Stereo Perception on an Off-Road Vehicle

A. Rieder, B. Southall, G. Salgian, R. Mandelbaum † H. Herman, P. Rander, T. Stentz ‡

Abstract— **This paper presents a vehicle for autonomous off-road navigation built in the framework of DARPA's PerceptOR program. Special emphasis is given to the perception system. A set of three stereo camera pairs provide color and 3D data in a wide field of view (greater 100 degree) at high resolution (2160 by 480 pixel) and high frame rates (5 Hz). This is made possible by integrating a powerful image processing hardware called Acadia. These high data rates require efficient sensor fusion, terrain reconstruction and path planning algorithms. The paper quantifies sensor performance and shows examples of successful obstacle avoidance.**

Keywords— **Stereo Vision, Real-Time Image Processing, Autonomous Off-road Navigation**

I. Introduction

Autonomous Off-Road mobility will be a key component for future combat systems. Autonomous systems must assess mobility hazards reliably, under all weather conditions, day and night, as well as in the presence of dust, smoke, and other air-borne obscurants. Although achieving this goal depends on a series of technical capabilities, off-road obstacle perception is commonly regarded the most critical component. "Without some minimum ability to detect obstacles, successful mobile mission completion will be extremely difficult."[1].

Current perception systems, however, are considered "too immature" [1] to handle the complex scenarios that an autonomous vehicle will face. In fall 2000, DARPA therefore founded the PerceptOR program "for advanced prototype perception systems for unmanned ground vehicles. ... The PerceptOR program is structured around advancing and understanding this critical enabling perception technology as it applies to robotic mobility. These robotic systems will minimize the use of human vision (teleoperation) for obstacle detection."[1].

In a response to the DARPA solicitation the National Robotics Engineering Consortium (an institute of Carnegie Mellon University), Sarnoff Corporation, Rockwell Science Center, Boeing Company and Redzone Robotics formed a consortium to design and built an innovative autonomous mobility system. Under the PerceptOR program, the research emphasis is on the development of new and effective perception and planning systems rather than the design of novel vehicle platforms to be guided by such systems. To this end, two Honda all terrain vehicles (ATV) were retrofitted with speed, steering and direction control; onboard positioning sensors (including GPS); general purpose computing for sensor data processing and autonomous navigation; special purpose computing and cameras for real-time stereo vision; a sensor head on a computer-controlled pan/tilt mount; ladar sensors, and wireless communications to a joystick controller and operator control station (OCS) for remote control and monitoring.

This paper describes our approach to perception for off-road mobility, and is organized as follows. After we provide an overview of related work (section II), section III describes the ATVs, concentrating on their sensors and other key hardware components for the perception process. Section IV outlines the algorithms that are used to process the sensor data, reconstruct the environment, detect driving hazards, control the actuators and avoid obstacles. In section V we quantify sensor performance and show examples of successful obstacle avoidance. Finally, we offer some conclusions and areas for future research (section VI).

II. State of the Art

Autonomous off-road perception and maneuvering has been a topic for research for several years; we will briefly outline some products of this research.

The Mobile Detection, Assessment and Response System (MDARS) [2] was designed and built to detect of intruders around warehouses, supply areas and commercial buildings. The vehicles incorporated a wide variety of sensors, including two different radar systems, a laser scanner and FLIR based stereo imaging system. Most of them were used for surveillance. The cooperation between indoor and outdoor systems was thoroughly investigated. However, outdoor activities took place in a controlled and mostly known environment.

MDARS supported the larger DEMO III program, which also built its off-road vehicle. Perception for mobility was based on stereo and color. The stereo system produced disparity maps at a rate of 6 Hz on a single PowerPC 750 microprocessor with 7x7 pixel correlation window and search range of 40 pixels on images at resolution level 1. (Resolution level 0 corresponds to the full 640x480 pixel image, level 1 corresponds to half size, and so on) [3].

