INTERNATIONAL JOURNAL OF INTELLIGENT CONTROL AND SYSTEMS
VOL.16,NO.2,JUNE 2011,52-66

Assessing Unmanned Ground Vehicle Tactical Behaviors Performance

Marshal A. CHILDERS, Barry A. BODT, Richard CAMDEN

Abstract- **The U.S. Army Research Laboratory Robotics Collaborative Technology Alliance (RCTA) has made significant advances in perception and planning that enabled true unmanned ground vehicle autonomous navigation. Perception advances included laser radar sensing and near- and mid-range perception algorithms. Planning aspects included global/local route planning, shared map data, and the ability to use time as a planning factor. True autonomous navigation is characterized as a capability for both planning the vehicle course and controlling the position of the vehicle on the route. This opened the door to a myriad of operational-like capabilities, which we call tactical behaviors. This paper provides an overview of the assessment process in the RCTA and illustrates a successful feedback loop with developers supported by periodic field assessment of developing technologies.**

***Index Terms*—Autonomous vehicles, algorithms, measurement, performance, field experimentation, tactical behaviors, unmanned ground vehicle, UGV, XUV, path planning, bi-directional information flow**

1. INTRODUCTION

At the conclusion of the U.S. Army Research Laboratory (ARL) Demo III program [1], unmanned ground vehicle (UGV) capability could be described as semi-autonomous mobility, moving from A to B following closely spaced waypoints, following the planned path within a specified distance (~50 m), and calling for operator assistance when needed. The ARL Robotics Collaborative Technology Alliance (RCTA) made significant advances in perception and planning that enabled true UGV autonomous navigation, which is characterized as a capability for both planning the vehicle course and controlling the position of the vehicle on the route [1,2]. Perception advances included laser radar (LADAR) sensing and near- and mid-range perception algorithms. Navigation planning aspects included global/local route planning, shared map data, and the ability to use time as a planning factor. This opened the door to a myriad of operational-like capabilities, which we call tactical behaviors. ARL periodically conducted field experiments throughout the RCTA to assess the progress and readiness of RCTA-developed technologies.

This paper provides an overview of the assessment process in this program and illustrates a successful feedback loop with developers supported by a periodic field assessment of developing technologies. We begin by discussing experimental philosophy in the program; that is, we provide a statistically grounded, responsive, and opportunistic assessment intended to incrementally move the program toward long-range technology goals. We follow by laying out the broad framework for the assessments performed, which includes envisioned challenges for UGVs in an operational setting, enabling technologies, and experimental methods. A more detailed discussion of three specific assessments is then given to illustrate how the experiment was conducted and what was learned. We conclude with an overview of the most important advances in the robotic technology and address some of the measurement challenges that persist in field experimentation of UGVs.

2. EXPERIMENTAL PHILOSOPHY

Performance assessment of robotic systems is a developing area. Since we have been engaged for several years in a development/assessment cycle, it is appropriate to share our approach to the task. To suggest this as *our philosophy* is perhaps an overreach; the ideas are not new, and maybe a collection of tenets is more to the point than a philosophy. Still, we approached the work in the manner presented here for several reasons, which were influenced broadly by our military setting, our statistical training, and special challenges of autonomous ground systems.

The RCTA was a sequel to the Demo III program in UGV research and development and represented a major portion of the Army investment in robotics technology. It was incumbent on the RCTA to be responsive to leadership and to periodically offer convincing arguments on the return on investment. Initially, the focus was on basic capabilities measured by, for example, the distance traversed autonomously and the number of operator interventions required; however, the focus soon shifted to notional scenarios consistent with tactical operations of an envisioned robotic system in complex urban and cross-country terrain—e.g., reconnaissance of a clearing beyond a forested region. Exposure to such situations addressed what the technology could or might do, but a convincing argument also had to be made on progress toward those goals.

To make defendable arguments, we turned to formal experimentation rather than demonstrations. The purposes of the experiments were to both benchmark progress and encourage innovation. Initially, progress was judged according to a technology readiness level schedule, especially with regard to testing in relevant environments [3]. We encouraged innovation by producing empirical evidence that would challenge developers—in a manner consistent with a feedback loop outlined in a classic text in statistics [4]—to revise their understanding and make advances in technology development. Annual reviews of all RCTA activity, including assessments, and frequent informal exchanges between assessment and development arms of the RCTA ensured the dissemination of assessment results. Two recent books concerning military system development [5,6] influenced our thinking by suggesting that operational relevance should be included in developmental testing, by emphasizing the importance of early discovery of failure modes in the acquisition cycle, by identifying rigorous statistical investigation as essential, and by stressing the importance of valid, reliable, and credible metrics. With regard to metrics, “Unless good decisions are made in this arena, no amount of analysis or interpretation will generate a successful experiment.” [6] However, that is not to say that a suitable metric has to be complex. In some instances, conformance with expected behavior as declared by subject matter experts or simulation prior to making runs was what was measured.

Autonomous ground systems offer special challenges for assessment. For example, we took the position that it was too ambitious to develop a metric to assess the degree of completeness or correctness of what the system perceived around it in a complex environment. This would have been even more difficult later in the program when various levels of a priori map data were fused with locally perceived information. Further, aside from how the robot planned in light of what it perceived, it is unclear what would constitute merit for such a metric. Rather, an expected behavior for the entire system was the focus when planning experiments. What the robot saw at the time of an unexpected or undesirable behavior was pursued after the fact through graphic displays of the perceived scene and other developer software tools synchronized with the time of the behavior.

Consistent with these principles in DoD system development, our general strategy was to (1) envision plausible mission profiles where robotic ground system technologies would be applicable. The Future Combat System at the time helped drive this by identifying roles for unmanned systems. The Unit of Action Maneuver Battle Lab at Ft. Knox helped define roles by advancing a reconnaissance mission scenario to illustrate how the technology might be used. (2) We identified first-step capabilities toward accomplishing a notional mission, which could be supported by some combination of perception, planning, and human-robot interface. (3) We assessed the capabilities using sound experimental practice that included a thorough consideration of the design space for independent factors in the context of the mission, nontrivial runs to explore that design space, and sensible response measures, e.g., speed and the number of emergency stops (e-stops). (4) Opportunistically, in consideration of technology advances and in partnership with developers, we would raise the bar on challenges in the next round of assessment, driving toward technology consistent with a fielded system that could accomplish the mission profiles in (1) and sometimes modifying aspects of those missions to make them more ambitious in light of technology advances achieved.

Others have considered guidelines for performance evaluation specific to unmanned systems [7-9]. Although we have not followed them overtly, there is overlap. Following initial RCTA experiments in 2002 and 2003, it became apparent to ARL that a solid framework for specifying and testing autonomous system requirements did not exist. The ARL and the National Institute of Standards and Technology funded an effort in this regard, which became known as ALFUS (Autonomy Levels for Unmanned Systems) [8]. ALFUS formulates, through a consensus-based approach, a logical framework for characterizing the unmanned system autonomy, covering issues of levels of autonomy (later renamed Contextual Autonomous Capability [CAC]), mission complexity, and environmental complexity. Using the ALFUS framework, we could parse the various aspects of tactical behaviors enabled through the research and development of the RCTA into the three aspects of UMS autonomy found in ALFUS. We discuss this more precisely later.

The Defense Advanced Research Projects Agency DARPA Grand Challenge competitions [10] offer another model for assessment and encouraging innovation. In the urban challenge, the time to successfully accomplish a number of urban-specific behaviors was the principal measure and was adjusted downward with penalties for behaviors judged inadequate according to the rules of the challenge. The level of a priori data assumed in the DARPA challenge competitions was well beyond what we assumed in many of our experiments. Indeed, whereas DARPA had the luxury of knowing where the robot would likely travel over the course, there was seldom a unique solution to a cross-country challenge presented in our experiments. We had to be willing to have operational definitions covering a wide range of “solutions” so that we could consistently score the success or failure of a run.

