Real-Time, Multi-Perspective Perception for Unmanned Ground Vehicles

Anthony Stentz, Alonzo Kelly, Peter Rander, Herman Herman, Omead Amidi
Carnegie Mellon University

Robert Mandelbaum, Garbis Salgian
Sarnoff Corporation

Jorgen Pedersen
Robotics Engineering Excellence

Contact:
Anthony Stentz
National Robotics Engineering Consortium
Robotics Institute, Carnegie Mellon University
#10 Fortieth Street
Pittsburgh, PA 15201
Email: tony@cmu.edu
URL: www.rec.ri.cmu.edu/projects/perceptor

ABSTRACT

The most challenging technical problems facing successful autonomous UGV operation in off-road environments are reliable sensing and perception. In this paper, we describe our progress over the last year toward solving these problems in Phase II of DARPA's PerceptOR program. We have developed a perception system that combines laser, camera, and proprioceptive sensing elements on both ground and air platforms to detect and avoid obstacles in natural terrain environments. The perception system has been rigorously tested in a variety of environments and has improved over time as problems have been identified and systematically solved. The paper describes the perception system and the autonomous vehicles, presents results from some experiments, and summarizes the current capabilities and limitations.

1. INTRODUCTION

In order to navigate autonomously, an Unmanned Ground Vehicle (UGV) must be equipped with sensors for measuring the terrain, software for interpreting the data, path planning to decide a safe course of action, and control to oversee that plan. Of these technologies, perception (i.e., sensing and data interpretation) are the most challenging. As part of DARPA's Perception for Off-Road Mobility (PerceptOR) program, "Team Blitz" assembled a perception prototype that was tested on an All Terrain Vehicle (ATV) platform. First results were previously reported [1]--this paper presents results from the second phase of the program.

The Team Blitz perception prototype design is based on the following principles:

• Sensor viewpoint greatly determines the difficulty of mobility analysis.

• Viewpoint fusion is required to simultaneously improve resolution and widen the wide field of view.

• Crash survivability is required because autonomous systems are not perfect; they will make mistakes.

• Multi-modal sensor fusion is required to deal with complex outdoor terrain.

Certain sensor viewpoints are fundamentally better for detecting certain types of hazards. For example, the detection of drop-offs and holes, and in particular the discrimination of lethal drops from traversable downslopes, is extremely challenging from the ground vehicle perspective. The challenge is that the disambiguation between the two cases – lethal drop-off vs. safe downslope – requires direct sensing of the down side of the potential hazard. With the UGV's low viewpoint, it must navigate extremely close to the potential hazard before knowing for sure whether or not the path ahead is safe. This observation led to a radical design element in the Team Blitz system, a Flying Eye (FE). The FE is an aerial vehicle that flies ahead of the UGV to detect holes, downslopes, and other difficult obstacles from an ideal, overhead perspective.

In general, viewpoint fusion uses multiple sensors to get more resolution across a meaningful field of view. No single off-the-shelf sensor can provide both at the same time, so multiple sensors are necessary for the near term. An angular resolution of at least 1 degree is critical for attaining a meaningful vehicle speed, and a 90-degree field of view is considered the bare minimum for operations in complex terrain that requires frequent tight turns.

Multiple sensors with significant baselines have other benefits as well. Notably, a single sensor can be entirely blocked by relatively little clutter in a scene (dust, smoke, grass, trees, etc.), whereas multiple sensors with separation increase the likelihood that at least one sensor will have a meaningful view. This also increases survivability. Smashing one's only sensor leaves the system blind, while smashing one of a set of sensors may hamper the system, but still leave it operational.

With the addition of the FE, Team Blitz added a major new aspect to viewpoint fusion, making use of FE data to benefit ground vehicle navigation. The viewpoints are radically different, but they observe the same terrain. The data must be registered to the ground vehicle in order for the UGV to make use of it.

Finally, it is our conjecture that no single sensor can provide sufficient information to completely deal with the complexities of outdoor environments. The UGV must understand geometric aspects of rocks, dirt, trees, ledges, holes, etc., as well as the non-geometric aspects of vegetation, airborne obscurants, water, and mud. No single sensor can discriminate all these features, let alone provide sufficient information for an autonomous system to safely navigate through terrain filled with this content.