In June 1999 the final demonstration of PRIMUS-C, a program for intelligent, mobile unmanned systems funded by the German DOD, showed a vehicle that maneuvered successfully both on and off-road [4], [5]. Perception for off-road mobility was based on a scanning ladar that supplied range information for 128 by 64 pixel at a resolution of 6 cm and a frequency of 4 Hz.

Another program that combines on- and off-road capabilities is the German/US cooperation AutoNav [6]. It evolved from years of successful on-road activities. Successful off-road maneuvering was repeatedly performed including live demonstration in front of large audiences. However, terrain structure was kept simple and quantitative analysis of the sensor performance has not yet been published.

Another approach originating from on-road activities is

† Sarnoff Corporation, Princeton, New Jersey
‡ National Robotics Engineering Consortium, Carnegie Mellon University, Pittsburgh, Pennsylvania

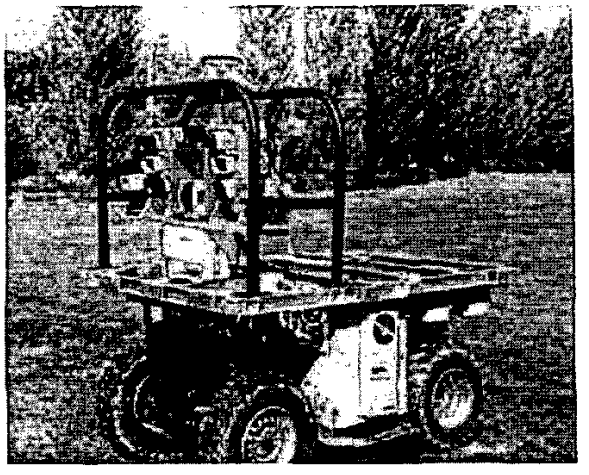


Fig. 1. Retrofitted ATV

given in [7]. This vehicle combines radar, sonar and computer vision to cover different areas ahead of the vehicle. However, the system is intended only to support the driver and has not achieved full autonomy.

III. Retrofitted ATVs

Two Honda all-terrain vehicles (ATV) were modified to comply with the requirements of an autonomous off-road vehicle. One of the vehicles is shown in figure 1.

The handlebars, seat, cargo racks and some of the shrouding was removed. The front cargo rack is replaced by a platform that holds the perception sensors. A ladar sensor and the three stereo camera pairs are mounted in the center using adjustable brackets so that position and angle of each sensor can be easily changed. Since these sensors have limited field of view, their brackets are fixed to a mount that can be tilted electronically to adjust to the current perception needs. This mounting is not intended to stabilize the sensors, but only to increase fields of view.

Computing and additional electronics and power supplies are distributed among the remaining space available on the vehicle, which includes the area behind the sensors, and the space the rider's legs would normally occupy.

All sensors and computers are protected against collision and vehicle roll-over by cages.

A. Sensors

At the current stage of implementation we use two different sensor modalities: ladar and stereo vision. All sensors are mounted in the front of the vehicle. Figure 2 shows the arrangement.

This paper mainly focuses on the stereo sensor. Information about the radar is available in [4].

The top and bottom video cameras on the left hand side of the vehicle are pointing to the right at an angle of 27 degrees. They are separated vertically by a baseline of 23 cm. Color cameras with a 2/3" sensor chip and 8mm lenses provide a field of view of more than 55 degrees. Identical

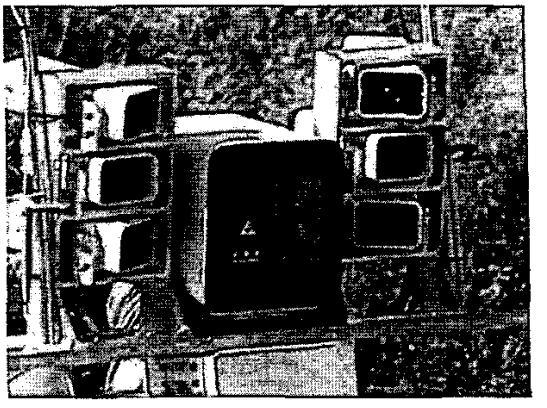


Fig. 2. Three stereo camera pairs and ladar sensor

cameras pointing to the left are used on the right hand side of the vehicle. The fields of view are arranged so that they overlap in the important region just in front of the vehicle and just touch at infinity. In total, we cover a field of view of more than 100 degrees.