Ultimately, this philosophy for experimenting in the RCTA was executed in field experiments, which can stress even the best design and analysis strategies.

3. FIELD INVESTIGATIONS FRAMEWORK

There is no recipe guaranteeing the success of field investigations in their role to advance the understanding of robotics. Field investigations are inherently difficult and sometimes messy, but venturing out from the safety of the lab is essential for system development (e.g., in the discovery of operational failure modes). The success of field investigations in that role is largely determined by the difficulty of the field challenge itself and the maturity of the technologies applied to meet that challenge, but it is also greatly influenced by the experimental philosophy of the program. Is it an experiment to stress the technology in application or a demonstration to showcase what is, for the most part, known? In the RCTA program the reward to developers of understanding the limits of performance of their particular technology was always deemed worth the risk of some failure. It is within this framework of challenge, technology, and experimental philosophy that the success of RCTA field experimentation should be viewed.

3.1 Tactical Behaviors Challenge

Over the length of the RCTA, as technology advanced, the definition of tactical behaviors expanded until it included the very ambitious capability of autonomous navigation on roads, trails, and cross-country and urban terrain while

- detecting, classifying, and avoiding static obstacles,
- detecting, classifying and avoiding people in various sizes, postures, and movements,
- detecting and avoiding moving vehicles,
- following the rules of the road,
- achieving a designated goal point in spite of obstacles or blocks on the original planned route, and
- minimizing the need for operator assistance.

Thus, the basic challenge was to accomplish simple mobility as a driven vehicle might accomplish over varied terrain.

Beyond simple mobility, however, using a notional reconnaissance mission as the context, autonomous navigation must be accomplished with a purpose, such as the following operational-like behaviors (see Fig. 1):

- avoid being observed
- observe an area of interest
- use terrain features
- understand military control measures

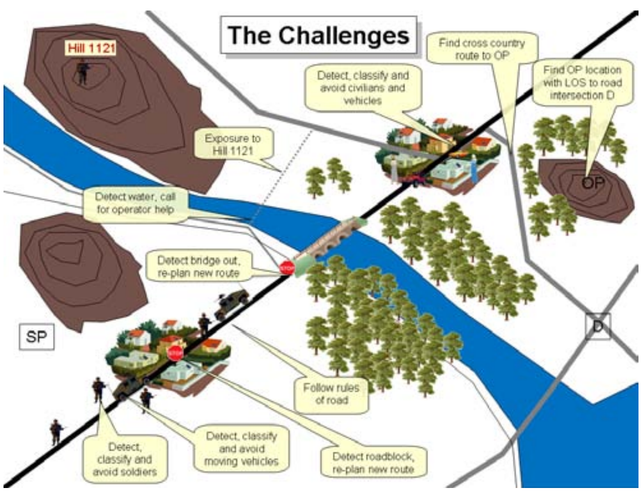

Fig. 1. Tactical behavior challenges (SP [start point], OP [observation point], LOS [line of sight])

Each challenge was experimentally investigated at some point in the program. The approach taken throughout the RCTA was to integrate developing technology and software onto the experimental unmanned vehicle (XUV) and then stress the *system* by exposing it to these progressively difficult operational-like challenges while collecting meaningful measures of performance [11]. Some highlights illustrating the process follow. More details for specific experiments are discussed in Section 4.

In early work, the focus was for the XUV to autonomously arrive at an assigned observation point (OP), with soldier operators determining the initial global route on the operator control unit (OCU), monitoring XUV progress, and remotely assisting the robot when it got stuck, relying only on the OCU information and teleoperation cameras. Success in autonomous mobility paved the way for consideration of a mission context [2]. Operators would then initiate, receive, and interpret a reconnaissance, surveillance, and target acquisition (RSTA) scan from the one *or two* robots under their direction. Operators were stressed with varying workloads and observations, and XUV performance metrics were collected, the results of which contributed to changes in the OCU/RSTA control to enable more effective and efficient soldier control [12].

Follow-on experiments examined the XUV capability to find and maneuver to a specified OP with a threshold visibility to a named area of interest (NAI) as determined by the local sensors. When at the OP (designated by the operator), the XUV visibility to the NAI was frequently blocked by local vegetation. This led to the development of the OP Finder—an algorithm that used locally sensed data to enable precise repositioning of the robot to improve visibility.

The autonomy challenge for a RSTA mission was increased further to include situations where the robot needing to travel 1 to 2 km would discover nearly insurmountable route challenges on the way to an NAI. In practice, these would occur for a variety of reasons (e.g., inaccurate or incomplete global map data used to set the route or downed trees and other local obstacles consistent with cross-country terrain). A representative and common cause of failure for autonomous navigation was the vegetated cul-de-sac, a naturally occurring area in the course with impenetrable vegetation on three sides and a depth exceeding the range of local perception. The XUV would enter a cul-de-sac and, finding no direct way toward the goal, would try to navigate out by backing up a limited distance along its path to gain a different perspective on potential local routes. This approach was only successful if the cul-de-sac was shallow enough that the new perspective would show a path out that would be within the deviation tolerance of the global route; otherwise, the operator was called for help. Detecting and solving a cul-de-sac or roadblock became a major metric of autonomous navigation success and was a driving force in the improvements of planners and in how map data was acquired and managed.

A further capability shortfall was the inability to navigate on roads following the normal rules of the road and avoiding moving vehicles and pedestrians. The Demo III Planner was incapable of handling moving obstacles because it could not reason with time. Successfully detecting and avoiding dynamic obstacles (vehicles and pedestrians) became another significant metric.

3.2 Enabling Technologies

In this section we describe the critical technologies developed to enable a UGV to overcome the operational challenges.

3.2.1. Platform and sensors

The platform used throughout the experiments was the XUV (Fig. 2). Sensors mounted on the XUV over the course of the RCTA program included proprietary General Dynamics Gen III and Gen IV LADARs; front, side, and rear SICK LADARs; a monocular teleoperation camera; stereo cameras for obstacle detection and classification; and an RSTA camera and rangefinder suite. The OCU (Fig. 3.) was mounted in a high-mobility multipurpose wheeled vehicle (HMMWV), which served as the control vehicle. The OCU provided a map for situational awareness and interfaces for planning and monitoring, teleoperation, and control of RSTA sensors [11].

Fig. 2. Experimental Unmanned Vehicle (XUV)

Fig. 3. Operator Control Unit (OCU)

3.2.2. Navigation (planning) algorithms

A High-Maneuverability Planner (HMP), capable of planning maneuvers in tight spaces (e.g., around an OP), and an OP Finder, which used local perception to find a location with maximum visibility to the area of interest, were developed after the early OP experiments.

The most significant solution consisted of bidirectional information flow (BIP) and a hybrid global/local planner [13-15]. Near-range data sensed by the XUV was used locally while simultaneously updating the global map. When called upon, the global planner would incorporate this new information, often leading to a feasible route opportunity to which the XUV could follow to the goal. Prior to this improvement, the XUV would not have been able to return to find that opportunity because it was limited by a prescribed backup procedure and by a specified maximum deviation from the initial route plan. To this latter concern, the XUV was untethered from the initial global route through a new planner that balanced costs from the local route (within sensor view) with global costs from the initial route and allowed more flexibility for the robot to advance toward the goal. Feature data was also leveraged to support the choice of a stealthy route relative to a predetermined threat location. RSTA operations were refined to include new strategies in local repositioning of the XUV at the OP to overcome foreground vegetation occluding the view to the area of interest. Building on this multilevel planning approach, the RCTA developed a Dynamic Replanner (DR) to provide several alternative routes for planner consideration each second, based on a continuous updating of the global map with locally sensed information. HMP enabled the XUV to transition to new plans, which often required tight initial turns [16].

