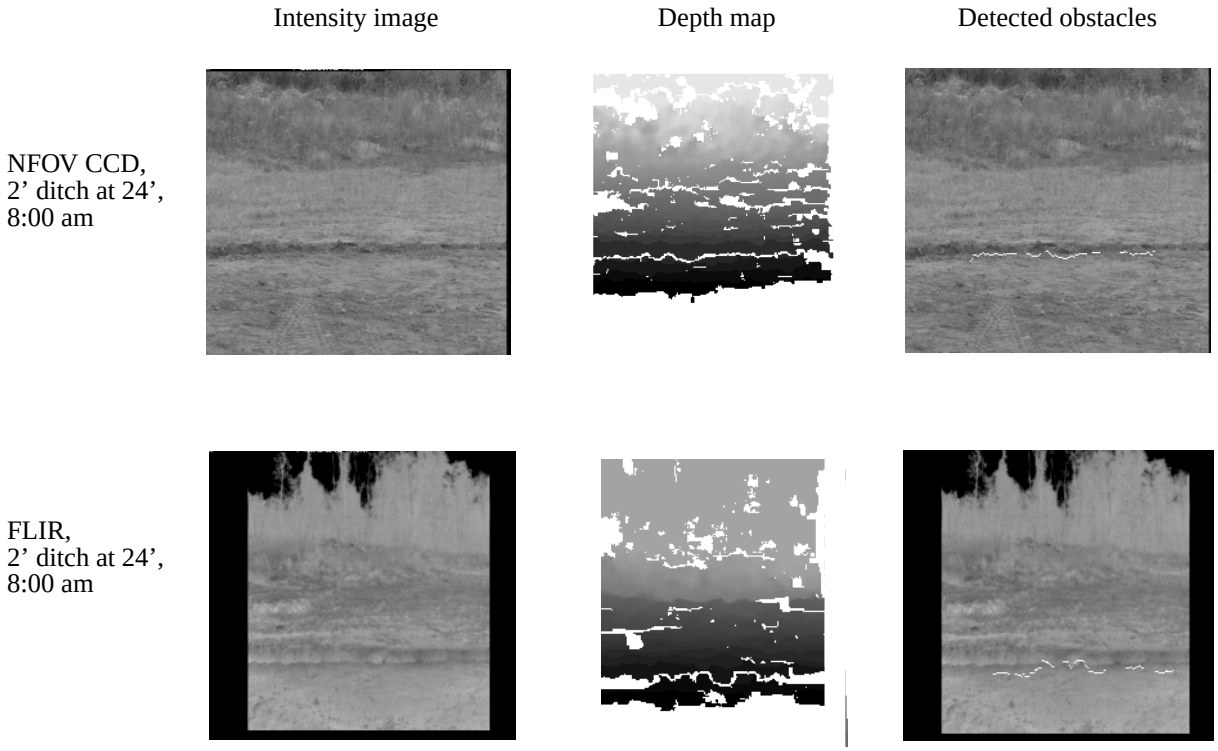


Figure 10: Sample negative obstacle detection results for CCD and FLIR stereo vision.

[3] "*Demo III Experimental Unmanned Vehicle (XUV) Program: Autonomous Mobility Requirements Analysis*", technical report, Robotic Systems Technology, Inc., April 1998.

[4] A. Thompson, "The navigation system of the JPL robot", *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-5)*, pp. 749-757, 1977.

[5] H. Moravec, *Obstacle avoidance and navigation in the real world by a seeing robot rover*, PhD thesis, Stanford University, September 1980.

[6] K. Olin and D. Tseng, "Autonomous cross-country navigation: an integrated perception and planning system", *IEEE Expert*, 6(4):16-30, August 1991.

[7] L. Matthies, "Stereo vision for planetary rovers: stochastic modeling to near real-time implementation". *International Journal of Computer Vision*, 8(1):71-91, July 1992.

[8] C. Elachi, *Introduction to the physics and techniques of remote sensing*, John Wiley and Sons, New York, NY, 1987.

[9] R. N. Colwell (ed.), *Manual of Remote Sensing*, American Society of Photogrammetry, Falls Church, VA, 1983.

[10] R. C. Bolles, H. H. Baker, M. J. Hannah, "The JISCT stereo evaluation", *Proceedings of the DARPA Image Understanding Workshop*, Morgan Kaufmann Publishers, pp. 263-274, April 1993.

[11] L. Matthies, A. Kelly, T. Litwin, and G. Tharp, "Obstacle detection for unmanned ground vehicles: a progress report". *Robotics Research 7*. Springer-Verlag, 1996.