The center cameras on each side form the third stereo pair. Black and white cameras with 12mm lenses provide higher resolution for the important area straight ahead of the vehicle. A baseline of 55 cm also results in higher depth resolution compared to the peripheral area covered by the side looking cameras.

All cameras are pro-scan, thus all 720 by 480 pixels are available for stereo computation, free of interlacing effects. Pro-scan also reduces the effects of aliasing, which improves the quality of stereo.

B. Acadia – Vision on a Chip

Each camera image consists of 720 by 480 pixels. Thus a single color image represents 675 kB of information. All six images add up to a total of 3.3 MB. To perform stereo at 10Hz requires that 33 MB of data have to processed at every second, resulting in additional 40 MB of range data (4 byte floating point representation).

These vast amounts of data cannot be handled by today's off-the-shelf processors. Sarnoff Corporation therefore developed custom-manufactured hardware that specializes on image processing tasks such as filtering and correlation. Figure 3 shows the latest generation in this development effort, started in 1984 [9].

Sarnoff's Acadia Vision Chip [10] is integrated together with two NTSC framegrabbers and a Motorola general purpose processor on a board that plugs into any PC. It grabs images from two cameras, processes them and transfers the results via PCI burst to the host PC. Acadia is capable of 80 BOPS. However, the set of operations is limited to some basic image operations:

- lookup tables and simple arithmetic / logic operations
- affine image warping

$$\begin{pmatrix} x \\ y \end{pmatrix}' = \begin{pmatrix} ax + by + c \\ dx + ey + f \end{pmatrix}$$

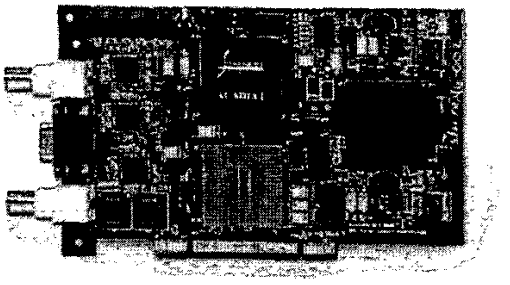


Fig. 3. Acadia – Vision on a Chip

- quadratic image warping

$$\begin{pmatrix} x \\ y \end{pmatrix}' = \begin{pmatrix} a + bx + cx^2 + dy + ey^2 + fxy \\ g + hx + ix^2 + jy + ky^2 + lxy \end{pmatrix}$$

- image correlation and peak finding.
- Gaussian and Laplacian filtering

Three Acadia boards are used to process the individual stereo pairs on each ATV. The results of Acadia's multi-resolution processing are three disparity maps at different resolutions (720x480, 360x240 and 180x120) for each stereo pair. At each level a search range of 32 pixel is searched for correspondences. Note, that this search range at level 2 corresponds to a search range of 128 pixel for the original images. The disparity maps for all levels are then transfered to the PC, where they are combined to a single disparity map with a resolution of 720 by 480. The disparity maps are then converted to range maps at a resolution of 360 by 240 pixels.

For terrain classification based on color, the color images are also transmitted from Acadia to the PC.

Each Acadia processor is capable of processing 720 by 480 images at 30Hz. The bandwidth of the PCI bus and the performance of the general purpose processor that is used to convert disparity to range results limits the performance of the system that combines three Acadia boards. It processes 15 image pairs (720 by 480) per second to produce an output of 2160 by 480 pixels at 5 Hz.

IV. Sensor Processing and Obstacle Avoidance

A. Stereo Processing

Stereo vision is a process of triangulation that determines range from two images taken from different positions. In our case we assume that the optical axes of the two cameras are almost parallel. The image translation is predominantly horizontal for the forward looking stereo pair and predominantly vertical for the side looking ones.