The only reasonable hope to solve this problem is to exploit multiple sensors with complementary characteristics. Team Blitz has pursued this strategy from the start of the program by including passive visible and infrared light sensing and active infrared ladar sensing.

Figure 1. First (left) and second (right) generation UGV test vehicles based on the Honda Rubicon ATV platform.

2. AUTONOMOUS NAVIGATION SYSTEM

Our design principles led to a solution consisting of a UGV equipped with multiple sensors with differing fields of view, and a FE with a downward looking sensor for early obstacle detection. This section provides a high-level description of the vehicles, sensors, and navigation software.

2.1 UNMANNED GROUND VEHICLE

The UGV system consists of a base ATV (Honda Rubicon), retrofitted for drive-by-wire operation, carrying perception sensors and computing hardware for semi-autonomous perception and navigation. Figure 1 shows the first generation of this vehicle (left) used in most of the program and the second generation (right) which replaced it late in Phase II. The second generation system has significantly improved the base vehicle stability, reliability, and terrainability with better onboard sensing, perception, and navigation.

The second generation vehicles use a pair of SICK single-axis ladars near the front of the UGV, with one scanning horizontally (across the UGV path) and the second scanning vertically (along the UGV path). The horizontally scanning laser can be tilted and the vertically scanning laser can be panned, creating the poor man's versions of two-axis ladars. The vehicles have two additional SICK ladars for special purposes. One looks rearward to guide backup maneuvers. The other sits in the center of the top of the UGV. We employ a slip ring to allow this vertically scanned ladar sensor to rotate continuously, providing a 360-degree view around the UGV. As with the other scanners, this scanner is programmable. For forward travel, we often scan +/- 90 degrees off the heading of the UGV, for example. A digital color video camera is mounted between the scanning SICK ladars. This camera adds color to the SICK ladar data, enabling color-based drivability classification of ladar data. The vehicles were equipped with mechanical bumpers to detect and stop for obstacles missed by the perception system. To improve robustness, these mechanical bumpers were later replaced by static, inertial bumpers (i.e., inertial sensors to detect impact forces) to perform the same function.

The processing architecture for the UGV is depicted in Figure 2. The vehicle controller performs closed-loop control of the vehicle's actuators to maintain a fixed curvature and speed. The path planner performs closed-loop control on UGV position by receiving UGV state from the state (pose) estimator and issuing commands to the vehicle controller. The mobility planner also performs obstacle avoidance based on the traversability assessments provided by the UGV perception system. If the Flying Eye is available, the mobility planner incorporates traversability maps produced by the FE perception system to improve obstacle avoidance and global navigation. Finally, if data about the area is available a priori, they are processed into prior maps offline. The path planner

