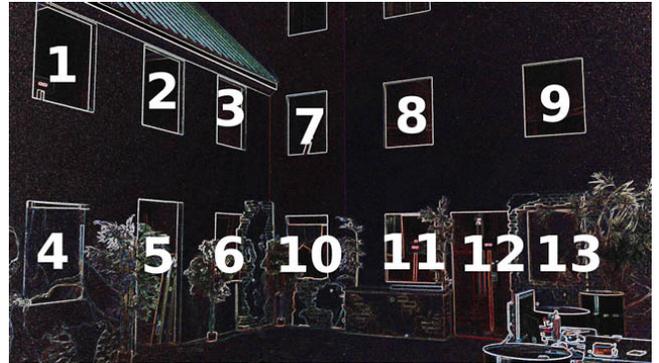
2) *Scenario-specific route generation*: The system allows the operator to create scenario-specific routes, where the operator 1) leverages understanding of the layout of the environment to generate sequences of viewpoints that maximize visual coverage while minimizing traversal and rotational costs and 2) interprets the visual cues to better prioritize viewing positions to ensure the viewing of higher probability locations given the constraints. Currently, this is done through an interface, but our testing methodology would extend to autonomously generated routes as well.

To easily generate and interact with this scenario-specific route, the operator is presented with a 3D representation of the environment as depicted in Fig. 4. While there do exist state-of-the-art technologies for autonomously generating photo-realistic 3D models of real-world environments, the implementation of such a system is beyond the scope of this work. The ground robot builds a three-dimensional point-cloud and traversal map of the environment; however, for the purposes of interpretability in testing our human-robot surveillance system, we present a 3D model to the operator that has been enhanced manually.

The interface allows the operator to quickly and easily add



(a) Test environment



(b) Test environment - labeled

Fig. 5. Urban testing environment. (a) depicts the complicated nature of the scene – targets can be hidden in any of the doors or windows. (b) Rooms are labeled to facilitate experiment design, setup, and execution.

and manipulate viewpoints including the position (x_i, y_i) , heading θ_i , and rotational velocity ω_i . The intention of this system is that is extensible to future work where the operator may modify automatically generated routes or change routes in an online setting to respond to new information about the environment.

V. RESULTS AND DISCUSSION

A. Environment

All experiments take place in an indoor test facility containing a multi-story urban street scene. The problem is made three-dimensional by utilizing the first and second stories of two perpendicularly oriented structures. The search team operates on the street outside the buildings’ two perpendicular external walls, and searches inside the buildings for targets by looking through 11 windows and 2 doorways. The visual and sensory texture of the environment is enriched with curtained windows, fake flora, textured wall placements, sandbags, steel barrels, and various objects placed inside rooms. The space behind each window or doorway is considered 1 room unit, giving 13 room units (Fig. 5); targets closest to a window or doorway are considered inside that corresponding “room.”

B. Control route generation

A naïve approach to the surveillance route was created as a baseline control against which to compare our algorithm.