Later work integrated knowledge of rules of the road into a separate Traffic Planner (TR), which enabled planning in time, thus placing the moving object precisely in the local map and predicting future locations in time. This enabled the XUV to maneuver correctly in the vicinity of dynamic objects.

Finally, the integration of the DR for off-road use and the TR for on-road use and the ability to transition between the two via the HMP provided seamless autonomous navigation in any environment during the RCTA Capstone Experiment [11].

3.3 Experimental Methods

Preparation for field assessments included consideration of test conditions and environment, experimental design, and analysis of response measures.

3.3.1. Test conditions and environment

In each experiment the principal variable conditions were chosen to reflect the technology being assessed, the capability that it provided, and the factors likely to influence XUV performance. For example, in the OP Finder experiments, the principal conditions were OP Finder algorithm (2 levels), % NAI to be observed (2 levels), and the range of the OP Planner (2 levels). In the most recent tactical behavior experiments, the principal conditions were planning algorithms (4 levels), a priori data (3 levels), and the number and location of blocked paths (3 levels).

All experiments were conducted at Fort Indiantown Gap, PA, Tactical Vehicle Maneuver Area-Bravo (TVMA-B). This maneuver area provided a limited but rich experimental environment, including roads, trails, trees, dense and light vegetation, and streams typical of a tactical maneuver area. This was considered an operationally relevant environment.

Many factors were uncontrolled by their nature. Most experiments carried over several days, hence weather varied. The time of year brought its own variations of snow, rain, and foliage. Between experiments, other military users of the terrain altered the environment by forging new trails and creating large ruts. During any one experiment, no single course was usually suitable to provide the terrain challenges necessary to examine all the technologies under scrutiny; thus, several courses were generally laid out to provide the venue to evaluate each experimental condition. Each course was usually given a unique name for identification, such as the Gold course or the Black course.

3.3.2. Experimental design

The experimental conditions were usually arranged in a factorial structure with replications to be run according to a randomized schedule to the extent practicable. A protocol for the test was established, and rules for how to handle and score anticipated situations in the field were established. The factorial integrity of the design was usually maintained; however, there was occasionally a deliberate modification based on existing knowledge and knowledge gained during experimentation so that range time was used well. Time and resources were always constrained in the field experiment. To maximize the collection of useful information and to save time, occasionally some replications were not completed when it was clear that additional runs would result in identical results.

For example, in the final (Capstone) experiment of the RCTA, conducted in fall 2009 on the Gold course, the progression of challenges could be considered sequential, in that the four-blockage condition was not attempted until the easier no-blockage and two-blockage challenges had been exhausted. In addition, the Autonomous Mobility (AM) Planner (i.e., without BIP and a global/local planner) was not carried forward to the four-block condition because its inability to solve the deep cul-de-sac problem created by this condition had been established in a previous experiment [16]. Taking the opportunity, we substituted the Mixed Planner (combination of the DR and Traffic Planners) that was concurrently performing well on the Red course for AM under the four-block challenge in the principal factorial design on the Gold course. When range time was made available, we conducted additional runs with DR and Mixed Planners under the four-block Gold course challenge [11].

One could view experimental design planning in terms of ALFUS. On the mission complexity axis, we varied the nature of the mission (movement to point or RSTA operations), distance traveled, number of waypoints provided, number of targets presented, and mission package workload. On the environmental complexity axis, we varied the terrain type, vegetation type, on/off roads, frequency of cul-de-sacs encountered, level of a priori data available, day/night operations, nature of obstacles (static or moving), and human obstacles (static, moving, various postures and groupings). On the level of autonomy (or CAC), we varied the degree of operator intervention allowed (teleoperation or none), planner capability (4 versions), type of OCU, UGV/operator ratio, and operator workload.

3.3.3. Analysis plan

An analysis plan accompanied the experimental design. When possible, if a suitable response measure could be determined, we conducted a standard analysis of variance. For example, in pedestrian detection studies, a variance-stabilizing transformation on the proportion of detections allowed an analysis of variance on the response. In other studies, time to complete a mission or time to teleop a robot out of trouble would support analysis of variance or regression. Lacking solid quantitative measures, we sometimes assessed performance in terms of the ratio of challenges overcome to the challenges presented. This was a key measure in the Capstone experiment. Despite planned analyses, sometimes the important findings were gained through exploratory data analysis of individual runs or segments of a specific run in a post-hoc analysis. Occasionally, behaviors were merely recorded and evaluated by developers and evaluators in a subjective manner; however those individuals were subject matter experts in the technology and/or military operations.

Throughout the program, there were struggles with finding measures of performance. Establishing meaningful response measures in an operational experiment is inherently difficult because it is not possible to control extraneous, potentially lurking variables not directly under study. Further, the experimental unit whose performance is to be measured is the system of integrated technologies as opposed to an isolated subsystem. Terrain, which clearly drives perception and planning, can only be selected through wholesale adoption of a course; another course with similar macro terrain features may yield a different outcome because of nuances at the micro level. The presence of dust, rain, and other environmental conditions cannot be controlled, nor can the malfunction of hardware not related to the technologies under study. When a system success occurs, all subsystems are credited with the achievement, but when a failure occurs, which is to blame? Toward the end of an acquisition cycle, one might not care which is to blame (failure or success will suffice), but upstream in the applied research phase it is important to both gauge system promise and provide subsystem feedback. Doing so often required dissection of performance within a run.

4. THREE FIELD ASSESSMENTS

We present three examples of RCTA experimentation in practice. The first is a two-part study addressing RSTA missions for the first time and span of control for operators over multiple assets. The second explores OPfinding. The third summarizes the Capstone experiment in tactical behaviors. A complete accounting of each experiment will not be given; rather, certain aspects will be highlighted to illustrate the previous discussion, with most emphasis on the most recent investigations.

4.1 RSTA and Span of Control 2005-2006

By May 2005, robot control had moved beyond a Unix window display—common at the beginning of the program—to a new OCU interface with improved features. Among those features were the capability of controlling more than one robotic asset and the capability for operators to issue spot reports of entities on the field by fixing a position on the OCU display corresponding to the entity location. The latter feature was to be used in conjunction with a live RSTA scan. That RSTA technology had not been integrated on the XUV, but a way to simulate the scan was developed so that both new features could be examined in an assessment.

An experiment was formulated to focus on the new features while, in addition, examining operator workload in the mission task of issuing a spot report on an RSTA scan. The experiment used a complete factorial arrangement of three factors at two levels in two replications. Two operators alternated in controlling the robot, and the OCU vehicle either remained in a fixed location or was *on the move*. The OCU motion was a potential stressor in operator workload. An additional workload stressor, the third factor in the design, was the task of acquiring a live RSTA feed from a small fixed-wing unmanned aerial vehicle (UAV). A total of sixteen runs were conducted.

As the XUV progressed over the route chosen by the operator, it would come in proximity to six preset locations on the course, each causing a simulated scan to appear to the operator as if it were a live RSTA scan (Fig 4.). These images were randomly presented. The images consisted of a panoramic stitched image of actual photos of the scene, which include military vehicles that were arranged so that they would appear in the scene in a left, right, or center position or the scene would be clear. At most, three vehicles were in one scene. Automated Target Recognition (ATR) was spoofed by using a photo editor to designate a suspected target with a box. Operators were told that the ATR was not 100% accurate; interpretation of the scene was not to rely on the ATR.

Fig. 4. Simulated RSTA image with three designated targets

The results of this study can be broken out according to human factors findings and technology evaluation. We only address the latter, even though the design of the study supported analysis of variance on an index used to evaluate workload. The response errors on a scan could be, for each of three locations on the scan, a missed target, a misidentified target, or a false alarm. In 96 scans and 288 image locations, only 30 errors were made. Over a third of those were missed M113 vehicles in hard-to-see locations in two scenes. With regard to spot report accuracy, there was no difference between fixed and on-the-move, but one operator tended to be more accurate. The impact of the fixed-wing UAV could not be evaluated because that aspect of the test was abandoned after three runs. Image stability impacted by wind made scene interpretation from the UAV too difficult, and spotty communication links prevented a continuous video feed.