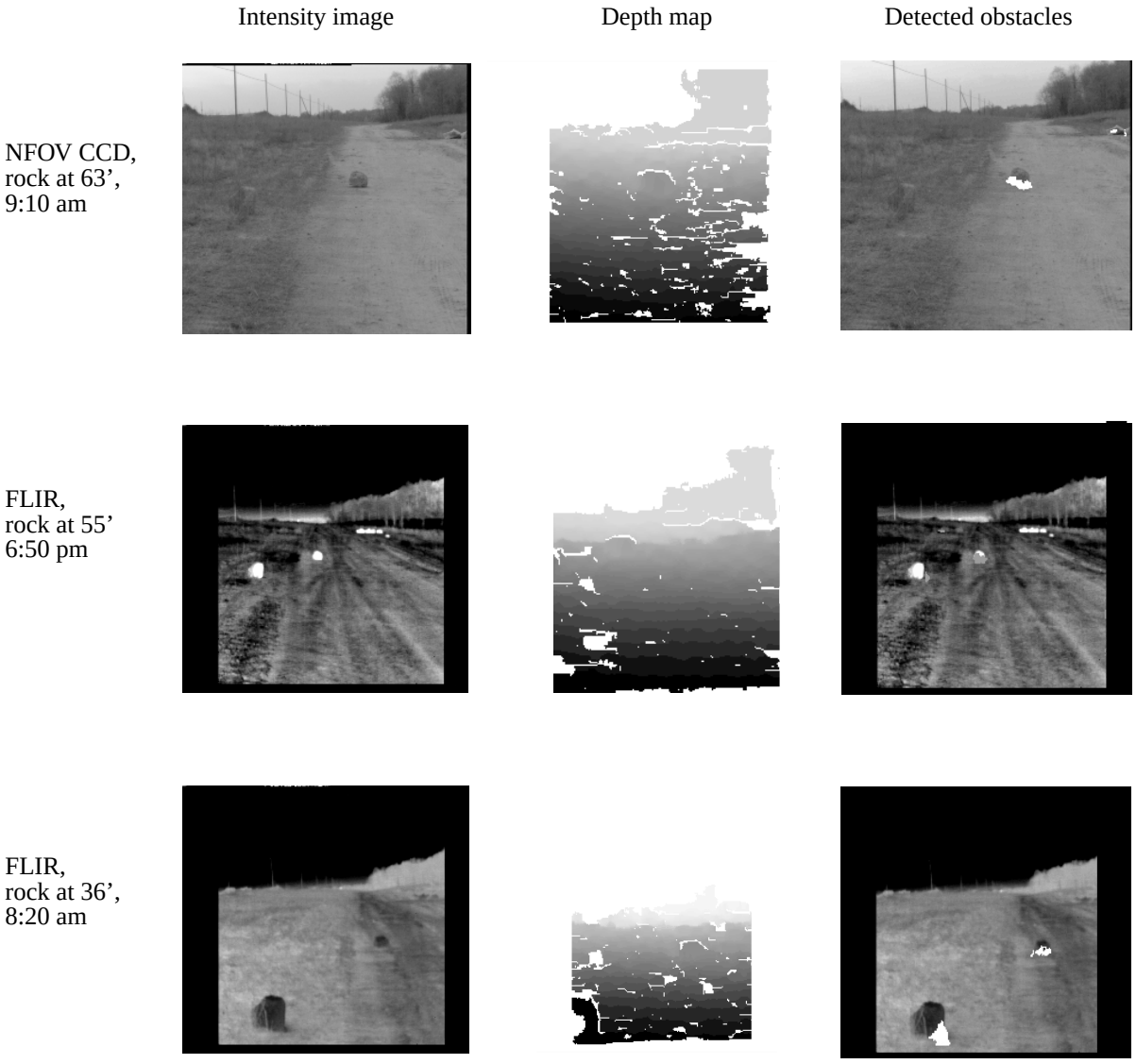


Figure 9: Sample obstacle detection results for CCD and FLIR stereo vision on 11" rock at various distances. Detection results are shown as white overlays in top and bottom image, grey overlay in middle image.

Acknowledgements

The authors would like to thank Chuck Shoemaker, Jon Bornstein, Gary Haas, and Tim Vong of the Army Research Lab, Fred Eldridge of the Tooele Army Depot, and Ish Sherali of the Aberdeen Test Center for their assistance in conducting this data collection. We also apologize to anyone inadvertently left off this list.

References

[1] L. Matthies and P. Grandjean, "Stochastic performance modeling and evaluation of obstacle detectability with imaging range sensors", *IEEE Transactions on Robotics and Automation*, 10(6):783-791, December 1994.

[2] A. Kelly and A. Stentz, "An analysis of requirements for rough terrain autonomous mobility", *Autonomous Robots*, 4(4), December 1997.

am, the contrast is much lower; detection has been achieved at 36 feet, with 11 pixels on target.

Although these are by no means definitive or exhaustive results, they still show detection at ranges equal to or exceeding the lookahead requirements discussed in section 2 for cross-country driving (60 feet for day driving at 20 mph, 25 feet for night driving at 10 mph). With the NFOV CCD stereo pair in results not shown here, the 11 inch rock has been detected at 87 feet (10 pixels high), but not at 114 feet (8 pixels high). Similar tests have not yet been conducted with the FLIR imagery.

For ditches, figure 10 shows good detection results for the 24 inch ditch seen 24 feet away for both the narrow field of view CCD stereo pair and the FLIR stereo pair. The ditch subtends about 15 pixels in the CCD range image and 6 pixels in the FLIR range image. Tests with the CCD-based imagery have obtained satisfactory detection at 36 feet, where the ditch subtends 7 pixels, but not at 42 feet, where it subtends 5 pixels. These ranges do not satisfy the lookahead requirements for stopping at 20 mph, although they do meet the requirements for small-angle turns.

To examine false alarm rates, the obstacle detection algorithms have been run on 100 images from the flat road data set, varying the size of obstacle the detector was tuned for. Preliminary results show that the number of false alarms for the NFOV CCD cameras drops to zero when the positive obstacle size exceeds 6 inches and the negative obstacle size exceeds 20 inches. Although this is promising, further testing is necessary, particularly on bumpier terrain.

6.0 Summary, Conclusions, and Future Work

This paper has reviewed the overall performance goals for autonomous mobility in the Demo III program, which include achieving obstacle avoidance at 20 mph for cross-country driving by day and 10 mph by night. The vehicle being designed for this program is expected to be able to negotiate positive obstacles up to 12 inches high and negative obstacles 24 inches wide. We summarized estimated requirements for lookahead distance at these speeds, which are 60 feet for stopping at 20 mph and 25 feet for stopping at 10 mph. Corresponding numbers for obstacle avoidance with small-angle turns are 35 feet and 22 feet, respectively. We also gave strawman angular resolution requirements for obstacle detection based on range data from stereo vision, assuming a detection criterion that requires the obstacle to subtend at least 5 pixels in the image. These numbers feed into ongoing design work to how many cameras, what image size (512x512 or 1024x1024), and whether or not high performance pan/tilt and gaze control are necessary to simultaneously satisfy requirements for angular resolution and field of regard. A key issue in this is to assess the realism of the "5-pixel rule".

We then described a large data collection of CCD stereo pairs, InSb FLIR stereo pairs, and LADAR that was conducted at the Aberdeen Proving Grounds in November and December 1997. The collection includes data sets for rocks from 6.5 to 13 inches high, ditches from 2 to 8 feet wide, a flat dirt road for measuring false alarm rates, and a dirt road with low hills for validating performance on hills. For the rocks in particular, data sets were collected at various times before and after dusk and before and after dawn to allow stereo vision performance evaluation with thermal imagery with varying degrees of contrast.