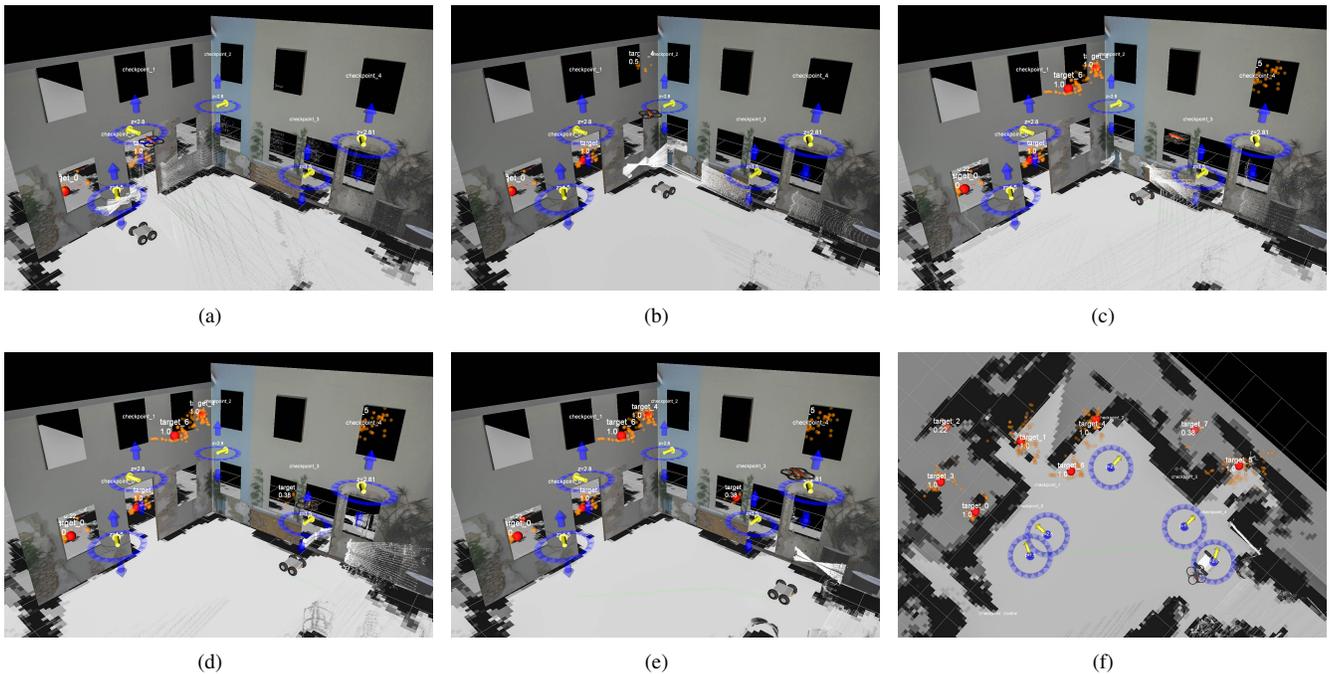


Fig. 6. Snapshots from a trial for threat distribution 1 and target placement 1. (a)–(e) depict the progress of the collaborative air-ground team as it makes observations in the environment. (f) shows an overhead view of the final target detections. Note that even though there are eight detected targets in this example, two of them are low confidence. Multiple localizations in the same room are treated as a positive detection.

By considering a “sphere-of-protection” model around the operator, a route is generated that uses high and low positions to scan the top and bottom stories of each wall from a central location. For each high and low position pair, the team assumes a position $\sim 1.5\text{m}$ away, perpendicular from the center line of the wall, and then scans ± 45 degrees to the left and right of that position, which allows it to view all windows on the wall from a central point.

C. Threat distribution

One crucial idea about real-world application of a surveillance task is that a human team member may be able to perceive *threat-cues* about the environment that would indicate areas of higher likelihood of a target being present as modeled in the \mathcal{T} term of our problem statement in Sec. III. Such cues would be highly task and situation dependent. For example, in a disaster response scenario, the human could perceive signs of life, written messages, disturbed debris, etc. In our threat-detection scenario, we use the term “threat” to indicate probability of target presence.

Since a general model of target probability would be quite complex for an arbitrary three-dimensional environment, we adopt a discretized model based on rooms in our test environment and introduce the variable r to indicate the relative level of threat for each room. We use two simple cues observable to the human operator to indicate a higher threat: a red light (symbolizing fire) indicating a high ($r = 10$) threat, and a white light (symbolizing smoke) indicating a medium ($r = 5$) threat; no cue indicates a normal ($r = 1$) threat. Because our purpose for this work was not to study a person’s ability to perceive threat, the meaning and

associated probability was made known to the human during route generation.