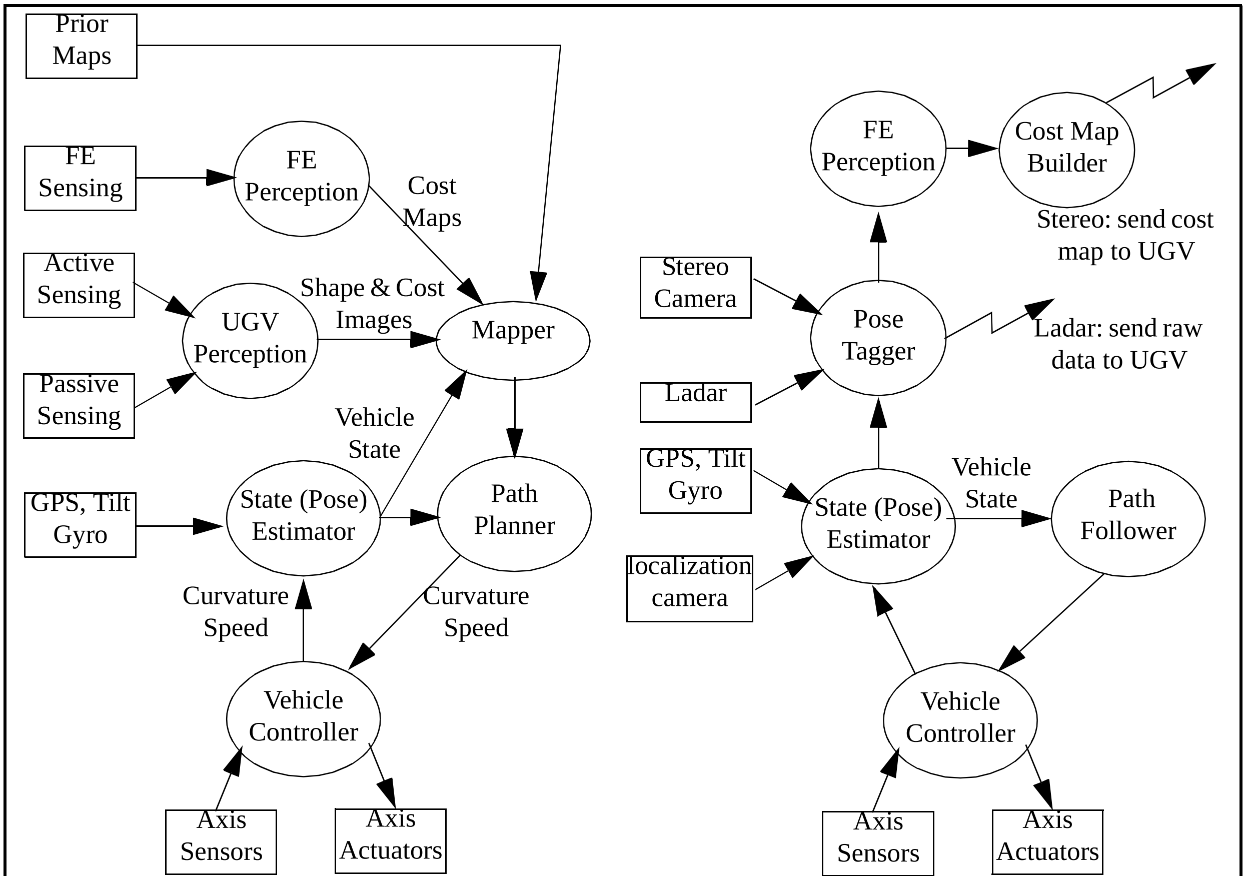

Figure 2. Processing diagrams for the UGV (left) and FE (right). The circles are software modules and blocks are data.

uses these maps to improve route selection.

2.2 FLYING EYE

The Flying Eye (FE) vehicle consists of a base helicopter retrofitted for fly-by-wire operation, along with perception sensors and computing hardware. Early in the phase, we used a manually-flown Bergen helicopter equipped with a downward looking stereo vision system to detect obstacles, tag them with position, and radio them to the UGV. Late in the phase, we switched to an autonomous Yamaha R50 helicopter, equipped with a single-axis scanning laser rangefinder, configured to scan lines on the ground orthogonal to the direction of travel at a rate of 20 Hz scans and 12 KHz pixels (see Figure 3). The raw scans were position tagged and transmitted to the UGV for processing. The UGV controlled the FE by computing its desired trajectory and commanding the FE to clear the path by flying through a series of way points. Presently, we have switched to an autonomous Yamaha RMAX helicopter to take advantage of its greater payload capacity, to accommodate both downward looking ladar and stereo for obstacle detection, and longer flight time.

Figure 3. The Yamaha R50 autonomous helicopter (left) and the Yamaha RMAX autonomous helicopter (right).

The processing architecture for the FE is shown in Figure 2. The figure is the composition of all architectures flown to date, including past, current, and planned. The FE uses a variety of sensors to detect obstacles, tag them with pose, and transmit them to the UGV as it flies a specified path.

3. NAVIGATION AND PERCEPTION SOFTWARE

The navigation software consists of two primary control loops: the Central Nervous System (CNS) and the deliberative autonomy loop. The CNS cycles very quickly (e.g., 100 Hz for some tests) providing a basic “reflex like” safeguarding function for the vehicle. The CNS monitors the vehicle bumper for collision, the inertial sensors for dangerous roll or pitch angles, the GPS for lost signal, and the transmission for stuck gears and stops the vehicle accordingly. The condition is reported to the deliberative autonomy loop or the operator for corrective action.

The deliberative autonomy loop is responsible for moving the vehicle to its goal in a safe manner. It cycles more slowly (e.g., 5 to 10 Hz), considering candidate steering arcs for the vehicle to drive in each cycle. Arcs are analyzed for safety using a forward simulation [2] and for goal acquisition using an incremental global path planner [3] and a winner is selected. In the forward simulation, the vehicle is convolved with the “cost of traversal” data reported by the perception subsystem and stored in the local map. These costs run a continuum from “safe” to “lethal”, with the safest arc being the one with the lowest aggregate cost. If no arcs are available without lethal obstacles, the vehicle stops and backs up.