The stereo imagery has been processed into range data at an angular resolution of about 1 mrad/pixel for the NFOV CCD's and 2.3 mrad/pixel for the FLIR's. According to section 2, this should be adequate to stop for positive obstacles of the requisite size and to do small-angle turns for negative obstacles. Initial performance evaluation results for stereo vision showed fairly good depth maps, except that low contrast FLIR imagery required increasing the size of the cross-correlation window. An 11 inch rock was detected at distances up to 87 feet with CCD imagery and 55 feet with FLIR imagery; this exceeds the estimated lookahead distance requirements for stopping. A 24 inch ditch could not be detected beyond 36 feet with CCD imagery; this is adequate for a small-angle turn maneuver, but not for the lookahead required to stop at 20 mph. Reasonably good detection performance is not achieved with 5 pixels on target, but is with 10. Preliminary results with false alarm statistics show negligible false alarms for these size obstacles on smooth surfaces; of course, the real test will be on more bumpy surfaces with vegetation.

In conclusion, initial performance evaluation results suggest that the obstacle detection performance needed for Demo III speeds can be met with stereo vision systems with reasonable IFOV and algorithms similar to those already in use, at least on semi-arid terrain. The stereo baseline for these experiments was 40 cm. The FOV of these systems does not fill the overall FOR requirement, so camera pointing (or multiple camera sets) seems to be required. This is a significant area for future research, particularly as it interacts with path planning and local map maintenance. Work in the coming year will also do performance evaluation on terrain like 2 course at APG. The much more difficult issue of obstacle detection sensors, algorithms, and performance evaluation in vegetated terrain will be addressed over the next two years.

cally assess the detectability of the obstacle sizes specified in section 2 at the lookahead distances estimated in section 2. Longer term goals of this work include developing and validating more sophisticated models of obstacle detectability, for example that relate probability of detection and false alarm to sensor design parameters (angular resolution and stereo baseline), illumination conditions, detector sensitivity, and design of the detection algorithm.

5.1 Quality of Range Data

Prior work on performance evaluation of stereo vision has shown good results on well-exposed, CCD stereo image pairs of outdoor scenes; that is, it has shown precision of disparity estimates better than 1/10 pixel [1], with 87% or more of pixels having acceptable depth estimates [2]. A key question for this data set is how well stereo works for low contrast conditions with FLIR cameras; for example, near the diurnal thermal crossover.

Some practical and logistical problems were experienced in this data collection. In particular, some data sets have less than ideal calibration, and some experienced inadvertent, sub-optimal setting of aperture or exposure time. We largely suppress these details here.

The stereo algorithm applied to generate depth maps was the real-time algorithm described in [11]. All imagery was acquired interlaced; hence, it has been processed as fields at 256x240 resolution.

We generally obtained fairly good depth maps, as can be seen in figures 9 and 10. Our primary observation so far concerns low contrast FLIR imagery; for example, the morning imagery shown in the figures. To get good depth maps for this imagery, we have had to increase the cross-correlation window size from the 7x7 pixels we typically use to 9x9 or 11x11. Part of this may be due to less than ideal calibration or exposure time settings; however, we anticipate that such adaptation will be required in some low contrast conditions regardless of calibration and exposure. Therefore, algorithms to do such adaptation automatically and efficiently will be an important part of future work.

5.2 Quality of Obstacle Detection

We have done initial testing with the rock, ditch, and flat road data sets to evaluate maximum detection ranges and to generate initial sets of false alarm statistics. Since these results will vary with the obstacle detection algorithms used, we briefly review these first.

5.2.1 Algorithms Applied

For this paper, we applied the column-oriented detection algorithms for both positive and negative obstacles that are described in [11]. For positive obstacles, for each pixel in the range image, we find a second pixel in the same column above it that would correspond to the top of a minimum-size vertical step, if such a step had its base at the first pixel. We compute the change in height and slope between these two pixels; wherever these quantities exceed a threshold, we label the first pixel as on an obstacle. This "step detector" is followed by a "blob filtering" stage that uses 4-connectivity to extract connected regions of positive obstacles; regions whose width falls below a threshold (typically about 4 pixels) are eliminated as likely false alarms.

The negative obstacle detection algorithm measures change in slope as a cue to presence of a ditch. Given a pixel that is a candidate to be at the front edge of a ditch, we fit one line segment to all pixels in that column below that point and another line segment to pixels above that point up to the apparent bottom of the ditch. If the change in angle between these two segments exceeds a threshold, and the width of the apparent ditch also exceeds a threshold, we declare the candidate pixel to be at the leading edge of a negative obstacle. Connected components are also formed, in this case with an 8-connected criterion since the leading edge of a negative obstacle is typically a linear feature in the image, rather than a compact blob as for positive obstacles. Again, blobs with width less than a threshold are eliminated.

5.2.2 Detection and False Alarm Results

Parameters were tuned to detect the minimum obstacle sizes discussed in section 2 (12 inch positive obstacles and 24 inch negative obstacles). Figure 9 shows example results from the rock data set, figure 10 from the ditch data set.

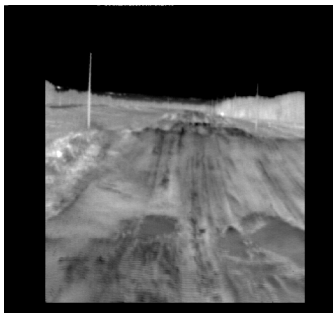
For rocks, we have been able to detect an 11 inch rock at a distance of 63 feet from the camera using the narrow field of view CCD stereo pair. In figure 9, the white region overlaid on the CCD image in the rightmost column shows where an obstacle was detected; there are no false alarms. In this case, the obstacle subtends about 13 pixels for the 256x240 level of resolution at which we processed the imagery.

As seen in figure 9, FLIR imagery of rocks at 5:50 pm has high contrast, especially between the rocks and the soil. Since the rocks appear white in the image, the obstacle detection overlay in the rightmost column is shown in grey; the rock is detected at 55 feet with no false alarms. Here the rock subtends 7 pixels. At 8:20

Photo of flat road site

Photo of 3-course site

WFOV CCD, NFOV CCD, and FLIR imagery from road



WFOV CCD, NFOV CCD, and FLIR imagery from 3-course

Figure 8: Representative images from flat road data set and 3-course data set. LADAR imagery was also acquired at these sites.