D. Experimental results

We conduct a set of experiments to explore and compare the performance of our collaborative air-ground system for a naïve surveillance route based on uniform models of threat and no understanding of complex environment geometry with a scenario-specific route where the operator can leverage a three-dimensional model of the environment and “threat-cues” placed in the environment.

Given the threat distribution \mathcal{T} , the operator generates a route. In our threat-based use case, we assume a short time limit τ for each search. Therefore, in order to make a valid comparison, the cost of the scenario-specific route must be less than or equal to the cost of the naïve control route. Note that time to generate routes are not included this calculation, as the actual cost of route generation would be highly specific to the application, and dependent on the interface and the operator. Because this work presents and examines the system, routes are generated by a single operator familiar with the interface. Future work to examine the performance of the system with multiple human participants is planned.

The experimental parameters are tuned such that exhaustively searching each window is not possible; instead, the human operator must decide where to direct the robot teammates for maximum effect.

A set of target placements \mathbf{Y} is generated after routes are created for a threat distribution. For $n = 5$ targets, target locations are generated via a weighted-random likelihood, using the threat distribution’s cue r values as weights, with

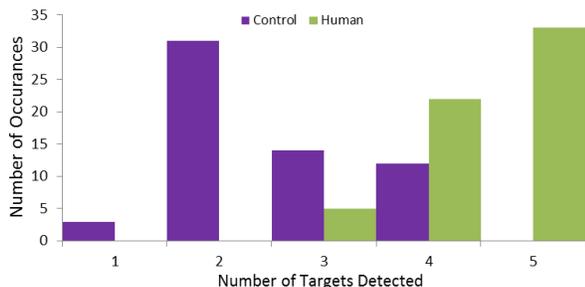


Fig. 8. Histogram of performance of control and scenario-specific routes.

no more than one target placed per room. In addition, to vary the difficulty of observing a target, for each i th target location, a difficulty $d_i \in \{easy, hard\}$ is generated with equal likelihood. Easy locations require target placement directly inside of a window or doorway opening, close to and centered on the opening. Targets in hard locations are placed to the side, set back (~ 5 m) from the opening, obscured (by curtains, plants or other obstacles), or a combination thereof. Because we are not attempting to solve a detection problem, we ensure that all targets are well-illuminated and facing the window/doorway opening.

Our experimental data consists of results from three threat distributions, with two target placement sets per threat distribution. In each target placement set, 5 targets were placed in 13 rooms. For each set of target placements, 10 runs using the baseline control route and 10 runs using the scenario-specific route were performed, for 120 runs total. Table II shows the configurations of runs, cues, probabilities, and placements. Figure 4 shows example control and scenario-specific routes for the first threat distribution. Figure 6 depicts the progress of a single experimental trial where the system successfully locates all of the targets in the environment.

Fig. 7 shows a complete breakdown of the results of all 120 runs. Our results show that, as expected, a scenario-specific route was able to outperform the baseline control in almost all aspects. There were, however, examples where the scenario-specific route missed easy targets. For example, in Threat Distribution 2 - Target Placement 1, in 2 of 10 runs, the scenario-specific route missed an easy target that the baseline route found. We note that these happened in situations where randomized target placements did not correspond to the higher-priority threat positions. In these situations, because of cost restrictions, the human made a decision to select a route that did not fully cover the less-likely target location. This highlights a key decision point that could be impacted by future work on intelligent route generation and human interaction.

A histogram of target detection counts is shown in Fig. 8. While performance of the system does exhibit variability due to stochasticity present in the control and sensing, there is a distinct shift in the performance distribution for scenario-specific routes. We believe that despite the high level of randomness in this complex environment, the human operator is able to implicitly choose patrol routes that are

more reliable. We also note that the average number of targets detected by the baseline control route was very near the number of “easy” targets in the target placement set, although not in all cases were the targets detected in easy placement locations. It is important to note that in only 55% of all runs did the scenario-specific route detect all targets. We observe that the trade-off between search priorities based on observed environmental cues and time cost can be critical in situations where human safety is at risk.