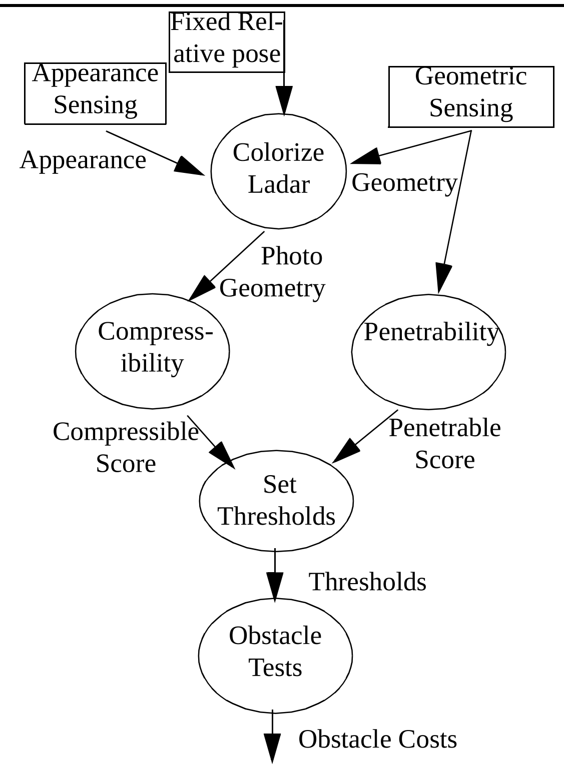

Figure 4. Perception System Block Diagram. In the data flows shown lower arrows include new information in addition to the information available immediately above it.

Figure 4 shows the perception software architecture. Essentially there are two sensing modes. Appearance sensing (passive) includes color cameras and FLIR cameras. Geometric sensing includes ladar (active) and either daylight or FLIR stereo (passive). The geometric sensing module processes the range data to detect geometric hazards (i.e., extreme slope, tree, vertical wall) and to compute traversability cost. The cost is loosely related to the probability of a mission failure if the UGV traverses the sensed region. This geometric processing is further modulated by using estimates of the non-rigidity or "compressibility" of the sensed terrain. These estimates relax the thresholds which are used in the obstacle tests. Intuitively, the tolerable slope of a soft bush is greater than for a hard rock.

There are two forms of rigidity calculations. The "compressibility" module uses registered appearance and geometry data to assess terrain compressibility. Examples include the use of range texture, color, and color texture in order to discriminate rigid rocks from pliable bushes. The "penetrability" module performs a density analysis on geometric data in order to assess whether ladar beams are penetrating a region which indicates that the region is of low material density.

Both the UGV's ladar/monocular camera and stereo vision systems provide the geometric and appearance data for obstacle assessments. Additionally, the FE's ladar provides geometric data. The results of these assessments are fused from multiple sources in both space and time.

3.1 UGV LADAR

The ladar system uses several time-of-flight laser rangefinders with registered camera data to determine the range and appearance of points in the scene. The approach has the advantages of measuring range directly, not requiring scene texture, and providing greater accuracy than stereo vision. It has the disadvantages of being an active sensor, acquiring data incrementally rather than in full frame, and providing range only (no appearance data). The latter weakness was rectified by registering a color camera with the ladars. The ladar/camera system is the primary navigation sensor for the UGV.

Early in the phase, we attempted to register raw ladar range data into a single coordinate frame (i.e., local map) and process the data for obstacles. Due to errors in vehicle position estimation, this approach resulted in data misalignment and false obstacle detection. To reduce our sensitivity to pose, we processed individual scan lines for geometric obstacles and registered the traversability costs into the local map, dramatically improving our performance. Late in the phase, we extended our perception system to use factors other than solid geometry (e.g., size and shape) to assess "drivability" or "traversability" of the terrain. We measured two indirect properties of the terrain: compressibility, a measure of object rigidity is based on its material composition, and penetrability, a measure of object density is based on how solid or porous it is. These properties were measured in the following ways:

- Color is a powerful cue to compressibility. A neural network was configured to learn typical classifications from color data. Not surprisingly, green objects (grass, weeds, bushes) tend to be compressible, while brown objects (dirt, rocks, tree trunks) tend to be rigid. Dead grass and moss-covered tree trunks are confounding cases, but color is still a powerful cue.
- Density is a powerful cue to penetrability. A special purpose ladar data processor was developed which processes multiple scans in order to count the number of hits and misses in a small volume of space. Here, we use

the often overlooked fact that a range measurement provides evidence not only of a reflecting surface, but of empty space between it and the sensor. By thresholding this density score, we get a classification of the compressibility of the voxel. Chain-link fences and concertina wire are confounding cases, but density is still a powerful cue. This module is also very effective at detecting overhangs.