In April 2006, a follow-on experiment was conducted. The focus of this experiment was span-of-control, but the RSTA mission was retained as the operator task. Simulated RSTA scans were expanded in the number of scan locations on the course and the library from which scenes could be randomly drawn. The newly integrated live RSTA on the XUV was also demonstrated at the end of some runs to view an NAI approximately 1 km away. The OCU interface had also been migrated to a Tablet PC, so there was a human factors question to explore on the difference between that and the full-size display. This experiment marked the first time that soldiers interacted with the OCU, with two participating from the Mounted Maneuver Battle Lab.

The experiment primarily examined operator control issues and repeated a portion of a course from an autonomous mobility test, Technology Readiness Level 6 (TRL-6), run four years earlier. This experiment was prior to the development of more advanced planners, so the

Demo III paradigm of backing up three times before calling for operator intervention was still in effect. The metrics shown were those used in the TRL-6 experiment. All of the major metrics applied in 2006 (average speed, kilometers between backups, and kilometers between operator interventions) showed improvements over the TRL-6 results (see Table 1).

Table 1: TRL-6 and April 2006 Experiment Rerun Results

TRL-6 Black Course*					
Experiment	Date	Dist (km)	Avg Speed (kph)	km btw backups	km btw teleops
TRL-6	Dec 02-Mar 03	100	5.2	0.23	1.5
Rerun	6-May	54.7	8.1	0.33	2.5
* cross country and trails					

It was observed that one operator could successfully manage two robots in the missions in this experiment. As expected, operator workload increased in two-robot missions. The HMMWV-mounted 21-inch OCU screen and the handheld tablet were both satisfactorily used by the soldiers. This study also resulted in many interface improvements offered by the soldiers as feedback to the OCU developers [12].

4.2 OP Finder Experiment 2008

4.2.1. Background

Soldiers use OPs to observe enemy locations or movements and locate potential targets. There are many aspects to finding and occupying a good OP as discussed in various Army field manuals (e.g., Headquarters, Department of the Army. *Reconnaissance and Scout Platoon*; FM 3-20.98; Washington, DC, 3 August 2009.) Two important aspects of a "Good OP" are being able to see what you want to see and not being seen by the enemy. Accomplishing this task autonomously with a UGV is a challenge.

A February 2008 assessment of OP Finders and HMP was conducted as a follow-up to the 2006 study because of several key technology advances: (1) two independent OP Finder algorithms had been further developed and integrated with a more capable HMP for precise relocations at the OP, and (2) a stabilized multiple field-of-view RSTA camera and laser rangefinder system had been completed and integrated onto the XUV and deemed ready for field use (see Fig. 5).

The capability to autonomously find and occupy an OP is important for the survivability of UGVs and is needed to relieve the operator of the burden of trying to assist the UGV to find a "Good OP" through tedious and communications-intense interactions with the UGV. If successful, this capability would also contribute to stealthy tactical movements and to increased ability to conduct overwatch and similar tactical movements.

Fig. 5. XUV with stabilized RSTA suite

4.2.2. OP scenarios

Four OP courses were selected to provide a variety of environmental challenges in the way of near-field obstacles, distance to OP, differences in elevation, and availability of a priori feature data. The four courses were (a) an OP at the crest of a small hill overlooking a roadway at 200 m, (b) an OP behind a row of trees in a low-lying marshy area at 100 m, (c) an OP at the intersection of two trails looking up at a road and assembly area 1400 m away, and (d) an OP among bushes and trees overlooking a building and road 1000 m away.

Each experimental run consisted of a short XUV move to an OP that overlooked an NAI that was populated with recognizable features, such as silhouettes evenly spaced along a horizontal line, road intersections, and a gravel pile. Operators were asked to use the RSTA camera to identify the silhouettes seen at each NAI. Silhouettes were green plywood with painted white heads and white numbers to ensure that the targets were readily visible to the operator, as illustrated in Fig. 6.

Fig. 6. Silhouettes within an NAI

When the XUV arrived within 30 m of the OP, the OP Finder algorithms were invoked to find an acceptable OP. When the XUV stopped moving, pictures of the NAI were taken automatically with the RSTA camera, stitched together in a mosaic, and sent to the operator at the OCU. Each algorithm reported its success or failure to achieve a good OP. The operator would look at the RSTA image on the OCU, identify targets, and take additional RSTA images as desired to clarify his assessment of the visibility to the NAI and the identification of targets.

4.2.3. OP finder experimental design

Two OP Finder algorithms were used in the experiment. The goal of each OP Finder was to find a location near the OP that provided a measure of visibility to the NAI exceeding a threshold value (70%).

An original experimental design of 2 algorithms × 3 sites × 4 replications was planned. Snowy weather and technical failures during the experiment window caused some deviation to the plan during the execution of the experiment. Based on the experience of the first several days of experimental runs, excursions were added to look at the visibility threshold value needed for achievement of a good OP and the scan size of the area in which the OP Finder looks for a new OP. A fourth OP-NAI combination (“The Barn”) was added to expand the variety of OP locations used. At each location we attempted a balanced set of runs. Seventy-two runs were completed with full data collection, which included RSTA scans, OCU logs, RSTA metadata, mission plans, screenshots from the OCU, algorithm-reported results, and observations from the observer/data collectors.

Four performance measures were obtained on each run. The first was a “yes” (Y) or “no” (N) as to whether a “Good OP” had been reported by the OP Finder algorithm. The second measure was the number of repositions that were made for each run, i.e., the number of times the XUV moved before finding a “Good OP.” The third was the number of targets reported by the operator based on his analysis of the RSTA pictures. The fourth measure, percentage of NAI coverage, was derived from a post-experiment analysis of the RSTA images taken at the final OP.

The RSTA picture manual analysis estimated the amount of coverage of the NAI that was actually observed within the RSTA photograph. Two independent raters manually estimated the percent of coverage of the NAI, and the average served as the coverage estimate.

Several extraneous occurrences during a number of the 72 runs rendered some of the data unsuitable for use in data analysis. These reasons included heavy snow and fog obscuring RSTA pictures, emergency stops, software crashes, and RSTA system failures. After these considerations, 57 runs were included in the data analysis.

4.2.4 OP finder results

The data analysis shows that both OP Finder algorithms significantly improve the quality of selection of an OP location with respect to visibility to the NAI. Using the algorithm-reported results, OP Finder A improved the OP-finding performance from 31% to 69% and OP Finder B improved the OP-finding performance from 24% to 76%. A summary of the algorithm-reported results are shown in Table 2.

Results reported by the algorithms were optimistic when compared to the RSTA analysis results, and the performance varied between algorithms. Based on all runs, algorithm A self-reported 69% (22/32) “Good OPs,” and the manual RSTA analysis of these runs yielded 25% (8/32) “Good OPs.” Algorithm B self-reported 76% (19/25) “Good OPs,” but the manual RSTA analysis of these same runs yielded 56% (14/25) “Good OPs.” This indicates that OP Finder B holds up better against the manual RSTA picture analysis but that the algorithm-reported NAI coverage is not well calibrated to the actual coverage.

Table 2: Results of finding a "Good OP"

	OP Finder Algorithm A	OP Finder Algorithm B
Total Runs	**32**	**25**
First stop is a "Good OP"	**10/32**	**6/25**
% "Good OP"	**31%**	**24%**
OP Finder is invoked	**22**	**19**
can't find a "Good OP"	**10**	**6**
finds a "Good OP"	**12**	**13**
Total "Good OPs"	**22/32**	**19/25**
Total % "Good OP"	**69%**	**76%**

4.2.5 Other results

OP Finder algorithm speed varied inversely with the search area (scan size) for a new OP. Several excursions with larger scan size gave no evidence that scan size was a significant factor.