While this work is focused on creating a capability for a human operator to interact with and create surveillance routes, and not the extent to which we can maximize performance, we believe these results show great promise for exploration in future work for this issue and the many other issues surrounding this challenging and complex problem.

VI. CONCLUSION

In this paper, we create a system using ground and aerial robots to execute patrol routes to surveil a complex 3D environment, identify threats and visualize them for the operator. We examine the contributions a human operator can make to a surveillance task by making decisions regarding viewpoint selection and route cost minimization and interpretation of scene semantic information to determine priorities and create a successful route for the robot team members’ patrol. Through a large number of experiments, we demonstrate that, in the face of time constraints, while a route decided by a human is able to outperform a naïve algorithmically-generated route, important trade-offs exist that can impact human safety.

Furthermore, we consider the specific problem statement and framework presented in this paper to be a valuable cornerstone for future experiments exploring the interactions that take place when a human is teamed with an autonomous robotic system. Specifically, this work will allow us to further examine the influence of time-critical decision making on surveillance efficiency. We plan to further investigate this in the context of highly complex environments, where autonomous route planning and perception of environmental information are challenging, and examine what improvement a human operator can contribute to an algorithmically generated route.

REFERENCES

- [1] H. Kitano, S. Tadokoro, I. Noda, H. Matsubara, T. Takahashi, A. Shinjō, and S. Shimada, “Robocup rescue: Search and rescue in large-scale disasters as a domain for autonomous agents research,” in *Systems, Man, and Cybernetics, 1999. IEEE SMC’99 Conference Proceedings. 1999 IEEE International Conference on*, vol. 6, pp. 739–743, IEEE, 1999.
- [2] R. R. Murphy, S. Tadokoro, D. Nardi, A. Jacoff, P. Fiorini, H. Choset, and A. M. Erkmen, “Search and rescue robotics,” in *Springer Handbook of Robotics*, pp. 1151–1173, Springer, 2008.
- [3] A. Bircher, K. Alexis, M. Burri, P. Oettershagen, S. Omari, T. Mantel, and R. Siegwart, “Structural inspection path planning via iterative viewpoint resampling with application to aerial robotics,” in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pp. 6423–6430, IEEE, 2015.
- [4] I. Vasilescu, K. Kotay, D. Rus, M. Dunbabin, and P. Corke, “Data collection, storage, and retrieval with an underwater sensor network,” in *Proceedings of the 3rd international conference on Embedded networked sensor systems*, pp. 154–165, ACM, 2005.

TABLE II

EXPERIMENT CONFIGURATIONS. \mathcal{T} = THREAT DISTRIBUTION, \mathbf{Y} = TARGET PLACEMENT, d = DIFFICULTY, r = THREAT WEIGHT.

Room	\mathcal{T}_1		\mathbf{Y}		\mathcal{T}_2		\mathbf{Y}		\mathcal{T}_3		\mathbf{Y}	
	Cue	r	d_1	d_2	Cue	r	d_1	d_2	Cue	r	d_1	d_2
1		1			curtain	5	hard	easy		1		
2		1				1			curtain	5	hard	
3	curtain	5	easy	hard	curtain	5		hard		1		
4	sandbag	10	easy	hard	sandbag	10	easy	hard		1		
5		1	easy			1		hard	curtain	5		hard
6		1				1			sandbag	10	hard	hard
7	curtain	5		hard		1				1	easy	
8		1	hard			1				1		
9	sandbag	10		hard	sandbag	10	hard	hard		1		
10		1				1	easy		sandbag	10	easy	easy
11	curtain	5	hard	easy		1				1		hard
12		1			curtain	5				1		
13		1				1	hard		curtain	5	easy	hard



Fig. 7. Results for control and scenario-specific routes, by threat distribution and target placement, sorted by targets detected per target placement. Blue indicates easy target locations, red indicates difficult target locations.