The challenge is to find correspondences in the two camera images. This is usually done by some manner of image correlation and peak finding. In our approach, called horopter stereo [11], we increase the robustness of this step by pre-aligning the left and right image. When the cameras are looking at a flat surface from two arbitrary positions then the right image can be reconstructed by applying a projective transformation on the left image. In the case of parallel optical axes, this projective is indeed an affine transformation.

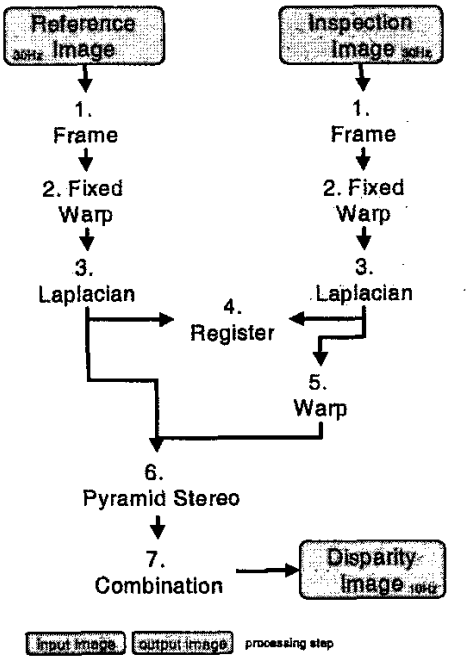


Fig. 4. Disparity computation on Acadia

Based on this, our algorithm works as shown in figure 4. In a first step we apply a quadratic warp to the right image. This is an approximation of the projective warp and takes care of vergence, rotation and translations between the two cameras. The parameters for this operation are determined by calibration. Since we can assume the optical axes to be almost parallel, we know that the projective is indeed close to an affine. Therefore the deviation of the quadratic approximation from the true projective is typically less than two pixels even at the borders of the image.

Now that the images are aligned, the search for correspondences can be restricted to horizontal image lines. However, changes in the vehicles position relative to the ground and the shape of the ground would still require a rather large search region. Therefore in a second step, we dynamically align the resulting left and right image. Since we have already compensated all other effects, an affine transformation is sufficient here.

Once the images are aligned like this, correspondences are computed. This is done at the original image resolution of 720 by 480, but also at lower resolutions of 360 by 240 and 180 by 120. Results at lower resolution are generally more robust to false matches that can occur in regions of low texture. At lower resolutions the search range is bigger, although results are commensurately less accurate.

The resulting three disparity maps at different resolutions are combined to a single one before range information is extracted. Here, the vergence angle α between the two

cameras has to be taken into account. Given the horizontal coordinates of corresponding pixels x_l, x_r in the left and right image the range z can be expressed as

$$z = \frac{bf}{d}, \text{ where } d = x_l - \frac{f \cdot (x_r - \tan(\alpha))}{f + x_r \tan(\alpha)}$$

Here b denotes the baseline and f the focal length in pixels, and d is the inter-image disparity.

B. Navigation

Effective autonomous mobility requires processing and models that span multiple scales of time and space. Traditional vehicle control represents the shortest time scale. It concerns itself with moving actuators to commanded positions and with following trajectories over the relatively short term. Our approach to autonomous navigation can be described in terms of two layers that exist above this layer: local navigation to safeguard the vehicle and global navigation to achieve mission objectives.

B.1 Local Navigation

Local navigation directs the control layer by considering the higher-level issues of obstacle avoidance and maintaining safe vehicle postures on the terrain. To do so, it must understand the capacity of the vehicle to move and its static and dynamic stability characteristics. Also, it must represent the environment on a scale capable of resolving the smallest terrain feature that can present a hazard and it must reason on the time scale of the vehicle's reactions.