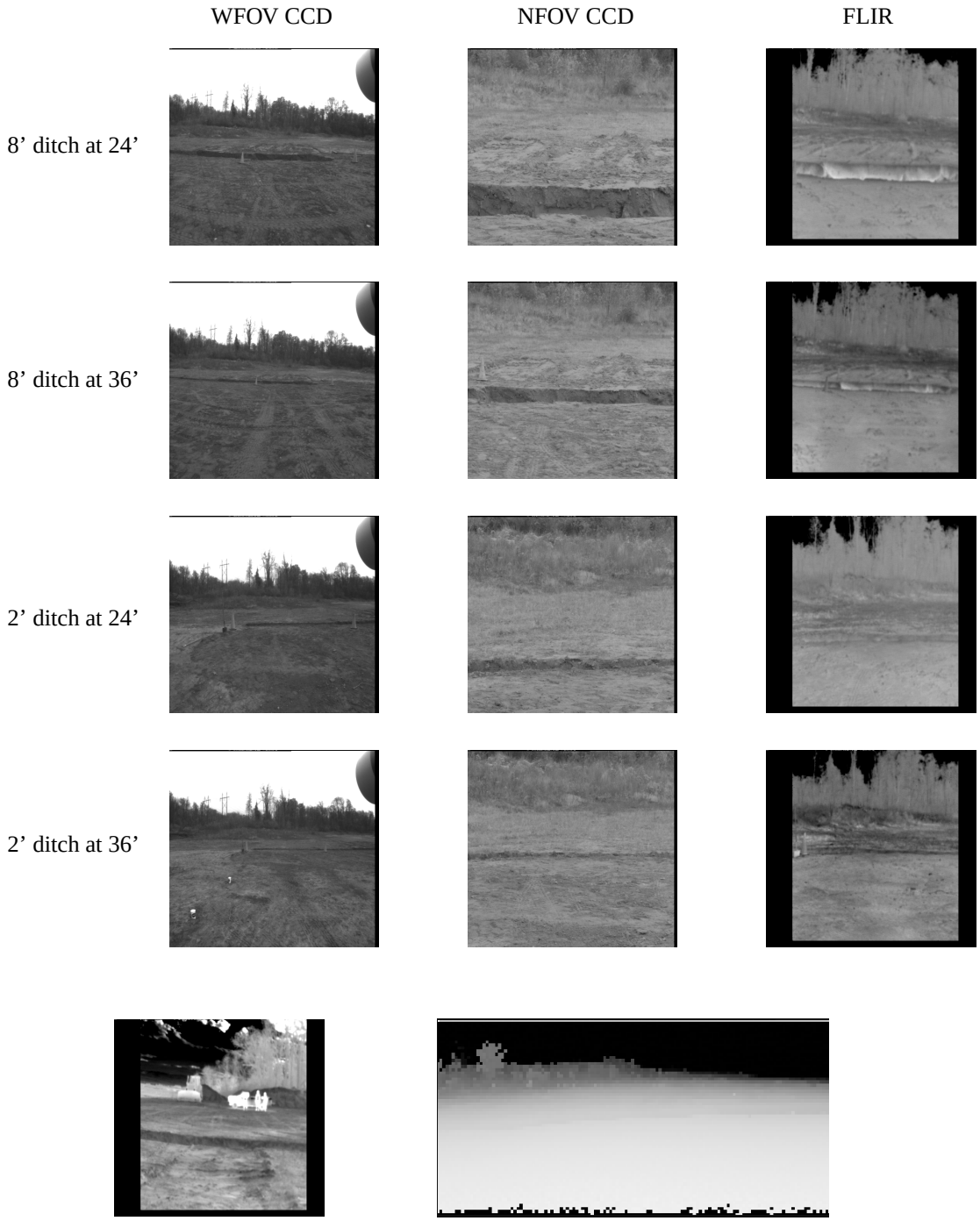
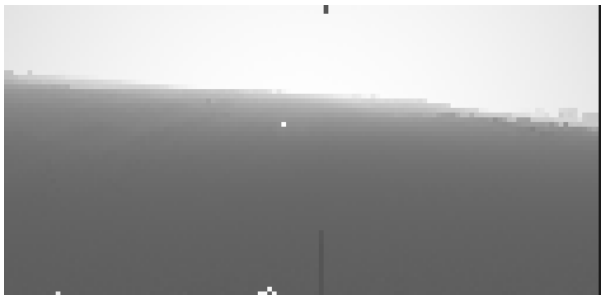


Figure 7: Representative images from ditch data set. Additional sequences were acquired approaching the ditches diagonally, instead of straight on.

(a) Photo of rock layout



(b) LADAR range image of rocks

(c) FLIR, 06:00

(d) FLIR, 06:50

(e) FLIR, 08:20

(f) FLIR, 17:00

(g) FLIR, 21:20

(h) WFOV CCD, 06:50

(i) NFOV CCD, 06:50

Figure 5:Rock data sets: (a) Documentation photograph, (b) LADAR image, (c)-(g) FLIR images at various times, (h) wide angle CCD image at 06:50, (i) narrow-angle CCD image at 06:50. Note varying contrast of FLIR imagery.

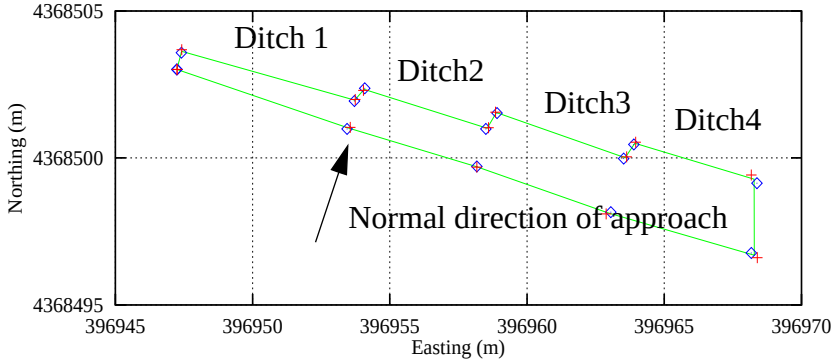


Figure 6: Left: surveyed location of ditches. Widths are 2', 4', 6', and 8' for ditches 1 through 4, respectively; lengths vary from 14' to 21'. Graph units are UTM coordinates. Right: documentation photograph.