Finding a “Good OP” may be dependent upon the “goodness” of the original OP location designated by the soldier. OP locations input by the operator were varied slightly (+/–10 m) around the original OP. There were several instances when the variation of the OP location subjectively appeared to the data collectors and operator to result in differing levels of visibility achieved by the OP Finder, although this was not confirmed by the data.

RSTA and OP operations are central to the ongoing assessments of UGV autonomous operations. Since this was the first experiment with a live stabilized RSTA suite and the automatic OP Finders, many technical issues that affected the operational performance of the OP Finders and the HMP were reported back to the developers for their consideration. These included observations on the calibration of the OP Finder self-reporting, the need for HMP to point the XUV toward the NAI after moving to the OP, and suggestions for operator-related improvements to the OCU-RSTA interface.

This exercise was based on a simplified concept of movement to an OP and exposed the need for an overarching concept of how OP operations with a UGV will eventually be conducted in an operational environment. Following this experiment, a senior noncommissioned officer and one of the experimenters generated a more complete OP Finder definition, including the required interplay between XUV and operator. In this process, we identified parameters that should be available to operators for their decision and use. As the OP tactical behavior is further developed, determining the best process for "OP finding" will include active consideration of the part that the operator will play in assessing the adequacy of the identified OPs.

4.3 Capstone Experiment

4.3.1 Background

Terrain challenges to autonomous navigation, such as the vegetated cul-de-sac, motivated the advances in planning technology mentioned earlier. A pilot study conducted in 2008 and a field experiment in 2009 explored two approaches to two-level planning, which helped to shape the Capstone experiment where these advances were assessed. Based on pilot studies and simulation runs, the performance of two-level planners proved to be sensitive to parameters balancing the weight allocated to local and global costs of the overall plan and to other parameters relating to the height and density of grass and brush returned by the LADAR. The tuning of planner parameters was too expensive to be accomplished in the field, so the experimental conditions were modeled in the Robotic Interactive Visualization Experimentation Toolbox, a high-fidelity Hardware-In-The-Loop simulator. Planner parameters were studied over a variety of conditions and scenarios. The simulation results enabled the planner developers to establish a reference set for expected performance of the planners in scenarios similar to ground truth at Ft. Indiantown Gap, PA (FTIG).

As the name implies, the Capstone Experiment served as the culminating experiment to assess autonomous tactical behaviors enabled by technologies developed under the RCTA. The primary focus was on the assessment of the ability to autonomously overcome difficult navigation challenges in a wide variety of terrain types, a priori data sets, dynamic (moving) obstacles, and man-made route blocks. Secondly, we incorporated other RCTA technologies such as water detection, terrain classification, RSTA operations, and OP finding as well as scenarios to examine operator span of control and workload.

The Capstone scenarios stressed the XUV autonomous capabilities beyond previous field experiments so that limits could be explored. With a notional reconnaissance mission as a framework, the XUV was required to autonomously navigate, using varying levels of a priori terrain knowledge, from a start point to a goal over roads, trails, and cross-country and urban terrain, and around road blocks, cul-de-sacs, and moving vehicles. XUV performances were compared to the autonomous mobility capability that existed at the beginning of the RCTA program, providing a measure of progress toward UGV autonomous navigation. The Capstone experiment measured the ability of newly developed technologies to provide autonomous navigation in the most diverse and relevant operational environment ever presented to an autonomous UGV. We report on the primary focus of the Capstone experiment: autonomous navigation, a priori data levels, and dynamic obstacles.

4.3.2 Two-level navigation planners

Researchers from the ARL RCTA developed several new approaches to combining local and global planning into a hierarchical approach to autonomous navigation. Demo III autonomous mobility planning was done at the local (XUV) level only. The requirement to operate in diverse terrain types and among dynamic obstacles was the driver to create autonomous planners capable of dealing with the challenges found in each terrain type.

Four planning algorithms were used in the Capstone: AM Planner, DR, TR, and a Mixed Planner that combined the DR and TR.

AM Planner. In the DEMO III planning mode (referred to as AM Planner), the XUV received an initial global route generated by a global planner and relied only on locally sensed information to advance along the route. This planning mode most closely resembles the performance at the conclusion of the Demo III program and is used as the baseline capability. The AM Planner considers the kinematic constraints of the XUV and plans the least-cost mobility route for 40 m ahead but has no information on conditions beyond the LADAR range.

Dynamic Replanner (DR). In the DR mode, the XUV received an initial route generated by a high-level planner based on a priori data. En route to the goal, the global map was continuously updated with locally sensed terrain data, and the DR responded with a continuous generation of new potential global routes. When a new global route substantially improved the route to the goal, the new global route was automatically sent to the XUV to replace the current route.

Traffic Planner (TR). The TR incorporates a priori road network data from the world model and generates planned XUV positions and heading over time in highly cluttered, dynamic environments. By utilizing a road network, the planner is capable of finding the safest route while obeying traffic rules and responding to "on-the-fly" perception of dynamic obstacles.

Mixed Planner. The Mixed Planner is a combination of the TR (used when on the road network) and the DR (used off-road). Transition from one planner to another is accomplished autonomously using HMP.

4.3.3 Dynamic and static obstacles

Safe operations requiring autonomous operation of an unmanned ground system in proximity to humans and vehicles required advancements in perception and intelligent control. Only with regard to road rules and other vehicles is it included in this study.

Dynamic obstacles served as a course challenge in the Red course and the Combined Arms Collective Training Facility (CACTF) course (Fig. 7.). In all cases, the obstacles were vehicles (HMMWV or truck) that would interact with the path of the XUV, forcing it to adjust its route in some manner (e.g., plan around, slow, or stop). Specifics of how dynamic obstacles were employed were unique to the course.

Fig. 7. Dynamic obstacles in CACTF course

Static obstacles were placed in the Gold and Red courses to create cul-de-sacs and roadblocks that required the XUV to find and execute alternate routes (Fig. 8.).

Fig. 8. Static obstacles in Gold course

4.3.4 A priori data

Autonomous navigation is naturally affected by the level of prior knowledge of the terrain (elevation and feature data in the global map). The RCTA Program is built on the assumption of the absence of a priori data because this is often an operational reality. It is important to know, however, what the operational impact is of varying levels of a priori terrain data. In the Capstone, three types of map information were used as factor levels in the experiment: (1) road network, (2) maps generated by rotary-winged UAV flyover data, and (3) maps generated by structure from motion. The third is omitted here.

Road network a priori data is shown in Fig. 9, and combined with 20-m elevation posts, it served as the baseline map. Map inaccuracies, owing to the continually changing face of the terrain, added to the realistic situation where a priori maps cannot be taken as ground truth.

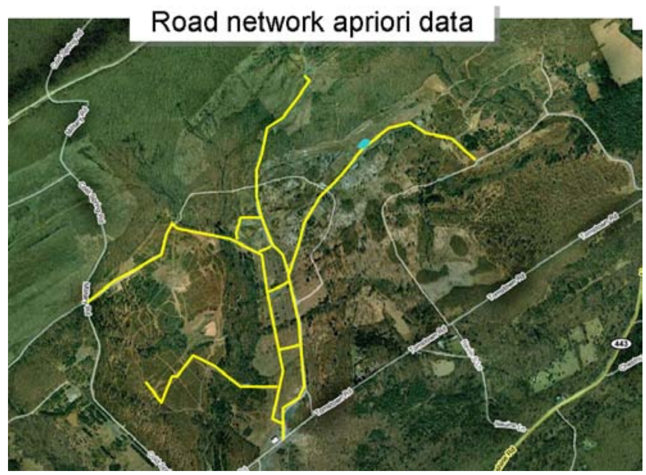

Fig. 9. Road network a priori data

Flyover map data was generated using an autonomous rotary-winged UAV and compact LADAR sensor (100 KHz, 265-m range, 5-mm accuracy).