These two measures enabled our perception system to drive through tall grass, weeds, and some bushes while still avoiding solid, incompressible objects like large trees.

3.2 UGV STEREO VISION

The stereo vision system uses a pair of cameras to triangulate range to points in the scene. The approach has the advantages of being passive, providing a full frame of data at once, and co-registering appearance data with the geometric data. It has the disadvantages of requiring appearance texture to work and is generally less accurate than ladar, especially at greater ranges. The stereo vision system uses similar algorithms for detecting terrain hazards such as positive obstacles and steep slopes, and these algorithms were shown to be sound; however, the system was limited in the type and size of obstacles it could detect. In particular, narrow vertical structures, such as trees, were a problem. We traced the problem to the range reconstruction process, and we spent most of our effort increasing the reliability and accuracy of this process through both hardware and software improvements.

In particular, we developed a new stereo camera mount to increase stability and to minimize correlation problems due to vibration. Second, we eliminated the assumption that the cameras had parallel optical axes (no vergence) and solved for the vergence parameters. Third, we solved for the lens distortion, rather than assuming there was none. These three improvements eliminated warping from the scene geometry. Fourth, we switched from a vertical camera configuration to a horizontal one, thus eliminating our dependence on what is generally weak vertical texture for vertical objects. Fifth, we switched from a ground plane horopter to a traditional vertical horopter, since the ground plane is not prominent in many complex scenes (e.g., dense woods) and is difficult to isolate.

3.3 FE LADAR

The Flying Eye ladar is a single-axis device that scans orthogonal to the flight direction. As the FE flies forward, the sensor “paints” the ground below it. The data is pose tagged and transmitted to the UGV for processing. The

FE ladar data is processed in a two-dimensional manner by fitting a plane the size of the UGV's footprint at every point on the terrain. The roll and pitch are analyzed to determine traversability. Large angles are scored as obstacles since they likely correspond to steep slopes, deep holes, or large positive obstacles. This approach worked very well in environments with sparse vegetation, including trees. However, dense vegetation was a problem. Additionally, we had difficulty with mis-registered data acquired during aggressive FE maneuvering. Processing tree canopy in aerial data is fairly analogous to the issue of processing occluding ground cover in ground vehicle imagery. In both cases, it is difficult to locate the surface which will support the vehicle. Further, assuming that the geometry is rigid leads to incorrect slope assessments both at the edges of canopy and inside the canopy. Our solution to the first problem was to simply detect canopy based on its signature of significant height variation and declare the region as unknown for planning purposes. For the second problem, we extended the processing to discard the associated ladar data during aggressive FE maneuvers in order to eliminate the generation of false obstacles that would be caused otherwise.

3.4 SENSOR FUSION

The various sensor modalities (i.e., UGV ladar, UGV stereo vision, FE ladar) provide traversability cost assessments of the same terrain at multiple times. These assessments must be fused together along with those assessments generated from prior map data, if it is available. At first, we fused data by averaging all costs over sensor modality and time, equally weighting all data, including older data with newer data. This strategy exhibited two problems: 1) response to "lethal" obstacles was slow since the obstacle would need to be seen many times before the average cost would climb to a sufficiently high level, and 2) those sensors with high data bandwidth would overwhelm those with lesser bandwidth for the same patch of terrain.

For the first problem, we introduced a "fast response" to lethal hazards resetting the number of measurements and setting the current cost to lethal whenever a lethal cost arrives. When a non-lethal arrives, we averaged as usual. The results were sporadic, so we introduced a recursive filter for combining new sensor data with the old value to produce a new value: (new value) = (alpha) * (old value) + (1-alpha) * (sensor data), where "alpha" is a constant (typically 0.5). This sensor fusion strategy exhibited much improved results.