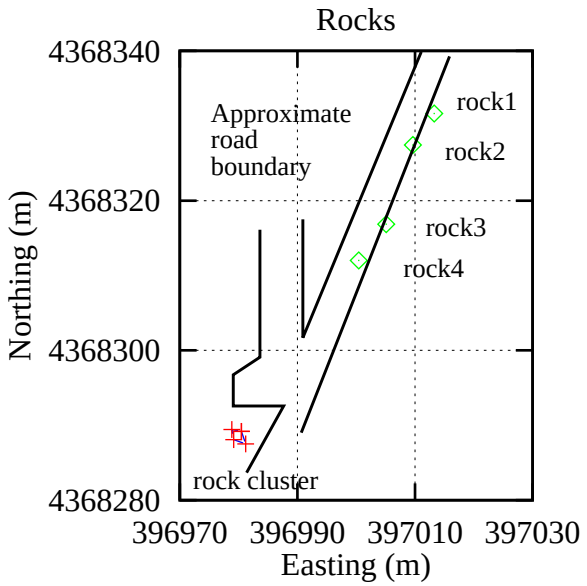


Figure 4: Sketch of dirt road layout with surveyed positions of rocks 1 to 4; heights of these rocks were 7”, 6.5”, 13”, and 11”, respectively. Graph units are UTM coordinates.

stored on a digital video disk. For image sequences grabbed on the move, a continuous capture loop snapped all three stereo pairs every 1/5 of a second.

4.1 Rocks

The rock data set (figure 4) used four rocks spaced at intervals of about 25 feet along the edge of a dirt road, with rocks 1 and 3 in vegetation just off the road and rocks 2 and 4 on the road; rock heights ranged from 6.5 to 13 inches. A cluster of boulders was also visible to the cameras 100 feet beyond rock 4. Figure 5 shows a sampling of the imagery acquired.

Data sets were acquired under the following conditions:

- Late afternoon/early evening: (a) vehicle stationary, advancing 6 feet per image set through the rock field and (b) driving at 5 mph through the rock field.
- Pre to post-dawn: (a) vehicle stationary and (b) driving at 5 mph
- Early afternoon (1:30 pm), driving at 10 mph.

Variations in contrast of the thermal imagery as a function of the time of day are evident in figure 5.

4.2 Ditches

To provide controlled conditions with good terrain ground truth, 4 ditches were dug in a row in a flat, gravel area, with widths from 2 to 8 feet and lengths from 14 to 21 feet (figure 6).

Data sets were acquired by driving the HMMWV toward the ditches, both with perpendicular and diagonal approaches, sometimes stopping to take the imagery and sometimes driving continuously at 5 mph. Data sets included:

- Mid-afternoon, with the vehicle stopping every 6 feet from 18 to 60 from the ditch; acquired sequences for all 4 ditches. Also acquired sequences for all ditches while driving at 5 mph.
- Mid-afternoon, 45 degree diagonal approach at 5 mph. Also, 5 mph perpendicular approach from the “back side”.
- Early morning (8:00 am), ditches 1 and 4 only, vehicle stationary only, perpendicular approach only, stopping every 6 feet.

Figure 7 shows sample imagery from each type of camera. Note that instances were observed in the thermal imagery where the visible side of the hole was both warmer and cooler than the level ground surface, depending on the direction of the sun.

4.3 Flat Road and 3 Course

Figure 8 shows sample imagery from data sets taken on the flat road and over the low hills on 3 course. Sequences were taken on the flat road at 10 and 20 mph, heading northward at roughly 1:30 pm, with 100 stereo pairs per sequence. Sequences were taken on 3 course at 5 and 10 mph, heading southward shortly after noon., with 100 to 200 stereo pairs per sequence. 3 course was wet, with distinct puddles at low spots.

5.0 Initial Results from Performance Evaluation

At the time of writing, performance evaluation of stereo matching and obstacle detection with this data set is work in progress; therefore, we report results to date, but expect to produce a more extensive report in the future. The focus in this paper is on results with stereo; we plan to compare stereo and LADAR performance in future work.

The short term goals of this work are to (a) evaluate the quality of the range data, especially for FLIR stereo through the thermal cross-over period and (b) empiri-

that is, trying to use both narrow and wide FOV stereo pairs at once to provide a foveal/peripheral capability. The thermal cameras had a 34x34 degree FOV, or 2.3x2.3 mrad IFOV at the detector resolution of 256x256 pixels. The LADAR available for this effort was the Dornier EBK LADAR, which has 128x64 pixels in a 60x30 degree FOV, or 8.2x8.2 mrad IFOV; its scan rate is 1 frame/second. Exposure times for the CCD cameras were between 1/60 and 1/125 of a second, with auto-iris lenses. The thermal cameras had a fixed aperture and were set up for approximately a 1 ms exposure time.

Our original intent was to collect imagery from all of these sensors simultaneously on November 11 and 12, 1997. However, technical problems with the LADAR forced postponement of the LADAR data collection to December 1 and 2. Site surveys were used to set up the test course to be as nearly identical as possible for the LADAR data collection as for the stereo collection, and matching data sets were acquired.

The Perryman area at APG was selected for data collection. Terrain available at Perryman includes several roadway "courses" offering different conditions and a variety of off-road areas. The roadway courses are all gravel or dirt surfaces in the following conditions:

- 1 course: a smooth road surface.
- 2 course: a moderately bumpy road surface, with bumps and ruts with heights and depths of several inches.
- 3 course: a dirt road with a series of low hills several feet high.
- 4 course: a road with multiple parallel, roughly sinusoidal, out of phase variations at an amplitude and wavelength intended to stretch the suspension of military vehicles to the limit (ie. order of two feet in amplitude).

Differential GPS coupled with INS was used to tag all imagery with the vehicle position and attitude at the time of image capture. To simplify logistics, all data sets were collected in the vicinity of a single GPS base station set up near 3 course; a map of this area is shown in figure 3. Data sets of rocks and ditches were acquired on available terrain next to 3 course. As a "control" case for evaluating false alarm rates, image sequences were obtained on a stretch of flat dirt road adjacent to 3 course. Since hill crests are known to be a source of difficulty for algorithms that detect negative obstacles, image sequences were also acquired on 3 course to take advantage of its sequence low hills. For this data collection, lack of time unfortunately prevented the collection of imagery on 2 course, which would have been valuable as a more challenging test of false alarm performance than the flat road sequences; we plan to use 2 course in future evaluations.