The onboard 3-D mapper produced dense, colorized 3-D point clouds, which were interpreted by terrain classification algorithms as terrain types and combined with the road network map (Fig. 10). This method provided the most accurate update of feature data free from contamination from maneuver area alterations and differences in foliage.

Fig. 10. UAV-generated terrain map of FTIG

The Capstone built on numerous technologies generated during the RCTA in addition to several new capabilities not previously assessed. The Mixed Planner was a significant new capability, combining the DR for cross-country use and the TR for on-road use. Autonomous switching between modes was also a new capability. Two of the Capstone courses ran through both terrain types, and the Mixed Planner was created for just this situation.

Three courses discussed in this paper (Gold, Red, and CACTF) were designed to be traversable by the XUV in the baseline AM mode most of the time. During the experiment, the scenarios were made increasingly difficult by placing man-made blocks at key locations on the course, creating cul-de-sacs or complete roadblocks and by introducing moving vehicles into the scenarios. No single test course was suitable to examine the wide variety of technologies and capabilities, so several courses were used.

4.3.5 Course scenarios

Gold Course. The Gold course (Fig. 11) contained cross-country terrain, trails, and roads. A natural cul-de-sac 240 m deep and 50 m wide occurred early in the scenario. That cul-de-sac had two trail entrances that could be discovered by the perception system and would allow egress at the end of the cul-de-sac. Trails served as convenient locations for man-made blockages. Depending on the experimental factors for each run, one or the other of those entrances would be included in the global plan or would serve as an alternative route to be explored. The latter situation arose, for example, when the a priori map data did not include the trails, and the initial global route was an impassible path through the trees. In the two-block scenario shown in Fig. 13, one of the blocks is placed at the entrance to the north trail. If an alternative route was required, there was ample maneuver room in the cul-de-sac for the XUV. The road network associated with the Gold course also contained several areas where the road could be easily blocked. The on-road block required a cross-country bypass or an on-road backtrack to continue the route. This is shown midway through the Gold course.

Fig. 11. Gold course with two blocks

The Gold course was used to examine the effects of three levels of planners (AM, DR, and Mixed), three types of a priori data (road network, structure from motion (SFM), and Rmax), and three levels of terrain challenges (no blocks, two blocks, and four blocks).

Red Course. The Red course start and end points were chosen so that the initial global plan would make use of the existing road network in the a priori data (see Fig. 12). To acquire the goal point, the XUV had to leave the road network for the final 50 m and travel on a cross-country segment. Fixed blockages and dynamic obstacles were included in the scenario. One block could be solved only with an on-road alternative route, and the other required an off-road, cross-country solution. In addition, each run involved interaction with two vehicles (pickup trucks) via choreographed movement. Three separate route pairs for these trucks were randomized over the runs. Interaction types within these routes required the XUV to pass by an oncoming truck, overtake and follow a truck, and react to a truck pulling in front of the XUV.

The Red course was used to examine the effects of three levels of planners (AM, Traffic, and Mixed), three types of a priori data (road, network, and UAV), and two levels of terrain challenges (no blocks and two blocks). The Red course runs included vehicles serving as dynamic obstacles.

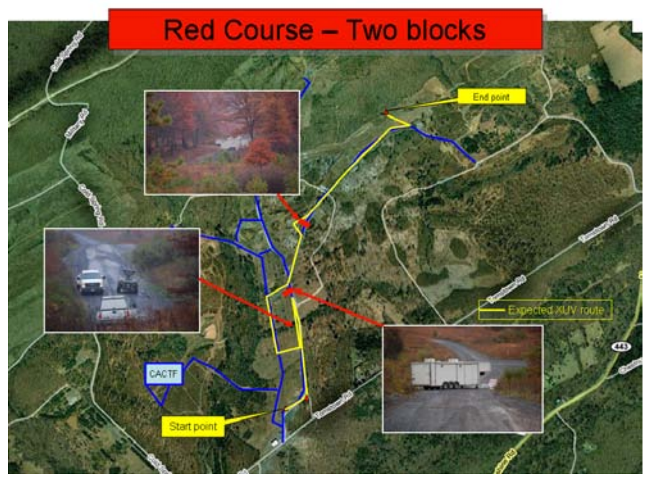

Fig. 12. Red course with two blocks

The experimental conditions exercised on each course were initially arranged in a factorial structure to be run according to a randomized schedule. On the red and CACTF courses, the design was adhered to; however, there were changes necessitated by findings during testing in combination with existing knowledge.

CACTF Course. The CACTF course consisted of one loop around the CACTF facility road network shown in Fig. 13. This was an urban course with buildings, open space, walks, intersections, and one traffic circle.

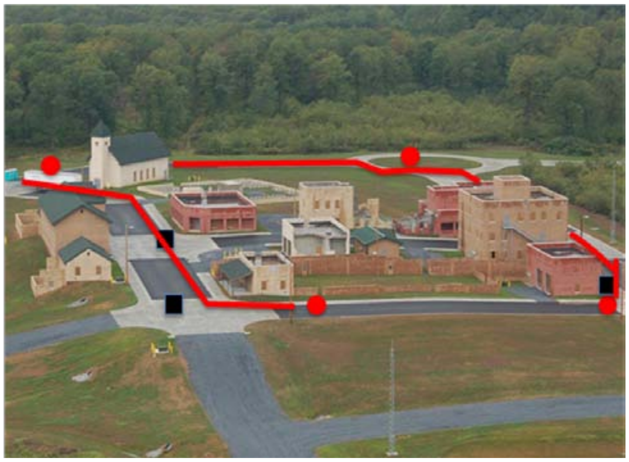

Fig. 13. CACTF course at Fort Indiantown Gap

The a priori data for the CACTF road network had two single lanes of opposing traffic. The start, end, and two intermediate waypoints were provided and are shown as circles in Fig. 13. The start and end locations were within 75 m on the east side of the course, and the XUV moved in a clockwise direction. The intermediate waypoints, far removed from the start and end points, were necessary to make the XUV travel to the west side of the course before returning to the goal. Based on these points, the global planner determined the initial route. Along the route, the XUV encountered one or two vehicles executing a choreographed movement that would interact with the path of the XUV. Event locations appear as squares in Fig. 13. There were no fixed blockages along the course for the principal block of runs—only those interactions with dynamic vehicles. Vehicle movements were constructed so that the XUV would need to find alternative paths constrained by rules of the road (see Fig. 14).

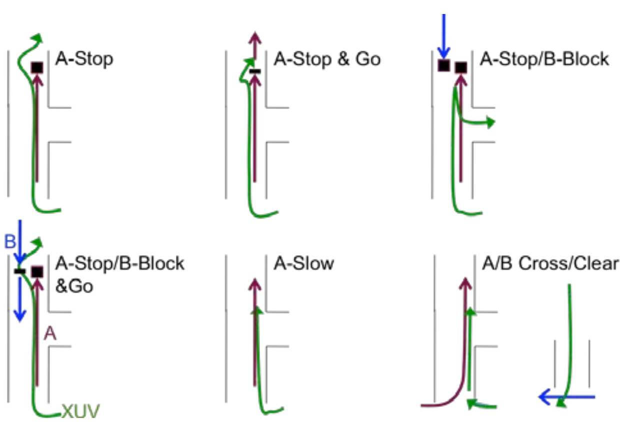

Fig. 14. XUV and dynamic obstacle movement patterns

The CACTF course was used to look at two levels of planners (AM and Traffic) in a constrained, "rules-of-the-road" scenario with six distinct movement patterns of the vehicles, A and B, interacting with the XUV. The movement names are suggestive of the interactions (A-Stop, A-Stop & Go, A-Stop/B-Block, A-Stop/B-Block & Go, A-Slow, A/B Cross/Clear). Some purposefully resemble movements in the DARPA Urban Challenge. This course required the XUV to process time and motion of objects in its path and to behave appropriately on a road network. It was expected that AM would not quickly overcome the shadowing created by objects passing in front of the XUV but that the newer planner would. The objective of each run was to avoid vehicles and navigate to the goal.