For the second problem, we treated each data source as a channel and time averaged the data within each channel. We fused data by taking, for each cell in the terrain map, the highest cost value for the cell across the different channels. This approach levelled the playing field for sensors with different data bandwidths, and it provided a clean way for combining sensor data with prior map data.

4. PERCEPTION SYSTEM TESTS

The Team Blitz perception system was tested in sparse woods in Virginia, desert terrain with washes, gullies, and ledges in Arizona; mountainous terrain with pine forests in California, and dense woods with tall grass and other vegetation in Louisiana. In an unrehearsed manner, the vehicles were given a set of waypoints to drive through autonomously and their performance was evaluated. The metrics evaluated included the time to travel the courses, the average speed, communications bandwidth (between vehicle and operator control station), level of remote operator involvement, and number of emergency stops. For example, a long time could indicate that the perception system had too many false positives. A large number of emergency stops could indicate too many false negatives.

Figure 5 shows the first generation vehicle operating in desert terrain in Arizona. Figure 6 shows a one-kilometer traverse in desert terrain. The black regions are obstacles detected in the prior map data. The white regions are areas that are easy to drive. The gray regions are of intermediate difficulty. The blue curve shows the intended path based on prior data. The red curve shows the actual path driven. The two differ because the UGV detected additional obstacles enroute that caused the path to be re-planned. The UGV did not experience any emergency stops during the traverse.

Figure 7 shows the vehicles in action at some other sites, including a first generation UGV detecting and driving across a “hasty bridge” over a ravine in mountainous terrain in California (left), and a second generation UGV driving through a grassy area in Louisiana.

Figure 8 and Figure 9 show some perception results. Figure 8 shows a scene of terrain with grass and weeds. The terrain is flat and drivable everywhere. Using geometric analysis alone, most of the terrain is classified as hazardous (red areas), since the vegetation is assumed to be rigid and solid and therefore would impede forward motion. But by using drivability analysis, these initial classifications are relaxed and the terrain is determined to be safe. Figure 9 shows two results from the FE, including a large trench detected from the air using stereo vision, and a set of smaller holes detected using ladar.

Figure 5. First generation UGV (left) drives through desert terrain. Overview of desert terrain (right).

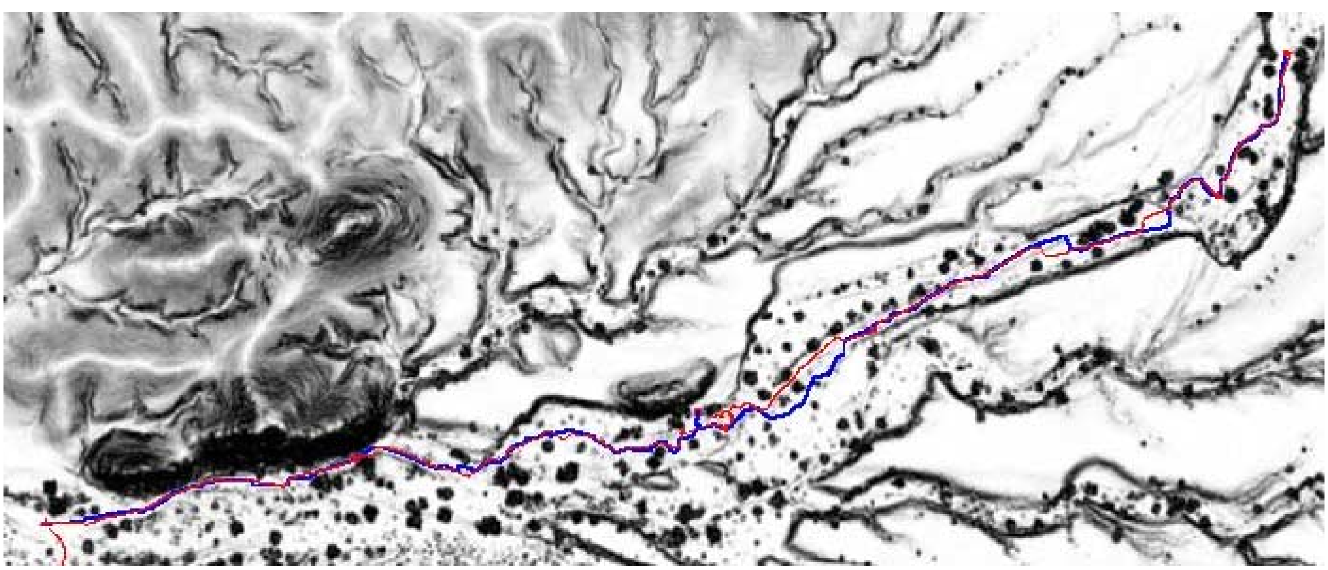

Figure 6. Plan view of one kilometer traverse through desert terrain without emergency stops.