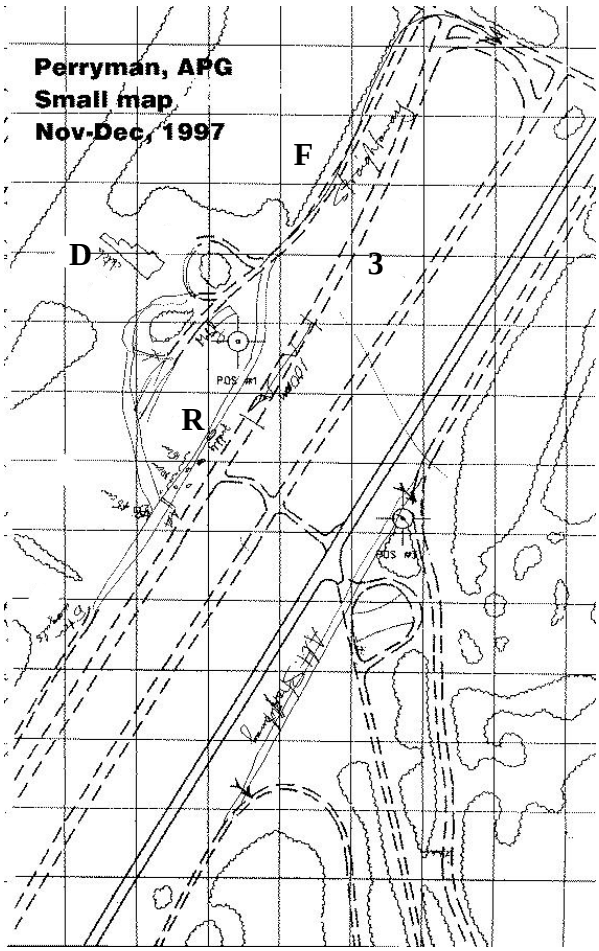


Figure 3: Map of a portion of Perryman facility at APG. Data sets described in this paper were acquired at locations marked "R" (rocks), "D" (ditches), "F" (flat road), and "3" (3-course).

It is well known that the contrast of thermal imagery varies cyclically over a 24-hour period, due to different thermal inertias and moisture contents of materials [8,9]. In particular, rocks tend to be brighter than vegetation during the day, and vice versa at night. Therefore, there are two times of the day when a "thermal crossover" occurs, where temperatures roughly equalize throughout the scene and the image contrast is very low. One of these cross-overs occurs shortly after dusk, the other shortly after dawn. In an effort to capture these difficult conditions, we captured thermal imagery from roughly 5 to 9 pm and 6 to 9 am. The rest of this section summarizes the characteristics of each data set marked in figure 3. Note that stereo imagery was digitized and

resolution requirement of 5 pixels on target, the required FOR greatly exceeds the FOV available from even a 1024x1024 camera if we are to stop for a 24 inch ditch at 20 mph. This will force the vision system design to use multiple cameras or cameras on a pan/tilt, or force the overall performance objectives to settle for detecting larger ditches or to assume that the lateral extent of such obstacles permits a small-angle turn as an avoidance maneuver, with the attendant shorter lookahead range.

All of this discussion leads to the conclusion that, even with a simple "pixels on target" criterion for defining obstacle detectability, the combined IFOV and FOR requirements for detecting obstacles at high speeds are hard to satisfy, especially for negative obstacles. This suggests that there will be a need for cameras on a pan/tilt, together with gaze control logic to decide where to look. It also underscores the importance of experimentally validating estimates for required numbers of pixels on target. Furthermore, it is desirable to derive a more sophisticated detectability model that relates the probability of detection and false alarm to the obstacle size and distance and to the angular and range resolution of the sensor. A very modest start on this was made in [1]. The data collection effort described below takes another step by providing a substantial data set for experimental evaluation of detectability with different sensors placed side by side and with good ground truth for the terrain. We also show initial performance evaluation results that support a "pixels on target" number between 6 and 10. However, more mathematical modeling and experimental evaluation of this issue is in necessasry and will be a subject for our future research.

3.0 Candidate Range Sensors

For autonomous navigation in daylight LADAR and stereo vision have been used for obstacle detection since the 1970's [4,5]. Radar and sonar have been pursued less for cross-country navigation because of their significantly lower angular resolution and other limitations. In the 1980's, LADAR progressed rapidly and was prominent in cross-country mobile robot projects [6], in part because stereo vision was very slow in comparison. Since 1990, however, stereo algorithms have been available that present a viable alternative to LADAR at modest working ranges (order 100 feet) [7]. Currently, engineering considerations do not make either sensor a stand-out winner for cross-country navigation, so the non-emissive nature of stereo vision has made it attractive to military projects where low signature is desirable. At present, stereo also affords a smaller sensor head than LADAR, which is attractive if pan/tilts are required; in addition, stereo still has the lead over LADAR in ability to do non-scanning range imaging, which has advantages for a fast vehicle on rough terrain.

Autonomous navigation at night is a new dimension to the problem. LADAR works at night even better than it does during the day; however, its active nature is still a concern for military applications. Options for doing stereo vision at night include:

- Active illumination, possibly in invisible (near infrared) wavelengths; this is unattractive for signature reasons.
- Use of image intensified CCD cameras; however, the high noise level in such images has discouraged their use.
- Use of thermal infrared cameras (FLIR), with one of the many available thermal imaging technologies.

Current efforts are focusing on the last of these alternatives. The options include cooled detectors in the 3-5 and 8-12 μm bands and uncooled detectors in the 8-12 μm band. Among cooled detectors, indium antimonide (InSb), sensitive in the 3-5 μm band, is attractive because of its extremely high quantum efficiency, availability in medium to large format arrays, and other factors. Two 256x256 InSb thermal cameras with 34 degree FOV optics are available to us as legacy from the Demo II program. Uncooled detectors are of interest because of their potential for much lower cost and smaller size; however, their lower sensitivity, much longer exposure times (30 ms, as opposed to around 1 ms for InSb), and other factors make their utility for robotic vehicle stereo vision unclear at present. A more quantitative comparison of thermal cameras for night stereo vision is in progress; given the availability of the InSb stereo pair and their high performance relative to other options, the effort described below focused on night stereo vision with InSb thermal stereo cameras.