- [5] O. Tekdas, V. Isler, J. H. Lim, and A. Terzis, "Using mobile robots to harvest data from sensor fields," *IEEE Wireless Communications*, vol. 16, no. 1, p. 22, 2009.
- [6] S. Shen, N. Michael, and V. Kumar, "Stochastic differential equation-based exploration algorithm for autonomous indoor 3d exploration with a micro-aerial vehicle," *The International Journal of Robotics Research*, vol. 31, no. 12, pp. 1431–1444, 2012.
- [7] S. Nuske, S. Choudhury, S. Jain, A. Chambers, L. Yoder, S. Scherer, L. Chamberlain, H. Cover, and S. Singh, "Autonomous exploration and motion planning for an unmanned aerial vehicle navigating rivers," *Journal of Field Robotics*, 2015.
- [8] M. Faessler, F. Fontana, C. Forster, E. Mueggler, M. Pizzoli, and D. Scaramuzza, "Autonomous, vision-based flight and live dense 3d mapping with a quadrotor micro aerial vehicle," *Journal of Field Robotics*, 2015.
- [9] P. a. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, and R. Parasuraman, "A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 53, pp. 517–527, Sept. 2011.
- [10] E. Galceran and M. Carreras, "A survey on coverage path planning for robotics," *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1258–1276, 2013.
- [11] P. Cheng, J. Keller, and V. Kumar, "Time-optimal uav trajectory planning for 3d urban structure coverage," in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pp. 2750–2757, IEEE, 2008.
- [12] M. A. Hsieh, A. Cowley, J. F. Keller, L. Chaimowicz, B. Grocholsky, V. Kumar, C. J. Taylor, Y. Endo, R. C. Arkin, B. Jung, *et al.*, "Adaptive teams of autonomous aerial and ground robots for situational awareness," *Journal of Field Robotics*, vol. 24, no. 11-12, pp. 991–1014, 2007.
- [13] M. Lewis, H. Wang, S.-Y. Chien, P. Scerri, P. Velagapudi, K. Sycara, and B. Kane, "Teams organization and performance in multi-human/multi-robot teams," in *Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on*, pp. 1617–1623, IEEE, 2010.
- [14] Y. Liu and G. Nejat, "Multirobot cooperative learning for semiautonomous control in urban search and rescue applications," *Journal of Field Robotics*, 2015.
- [15] B. A. Johnson, V. An, and H. Qi, "Unsupervised 3d scene understanding and prediction to enable adaptable approximate solutions to the art gallery problem and watchman route problem," in *SPIE Defense+ Security*, pp. 94541L–94541L, International Society for Optics and Photonics, 2015.
- [16] J. Vogel and B. Schiele, "Semantic modeling of natural scenes for content-based image retrieval," *International Journal of Computer Vision*, vol. 72, no. 2, pp. 133–157, 2007.
- [17] M. R. Greene and A. Oliva, "Recognition of natural scenes from global properties: Seeing the forest without representing the trees," *Cognitive psychology*, vol. 58, no. 2, pp. 137–176, 2009.
- [18] H. González-Baños, "A randomized art-gallery algorithm for sensor placement," in *Proceedings of the seventeenth annual symposium on Computational geometry*, pp. 232–240, ACM, 2001.
- [19] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "Ros: an open-source robot operating system," in *ICRA Workshop on Open Source Software*, 2009.
- [20] J. Gregory, J. Fink, E. Stump, J. Twigg, J. Rogers, D. Baran, N. Fung, and S. Young, "Experiments toward the application of multi-robot systems to disaster-relief scenarios," Tech. Rep. ARL-TR-7423, ARMY RESEARCH LAB ADELPHI MD COMPUTATIONAL AND INFORMATION SCIENCES DIRECTORATE, 2015.