Figure 7. First generation vehicle (left) drives over hasty bridge. Second generation vehicle (right) drives through grassy field.

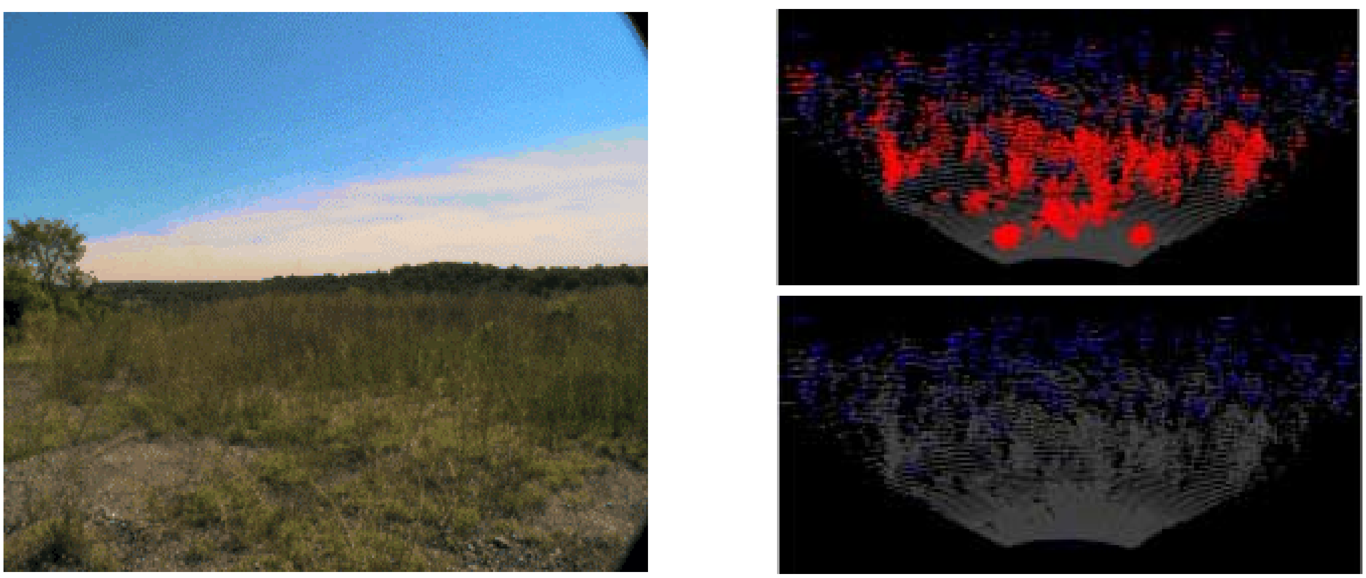

Figure 8. The picture on the left shows flat terrain covered with grass and weeds. The terrain is completely traversable. The top right shows the traversability classification for this terrain using geometric data only. The red areas are false obstacles; the blue areas are unknown. The bottom right shows an updated classification taking "drivability" metrics into account. Note that the vegetation has been eliminated as obstacles.