4.3.6 Results

Performance was assessed according to whether or not presented challenges were overcome in each instance and whether the goal point was reached. The approach taken was to identify mobility challenges along each route and record whether or not they were overcome. Except in a few exploratory cases, the first mobility challenge that defeated the system ended the run. Consequently, successful performance resulted in more challenges being presented. Achieving the goal point or not was also recorded, but as that binary outcome was subject to other issues (e.g., communications, platform failure) it is not regarded as a good measure of performance. To assist in this approach, usually two or three data collectors participated in each run, making notations about what the robot was doing in each situation. That information was combined in a database and provided a drill-down capability for use in after-action review.

The analysis strategy was simple. The fraction of challenges overcome to those presented were summarized for each treatment combination, tabled, graphed, and interpreted via the database of observer comments. It is important to note that the denominator was fixed to challenges encountered and not challenges possible. Sometimes challenges later in the course were not presented because of failure modes unrelated to the technologies under study.

Gold Course. The Gold course presented up to five challenges during a run: (1) find the open exit at the north trail of the cul-de-sac or avoid this trail if blocked, (2) find the open exit at the south trail of the cul-de-sac or avoid this trail if blocked, (3) discover and avoid the roadblock on the main tank trail by moving through trees to the right, (4) discover and avoid the roadblock on the tank trail entering the canopy area by backing up and going off-road to the left, and (5) discover a path up the hill to the helicopter pad. Results are included in Table 3.

As the difficulty of the course increased from 0 to 2 blocks, AM performance degraded and the DR solved every challenge. With 4 blocks on the course, DR solved 16 of 17 challenges and Mixed Planner solved 25 of 26 challenges. In a previous study, it was shown that AM could not solve any of the 4 blocks. Two-level planners are very effective in solving terrain challenges when faced with unanticipated terrain blockages in both cross country and on-road situations (62/69).

Table 3: Gold Course Terrain Challenge Results

Count of terrain challenges solved (Gold course)									
Blocks	No blocks			Two blocks (1 block@cul-de-sac and 1 road block)			Four blocks (2 blocks@cul-de-sac + 2 road blocks + path to helo pad)		
Planner	AM	DR	Mixed	AM	DR	Mixed	AM	DR	Mixed
Road network only apriori data	1/2	2/2		0/2	4/4			6/7	2/2
Rmax + road network apriori data	2/2	2/2		3/4	4/4			10/10	23/24

Without a priori data, AM solved 1 of 4 challenges, DR solved 12 of 13 challenges, and Mixed Planner solved 2 of 2 challenges. With Rmax a priori data, AM solved 5 of 6 challenges, DR solved 16 of 16 challenges, and Mixed Planner solved 23 of 24 (96%) challenges. The most important observation here is that AM performed poorly without a priori data while both DR and Mixed Planners performed equally well with and without good a priori data. This is consistent with the RCTA intent of operating well in unknown and unmapped terrain.

Red Course. The performance on the Red course is summarized in Table 4. With no blocks on the course, all planners performed well, although AM was slightly lower in percent of challenges solved. With two blocks, AM failed on every challenge, TR solved half of the challenges, and Mixed Planner solved 90% of the challenges.

Table 4: Red Course Terrain Challenge Results

Challenges (blocks only) solved (Red course)						
Blocks	No blocks + Two traffic vehicles			Two Blocks + Two traffic vehicles		
Apriori data/Planner	AM	Traffic	Mixed	AM	Traffic	Mixed
Road network only	1/2	1/1	1/1	0/2	2/4	3/4
Rmax + road network	2/2	2/2	2/2	0/2	2/4	6/6

When the difficulty of the course was increased with the addition of two blocks, the Mixed Planner far outperformed the Traffic and AM Planners because it could transition from the TR to the DR when required to generate an off-road path around the road blocks.

Without a priori data, the Mixed Planner outperformed the Traffic and AM Planners because it could transition from the TR to the DR when required to generate an off-road path around the road blocks.

With either a priori data set, the Mixed Planner solved more challenges than the TR, which, in turn, solved more challenges than the AM Planner. The Rmax data set increased the performance of each planner about uniformly, showing that accurate a priori data assists all planners in solving terrain challenges. With Rmax a priori data, the Mixed Planner far outperformed the Traffic and AM Planners because it could transition from the TR to the DR when required to generate an off-road path around the roadblocks.

CACTF Course. The CACTF course presented up to three challenges during a run. Each vehicle encounter was considered one challenge. Finding the way to the goal was another challenge. Also important was the manner in which the planning challenge was met. It was expected that the AM Planner, not having any notion of the rules of the road, would continue to take a low-cost local route, which would often lead it off the road onto sidewalks and through yard areas. To prevent damage to the CACTF facility, these runs were administratively stopped at the point the XUV left the road. In order to score perfectly on the run, the XUV had to successfully plan around challenges and reach the goal without being administratively stopped for leaving the road.

The results of the CACTF course trials are given in Table 5. Entries are the number of challenges met over the number of challenges presented. In all runs, the XUV was able to avoid and pass vehicles intersecting its path. Under the AM Planner, the XUV had to be administratively stopped for each run, and failure to achieve the goal is the reason why each AM ratio shows fewer solved challenges than presented challenges. Shadowing and pursuit of a low-cost solution without regard to the road hindered the performance of the AM Planner, but it was able to physically clear traffic blockages.

Table 5: CACTF Course Ratio of Traffic Challenge Outcome

Vehicle Movement/Planner	AM	Traffic
A-Stop	2/4	4/4
A-Stop & Go	2/4	4/4
A-Stop/B-Block	4/6	4/6
A-Stop/B-Block & Go	4/6	6/6
A-Slow	2/4	4/4
A/B Cross/Clear	4/6	4/4

The TP handled the six movements almost perfectly in accordance with normal traffic behavior. Only when the road was blocked in the A-Stop/B-Block condition was a challenge not met. In that instance, the XUV planned a path around vehicle A to the right on the edge of the sidewalk.

5. DISCUSSION

Assessing the readiness of UGV technologies is a challenge in an environment where every major factor seems to be changing—technology improves, new capabilities are achieved, terrain features change with seasons and weather, and the XUV has inherent mobility limitations. Radio communications between the XUV and OCU controller, while never a controlled factor in an experiment, often exerted unforeseen influence on operational aspects of the experiment. Technology innovations are often on the edge of readiness when taken to a field experiment, and developers are often still refining the algorithms and parameters that affect performance. This is a normal and highly useful practice for developers but creates difficulties when the objective is to assess a technology in a structured experiment where responses are measurable and attributable. Every experiment required flexibility on the part of developers and experimenters.

The metrics applied during the course of these technology assessments provided a qualitative measure of UGV-integrated performance and in some instances a quantitative evaluation of subsystem technologies. These assessment results kept ARL and Army leadership informed on the state of the art in UGV research and ensured that research funds were properly invested. Considering the nature of applied research with technology at various levels of development and maturity, it was assumed that the metrics used to evaluate system performance had to take into account end-to-end performance in a relevant environment if those measures are to be meaningful with respect to assessing capabilities in tactical behaviors. It is a worthy goal to derive additional metrics that can be applied across experiments so that improvements in performance can be assessed over time. We were able to accomplish this with very basic capability measures, but for more comprehensive measures capable of capturing the essence of a military function, this remains an open question.