4.0 Data Set Description

To compare various day and night range sensing alternatives, we equipped a HMMWV with two stereo pairs of CCD cameras, the stereo pair of InSb thermal cameras, and a LADAR. One CCD stereo pair had a narrow FOV ("NFOV") of 20x15 degrees, giving a 0.68x0.5 mrad IFOV (HxV) at full digitizer resolution of 512x480 pixels; this was set up to reasonably match the angular resolution requirements arrived at in section 2.0. The second CCD stereo pair had a wide FOV ("WFOV") of 85x64 degrees, or 2.9x2.3 mrad IFOV; this matched the angular resolution in Demo II, which allows a useful comparison to previous performance and lets us evaluate the concept of "dual field of view" --

while the vehicle turns to avoid the obstacle. For very wide obstacles, this is the distance required to execute a 90 degree turn at the given vehicle velocity [2]; this distance can be quite large. For very narrow obstacles that can be avoided by small-angle turns, it can be shown that at high speed it is possible to steer around the obstacle in much less distance than it would take to stop for it; therefore, this is an important maneuver at high speed. Note that the maneuver distance for turning also depends on the vehicle velocity, because the minimum possible turning radius is limited by requiring that the vehicle doesn't slide or tip over in high-speed turns [2]; therefore, vehicle width and center of gravity location also factor into the lookahead distance for turning maneuvers [3].

For the Demo III program, given currently assumed parameter values of 0.65 to 0.85 for the friction coefficient, 0.5 seconds for the combined reaction times, grades from 7% (on-road) to 30% (off-road), and a buffer distance of 8 feet, the lookahead distance for stopping is taken to be about 110 feet at 40 mph, 60 feet at 20 mph, and 25 feet at 10 mph. For steering around obstacles less than 3 feet wide, the lookahead requirement is 65 feet at 40 mph, 35 feet at 20 mph, and 22 feet at 10 mph. These are minimum values that provide a single "look" at the obstacle before a decision to avoid it must be made.

Given obstacle size and lookahead distance, we can estimate the required resolution of the range sensors. Ideally, these estimates should be based on a performance model that relates the size of the obstacle and the lookahead distance to the key design parameters of the sensor. For stereo vision, these parameters are the angular resolution of the camera and the length of the stereo baseline; for a LADAR, these are its angular resolution and its range resolution. Unfortunately, such models for obstacle detectability are not yet well developed, let alone validated experimentally. For expedience, some prior robotic vehicle programs (eg. Demo II) used simpler design criteria based on empirically determined rules for the number of pixels that must be subtended by an obstacle for it to be reliably detectable. Thus, for a rule of thumb that required an obstacle to be at least 5 pixels tall, the obstacle size and lookahead distance could be used to determine a required angular resolution of the camera.

Using this simple "pixels on target" criterion, it can be shown [2] that detecting a positive obstacle requires an angular resolution of

$$\theta_p = H/(NR)$$

where H is the height of the obstacle, N is the number of pixels it subtends, R is the range to the obstacle, and θ_p is the angle subtended by one pixel, or the "instantaneous field of view" (IFOV). For negative obstacles (ditches), foreshortening of the projected width of the ditch on the image leads to a resolution requirement of

$$\theta_n = CW/(NR^2)$$

where C is the height of the camera above the ground and W is the width of the ditch.

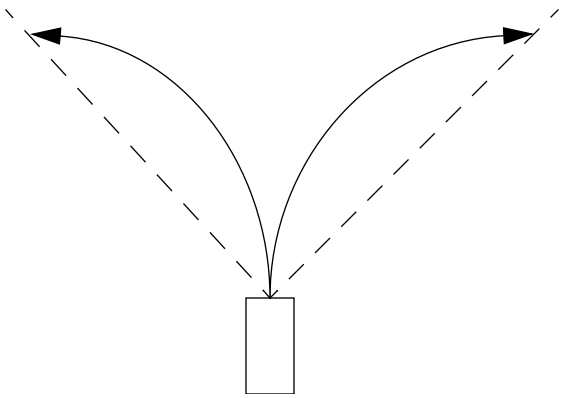


Figure 2: Total field of regard required for "stopping turns" that bring vehicle to a halt along minimum radius turns in either direction.

Taking the obstacle size and lookahead distance numbers from above and assuming detectability with 5 pixels on target, we find that stopping for a 12 inch positive obstacle at 20 mph (60 foot lookahead) requires an angular resolution of about 3.3 mrad/pixel, whereas stopping for the same obstacle at 40 mph (110 foot lookahead) requires 1.8 mrad/pixel. For a 512x512 camera, this implies a field of view (FOV) of around 97 degrees in the first case and 53 degrees in the second case (ie. 512 x IFOV). Stopping for a 24 inch negative obstacle, on the other hand, requires an IFOV of 0.33 mrad/pixel at 20 mph, which translates into an FOV of only 9.7 degrees. If we can execute a small-angle turn around the same obstacle (35 foot lookahead), the IFOV required is about 1.0 mrad/pixel, which permits roughly a 30 degree FOV.

To understand the significance of these numbers, we must relate the FOV's implied by the angular resolution requirement to the total "field of regard" (FOR) necessary to see all of the terrain that the vehicle could steer into before it came to a stop for an obstacle (figure 2). That is, the FOR must include all terrain covered by minimum radius turns in both directions, out as far as the stopping distance along those turns. It can be shown [2,3] that, at 20 mph, the total field of regard required for such "stopping turns" is roughly 75 degrees for the reaction time assumed here. Hence, given an angular

data set is currently being used to evaluate existing obstacle detection algorithms against such variables as obstacle size and distance, vehicle driving speed, and day or night imaging conditions.

The balance of this paper reviews the performance requirements of obstacle detection sensors for Demo III, as currently understood (section 2), describes the range sensors being considered (section 3), describes the data set collected at APG (section 4), and summarizes preliminary obstacle detection results obtained to date (section 5). The results show promise for meeting the performance requirements, but much work remains, both in completing a characterization of the sensors with this data set and in refining the detection algorithms to improve performance. Section 6 summarizes the results to date and comments on directions the future.