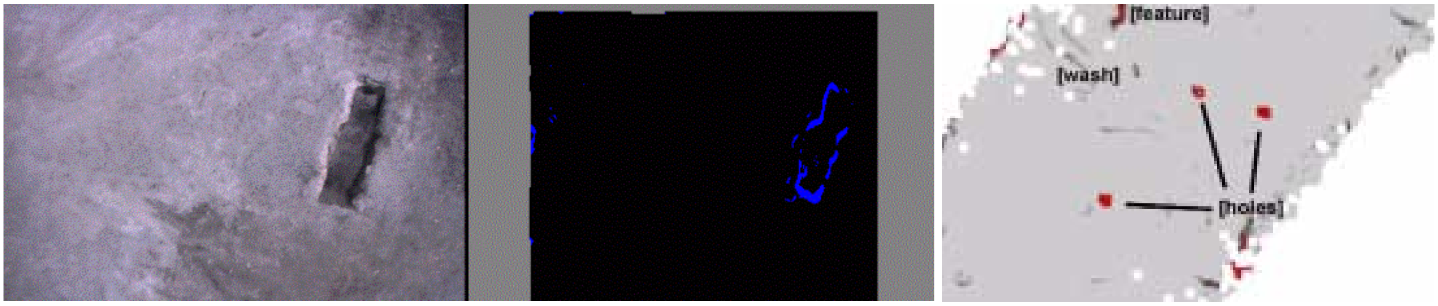

Figure 9. The left image shows a photo of a coffin-sized hole taken from the FE at 10 m altitude. The center image shows the hole detected as an obstacle using stereo vision. The right image shows several holes the size of an excavator bucket detected with ladar from the FE at 25 m altitude.

5. SUMMARY OF PERFORMANCE CAPABILITIES AND LIMITATIONS

Our navigation system was capable of detecting and avoiding large positive obstacles, such as trees, rocks, and large bushes; walls and other vertical structures; steep slopes that would cause tipover; and large negative obstacles such as cliffs and deep gullies (when using the FE). Note that these obstacles are discernible using geometric analysis alone. Later in the phase, we extended our perception system to detect and ignore vegetation such as tall grass, weeds, small bushes, and tree limb overhangs.

We have targeted the following areas of improvement for our perception system:

- Small positive obstacles: thin trees, poles, and small, sharp rocks (i.e., posing a tire hazard). The biggest problem is getting dense enough and accurate enough range resolution to detect these fine features.
- Small negative obstacles: tire-sized holes, long and thin ruts. Sensor vantage point continues to be the primary limitation. Small holes are typically not visible to the UGV until it is very close--perhaps too close to stop. The FE offers the optimal perspective, but higher resolution sensing and smoother scanning are needed to ensure proper coverage of these hazards.
- Non-geometric objects: objects for which factors other than geometry (such as material property) primarily determine their traversability. Examples include water, mud, ice, snow, bramble, wire, and fences. We made some progress detecting and ignoring certain types of vegetation, but more work is needed.
- Air/ground obscurants: dust, smoke, rocks in grass. The ladars and stereo vision systems used in our perception system were susceptible to classifying air obscurants as positive obstacles. The UGV would stop and wait for

the obscurant to clear, but this reduced average forward speed. Obstacles that were obscured by ground cover, such as rocks in the grass, were not visible to the navigation sensors. At times they were detected by the bumper during collision and the vehicle was able to recover autonomously.

- Passive night vision: our perception system had no such capability, but we did investigate the use of stereo vision using Forward Looking InfraRed (FLIR) cameras.

Team Blitz made substantial progress in PerceptOR Phase II toward fielding a fully autonomous UGV, but many problems remain to be solved.

6. ACKNOWLEDGEMENTS

This work was sponsored by DARPA, under contract “Perception for Off-Road Mobility (PerceptOR)” (contract number MDA972-01-9-0016). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies or endorsements, expressed or implied, of the U.S. Government. This document was approved for public release, with unlimited distribution.

The authors thank all members of Team Blitz, including Mike Bode, Mark DeLouis, Antonio Diaz-Calderon, Todd Dudek, Keith Hanna, Jim Ketterer, Eric Kratzer, Jeff McMahill, Ryan Miller, Bryan Nagy, Andre Rieder, John “Ben” Southall, Randy Warner, Nick Vallidis, Jon Woytek, and Marc Zinck. We would also like to thank the Government Team, including Scott Fish, Eric Krotkov, Mike Perschbacher, and Jim Pippine.

7. REFERENCES

[1] Stentz, A., Kelly, A., Herman, H., Rander, P., et al, “Integrated Air/Ground Vehicle System for Semi-Autonomous Off-Road Navigation,” Proceedings of AUVSI-2002, July 2002.

[2] Kelly, A., Stentz, A., “Rough Terrain Autonomous Mobility - Part 2: An Active Vision, Predictive Control Approach,” Autonomous Robots, 5, 163 - 198, 1998.

[3] Stentz, A., “Optimal and Efficient Path Planning for Partially-Known Environments,” Proceedings of the IEEE International Conference on Robotics and Automation, May 1994.