We incorporated the Ranger navigation system [12] to address the local navigation need. Ranger is a powerful navigation system that considers a variety of factors in planning vehicle motion, and is built to allow easy incorporation of new sensors and new evaluation metrics. Ranger runs as a real-time module performing a simple loop of receiving any available sensor data, incorporating that data into the terrain model, evaluating the terrain for discrete obstacles, and then performing a forward simulation of the vehicle traversing the terrain across a set of possible arcs. Each arc is scored by the relative safety it provides the vehicle, considering issues of rollover, collision with terrain, high centering, and the confidence of the terrain that would be traversed.

If all arcs are shown to be unsafe, Ranger can bring the vehicle to a halt. If the final motion before stopping the vehicle still reveals no safe path, Ranger can back up, reversing the direction along the arcs already traversed, until a clear path forward is found. Ranger could also alert a remote operator.

B.2 Global Navigation

Global navigation directs the local layer by considering the higher-level issues of achieving waypoint goals and doing so in a manner that optimizes an objective function that characterizes different degrees of mission success. Myriad issues such as vehicle coordination, time elapsed, distance traveled, fuel use, visibility, threat proximity etc. must be managed. This layer must represent the environment on a spatial scale and reason on a time scale sufficient to resolve changes in the objective function while respecting practical limits on computing resources. During Phase I, we incorporated the D* [13] path planning system to address the global navigation need. D* is an efficient, high-speed planning system capable of planning missions with the entire spectrum of potential pre-mission information, from knowing the complete terrain model at fine detail to knowing nothing about the terrain. D* also supports in-transit updates to its terrain model, allowing the global planning system to incorporate new information as it becomes available.

V. Experimental Results

For the appropriate use of any sensor data, it is vital to understand the sensor's performance when measuring a particular quantity of interest. For the autonomous navigation task, it is necessary for us to understand the minimum range at which the stereo vision system can reliably detect both positive and negative obstacles; the minimum detection range will determine the types of maneuver that the vehicle can perform between successive observations of its environment.

In the following we present results from a quantification of the stereo vision system's measurement of range to positive obstacles, an assessment of the detection of negative obstacles, and finally an example of autonomous navigation on off-road terrain.

A. Evaluation of stereo vision

A vehicle must be able to detect and avoid hazards in its surrounding environment in order to navigate successfully. For unmanned ground vehicles, driving hazards can be placed into two major groups: positive obstacles such as trees, tree stumps and rocks, and negative obstacles, such as ditches and pot-holes.

A.1 Positive obstacles

For a quantitative evaluation of stereo performance, we used two objects, a tree stump and a board, as illustrated in figure 5. The tree stump is 25 cm tall and has a diameter of 15 cm; it was chosen as an example of an obstacle that might be found in grassy or wooded off-road environments. The board, 52 cm wide and 30 cm tall, is flat. For a perfect sensor, the average range over the surface of the board will correspond to the range to the board's center; thus we de-couple errors owing to sensor performance from any ambiguity arising from target shape.

During the experiment, the two targets were moved together in front of the stereo sensor at 16 different distances between 2.7m and 16.2m in 0.9m increments. At each step, the distance between the two objects was 37 cm.

For each distance, the depth map recovered from stereo was manually segmented to select regions corresponding to the two objects.

Depth measurements for points inside the two regions were used to compute average distance to that object. Fig-