We found that to formalize assessments as experiments rather than demos is a winning strategy, and good scenarios created in the planning of an experiment are as important as metrics to furthering understanding. There is always much to be gained by developers and testers from the consideration of basic questions concerning, for example, design factors, variables to be controlled, and response measures. We exercised the system over a wide range of representative scenarios for relevant environments and were opportunistic in the application of findings to raise the bar for subsequent development and assessment. Much can be learned from a good scenario by simply recording whether the challenge was or was not met.

In the cycle of model, deduction, data, induction, and remodel, there is more than one path to the goal. We would advocate broad notions of advancement to guide in a progression of rigorous experiments, thoughtful reflection at each stage, and openness to revise both the goals and the path. Frameworks for performance assessment and some standard metrics may provide a useful structure for this activity, especially for those unaccustomed to formal experimentation. We would caution against overscripting to favorite metrics or an idealized scenario because either is likely to lead developers to a brittle solution that has not properly taken into account the complexity of the environment.

In this paper we have provided an overview of the UGV assessment philosophy and practice associated with the RCTA program in the area of tactical behaviors. Details of specific field experiments highlighted the vast benefits to understanding afforded by a vigorous assessment program. The information provided to developers in a healthy feedback loop impacted both the direction and speed of technology advances. A program that began having only demonstrated semiautonomous mobility concluded having shown significant promise in full autonomy consistent with real operational challenges for the Army.

ACKNOWLEDGMENTS

The authors thank the engineers and technicians of General Dynamics Robotic Systems for their support in the field during preparation and execution of each technology assessment.

REFERENCES

[1] Bodt, B., and Camden, R. (2004), "Technology Readiness Level 6 and Autonomous Mobility." *SPIE Defense & Security, Unmanned Ground Vehicle Technology VI*, Vol. 5422, September 2004, pp. 302-313.

[2] Camden, R., Bodt, B., Schipani, S., Bornstein, J., Runyon, T., French, F., Shoemaker, C., Jacoff, A., and Lytle, A., *Autonomous Mobility Technology Assessment Final Report* (ARL-TR-3471), U.S. Army Research Laboratory, Aberdeen Proving Ground, MD, 2005.

[3] Government Accountability Office (GAO), *Best Practices: Better Management of Technology Can Improve Weapon System Outcomes*, GAO/NSIAD-99-162, Washington, D.C., 1999.

[4] Box, G., Hunter, J., and Hunter, W., *Statistics for Experimenters*, (2nd ed.), John Wiley and Sons, Hoboken, NJ, 2005.

[5] Cohen, M., Rolph, J., and Steffey, D. (ed.), *Statistical Testing, and Defense Acquisition*, National Academy Press, Washington, DC, 1998.

[6] Stenbit, J., Wells, L., and Alberts, D., *DoD Code of Best Practice: Experimentation*, Command and Control Research Program , Washington, D.C., 2002.

[7] Durst, P., "Performance Evaluation and Benchmarking for Unmanned Ground Vehicles," 2010, NDIA Ground Vehicle Systems Engineering and Technology Symposium, Dearborn, MI.

[8] Huang, H., Pavek, K., Albus, J., and Messina, E., "Autonomy Levels for Unmanned Systems (ALFUS) Framework: An Update," *SPIE Defense and Security Proceedings Vol. 5804*, 2005.

[9] Huang, H., Messina, E., and Jacoff, A., "Performance Measures Framework for Unmanned Systems (PerMFUS)", Performance Metrics for Intelligent Systems Workshop *(PerMIS)*, Gaithersburg, MD, 2009.

[10] Satterfield, B., Choxi, H., Salamon, A. and Drewes, P. (2008), "Advancing Robotics: The Urban Challenge Effect," *Journal of Aerospace Computing, Information, and Communication, Vol. 5, pp 530-542*.

[11] Bodt, B., Childers, M., and Camden, R., "A Capstone Experiment to Assess Unmanned Ground Vehicle Tactical Behaviors Developed under the Robotics Collaborative Technology Alliance," presented at AUVSI North America 2010 Conference, Denver, CO, 24-27 August 2010.

[12] Hill, S., and Bodt, B. (2007), A field experiment of autonomous mobility: Operator workload for one and two robots. *Proceedings of the 2007 ACM/IEEE Conference on Human-Robot Interaction*, Arlington, VA, March 8-11 2007 (pp. 169-176).

[13] Childers, M., Bodt, B., Hill, S., and Camden, R. (2008), "The Impact of Multi-Level Path Planning on Unmanned Ground Vehicle Tactical Behavior." 26th Army Science Conference, Orlando, FL, December 2008.

[14] Childers, M., Bodt, B., Hill, S., Dean, R., Dodson, W., and Sutton, L. (2007), "Assessing the Impact of Bi-directional Information Flow in Unmanned Ground Vehicle Operation: A Pilot Study" Performance Metrics for Intelligent Systems Workshop (*PerMIS)* Gaithersburg, MD, 2007.

[15] Childers, M., Bodt, B., Hill, S., Camden, R., Dean, R., Dodson, W., Sutton, L., and Sapronov, L., *Unmanned Ground Vehicle Tactical Behaviors Technology Assessment* (ARL-TR-4698), U.S. Army Research Laboratory, Aberdeen Proving Ground, MD, January 2009.

[16] Childers, M., Bodt, B., Hill, S., Camden, R., Gonzalez, J. P., Dean, R., Dodson, W., Kreafle, G., Lacaze, A., and Sapronov, L., *Unmanned Ground Vehicle Two-Level Planning Technology Assessment* (ARL-TR-5331), U.S. Army Research Laboratory, Aberdeen Proving Ground, MD, September 2010.

Marshal A. Childers is the Acting Team Leader for the Integration & Assessment Team, Autonomous Systems Division, of the U.S. Army Research Laboratory. Mr. Childers holds B.S. and M.S. Degrees in Mechanical Engineering from the University of Maryland, Baltimore County and in 2001 began as a Civilian Employee at ARL, working on munitions concepts, design, and experimentation. He began working on ARL robotics programs in 2005 and since 2007 his primary responsibility has been to lead unmanned ground vehicle technology assessments which focus on performance in environments that are relevant to the military.

Barry A. Bodt received a B.S. in Mathematics from the University of Maryland, College Park in 1980 and a M.S. and Ph.D. in Statistics from the University of Delaware in 1985 and 1989, respectively. He joined the Ballistic Research Laboratory in 1979 and served as a Mathematical Statistician for that lab and its successor, the Army Research Laboratory. He has led analytical efforts in a variety of applications—ballistic protection, network intrusion detection, chemical agents, toxicology, target detection, and war game simulation to name a few—and was awarded the Army Wilks Medal in 2006 for his contributions to Army Statistics. Since 2002, he has been integrally involved in ARL's robotics program, with a responsibility to develop operationally relevant and statistically sound assessments of maturing technologies for unmanned ground systems.

Richard Camden received a BS in mathematics from the Indiana University of PA in 1967 and an MS in mathematics from the University of Tennessee in 1969. After serving two years in the US Army, he worked at Brunswick Corporation as a statistician prior to joining the US Army Human Engineering Laboratory in 1974 as a mathematician. From 1997 to 2000 Mr. Camden served as the Technical Director of the OSD Joint Logistics Advanced Concept Technology Demonstration which earned the Army Superior Unit Award for logistics support to Operation Joint Endeavor in Bosnia. Mr. Camden received the ADPA Artificial Intelligence Individual Achievement Award in 1991 and the DA Meritorious Civilian Service Award in 1997. In 2001 he moved to the Robotics Program Office of ARL where he served as a Project Officer, responsible for oversight of assessment activities of autonomous mobility technologies for unmanned ground vehicles. Since 2006 he has been an Analyst with MPRI providing technical and operational support to ARL in the design, conduct, analysis, and documentation of field experiments to measure technology readiness for unmanned ground vehicle technologies.