2.0 Obstacle Detection Sensor Requirements

Currently, the most robust approach to obstacle detection for off-road navigation is geometric analysis of range imagery. Other methods are conceivable that use intensity, texture, or motion cues from monocular imagery; however, for now these methods are primarily heuristics that could suggest the potential presence of an obstacle without being able to make a definitive judgement. Therefore, our interest here is in estimating the resolution required of imaging range sensors to detect obstacles of a given size at the lookahead distances necessary for the driving speeds desired by the Demo III program. These resolution requirements will determine the feasible fields of view for the camera systems; given the overall angular field that must be observable (the "field of regard"), the combination of feasible field of view and required field of regard will determine whether or not we need one stereo pair on a fixed mounting platform, versus multiple stereo pairs or a pan/tilt. Clearly, these alternatives have very significant impacts on the design of the entire vision and control system.

A process for deriving such requirements was described in [1] and extended in [2] and [3]. Resolution requirements are determined by the size of an obstacle and the lookahead distance at which it must be detected. The minimum terrain feature size that is considered an obstacle is determined by the vehicle size and velocity. The nominal vehicle chassis to be used in the program, shown in figure 1, will have 29 inch tires and a ground clearance of 12 inches; ignoring speed-related variations, the nominal minimal obstacle sizes for this vehicle are 12 inches high for "positive" obstacles (eg. rocks) and 24 inches wide for "negative" obstacles (eg. ditches) [3].

Figure 1: Nominal vehicle for Demo III program

Lookahead distance requirements are determined by the vehicle velocity and reaction time, the coefficient of friction between the tires and the ground, the type of avoidance maneuver being executed (stopping or steering around), and related vehicle dimensions. The lookahead distance (d_l) is composed of a constant buffer distance d_b, a reaction distance d_r, and a maneuver distance d_m that depends on the type of avoidance maneuver. The buffer distance accounts for the distance from the cameras to the nose of the vehicle, plus any desired safety margin in the stopping distance. The reaction distance is the distance the vehicle moves before the obstacle avoidance maneuver is initiated; this is the initial velocity v times the sum of the computing time t_c and any mechanical actuation latency time t_a involved in engaging the brake or changing the steering angle. Nominal values for t_c and t_a are between 0.1 and 0.25 seconds each. Since the vehicle must actually see past an obstacle in order to have enough room to stop for it, the computing latency time is usually doubled in estimating the lookahead distance [1]; greater multipliers may also be applied to allow time for multiple "looks" to suppress false alarms. For a stopping maneuver, the maneuver distance is the distance required to come to a stop once the brake is engaged. On flat ground, this is $v^2/2\mu g$, where v is the initial velocity, μ is the coefficient of friction, and g is the gravitational acceleration; on hills, the μ term is reduced by a function of the slope angle [3]. For dry surfaces, experimental testing yields values of μ between 0.65 and 0.85 [3]. The overall lookahead distance for a stopping maneuver on flat ground is:

$$d_l = d_b + v(2t_c + t_a) + v^2/(2\mu g)$$

For turning maneuvers, the maneuver distance term is replaced with a term that models the distance travelled

Performance Evaluation of UGV Obstacle Detection with CCD/FLIR Stereo Vision and LADAR[1]

Larry Matthies, Todd Litwin, Ken Owens, Art Rankin
Jet Propulsion Laboratory, Pasadena, CA 91109

Karl Murphy, David Coombs, Jim Gilsinn, Tsai Hong,
Steven Legowik, Marilyn Nashman, Billibon Yoshimi
National Institute of Standards and Technology
Gaithersburg, MD 20899

Abstract

The next phase of unmanned ground vehicle (UGV) development sponsored by OSD (the Experimental Unmanned Vehicle, or "Demo III" Program) aims to enable round-the-clock operation with autonomous, cross-country navigation at speeds up to 40 mph on roads, 20 mph by day off road, and 10 mph by night off road. This paper reviews the obstacle sizes that must be detected and lookahead distances that are required to support such driving speeds, then describes data collection and performance evaluation efforts with three different range imaging systems: (1) stereo vision with CCD cameras, (2) stereo vision with InSb FLIR cameras operating in the 3-5 μm band, and (3) the Dornier EBK LADAR,. The LADAR and the FLIR stereo are applicable to day and night operation; the CCD stereo to day operation only. An extensive data set for this performance evaluation was collected at Aberdeen Proving Ground in November and December 1997. This paper describes the data set and preliminary obstacle detection results obtained with it. Ongoing performance evaluation with this data set will guide obstacle detection sensor selection and development for the Demo III Program.

1.0 Introduction

The Experimental Unmanned Vehicle Program, informally known as the Demo III Program, is developing unmanned ground vehicle technology to enable semi-autonomous vehicles to perform a variety of missions. Particular emphasis is being placed on reconnaissance missions for vehicles attached to scout platoons. Such missions will require semi-autonomous navigation with obstacle avoidance both on and off-road.

A key goal of the Demo III program is to advance the technology of autonomous mobility to enable autonomous driving by the end of the program at speeds up to 40 mph on-road, 20 mph off-road in daylight, dry conditions, and 10 mph off-road at night or in wet weather. These capabilities are to be developed and demonstrated in three stages: first, navigation in semi-arid terrain at speeds up to 10 mph cross-country (Demo Alpha, summer of 1999), then navigation in vegetated terrain at speeds up to 20 mph cross-country (Demo Bravo, summer of 2000), and finally navigation covering the full goal of the program (Demo III, summer of 2001).

Sensors for obstacle detection in these conditions will need to perceive the geometry and material type of the terrain and any ground cover during the day or night and in limited adverse weather (ie. in rain). Imaging range sensors will be used to perceive terrain geometry; multispectral imagery, image texture analysis, and other techniques are being explored for material typing. Range sensors currently under evaluation include LADAR, stereo vision with CCD cameras for daylight operation, stereo vision with FLIR cameras for night operation, and radar for conditions with poor visibility.

Previous programs have demonstrated semi-autonomous navigation with these sensors, but at much lower speeds, for larger obstacles and with shorter detection ranges than are required for this program. Therefore, a systematic effort must be undertaken to evaluate the performance of candidate sensors against the defined goals of the program. The first step in such an evaluation was taken in November and December of 1997 in an extensive collection of LADAR and stereo pair imagery at the U.S. Army Aberdeen Proving Grounds (APG). This

1. This work was sponsored by the Joint Robotics Program of the Office of the Under Secretary of Defense for Acquisition and Technology (JRP OUSD(A&T)) and monitored by the U.S. Army Research Laboratory.