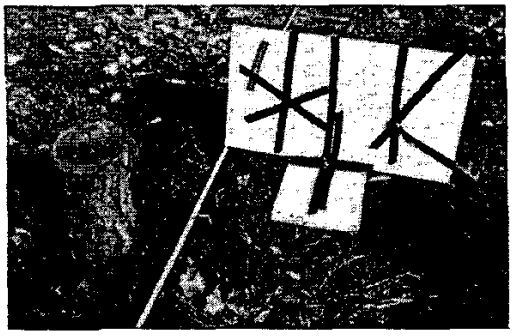


Fig. 5. Objects used for measuring stereo performance on positive obstacles

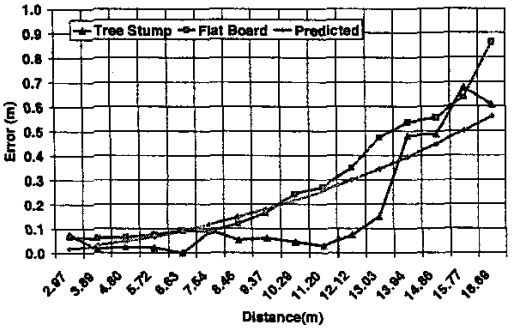


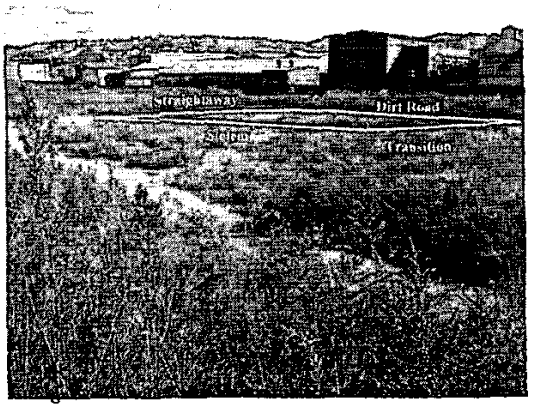
Fig. 6. Depth measurement errors for the two objects

ure 6 shows a plot of the absolute error between recovered and ground truth positions for each of the two objects (red triangles for the tree stump and pink squares for the flat board). As expected, the errors increase with distance to target. The theoretical depth error owing to a one pixel disparity error is also shown (green diamonds); the measured error values follow a similar trend.

Note that well-differentiated depth measurements for the two individual objects are only possible for distances up to about 10m, at which point the sum of the absolute error to each target approaches the inter-target separation. In addition to the measurement noise, another factor contributing to the decrease of depth resolution with distance is the fact that the image area occupied by an object decreases quadratically with distance. This means fewer pixels on target (in the example above, the number of pixels used to compute average depth on the flat board varied between 3600 at 2.7m and 96 at 18 m). If the image area occupied by an object is small, it is likely that the integration window used in the stereo algorithm will cover not just the object but also background points, which would generate incorrect disparity values.

A.2 Negative Obstacles

For testing the stereo performance on negative obstacles, we used a hole with dimensions 0.6 x 1.2 m and 0.45 m deep. In addition to this hole, the scene imaged also contained a large positive obstacle, a mound of soil.



The ATV was moved at various distances from the hole, between 5.5 m and 14.5 m, observing the hole from a diagonal direction. The top row of figure 7 shows reference images at 9, 10.5, 12 and 14.5 m, respectively. The bottom row shows a top-down view of a false-colored wire-frame rendering of the 3D shape recovered from stereo for the area in the green rectangle in the image above. The ground is colored in blue, with warmer colors corresponding to points higher above the ground. The area inside the green rectangle contains the hole and the mound of soil (a positive obstacle) to the left of it. Both are visible in the recovered 3D shape from 9m and 10.5m (two left-most images in Figure 7): the hole corresponds to the distinct dark blue patch in the bottom-right part of the wire-frame image. The mound of soil corresponds to the red patch in the upper left region.

At 12 m (third column from left in Figure 7) the hole is barely noticeable (the region corresponding to the hole in the wire-frame image is not distinctly differentiated from the surrounding area), and at 14.5m (rightmost column) it is not visible.

B. Obstacle Avoidance

We integrated local and global navigation based on formulating a cost function to be optimized by searching over a number of candidate trajectories. Candidate local trajectories extending somewhat beyond the vehicle reaction distance are joined to an equal number of global trajectories that continue to the next waypoint in the global motion plan. Naturally, the costs associated with very unsafe terrain traversal are high. When several alternative trajectories are safe, the net effect is to choose the one which best optimizes the objective function. On barren terrain, the objective function may change little over the field of view the local navigation system. However, on rough terrain, such as an area with a canyon, the global navigator will direct the system out of trouble.

Consider figure 8, which shows a path at the Buncher site near the NREC. A closed trajectory consisting of 3 waypoints at the corners of a triangle was specified to the D* algorithm as the mission. The length of the loop is approximately 146 meters and a single circuit is typically

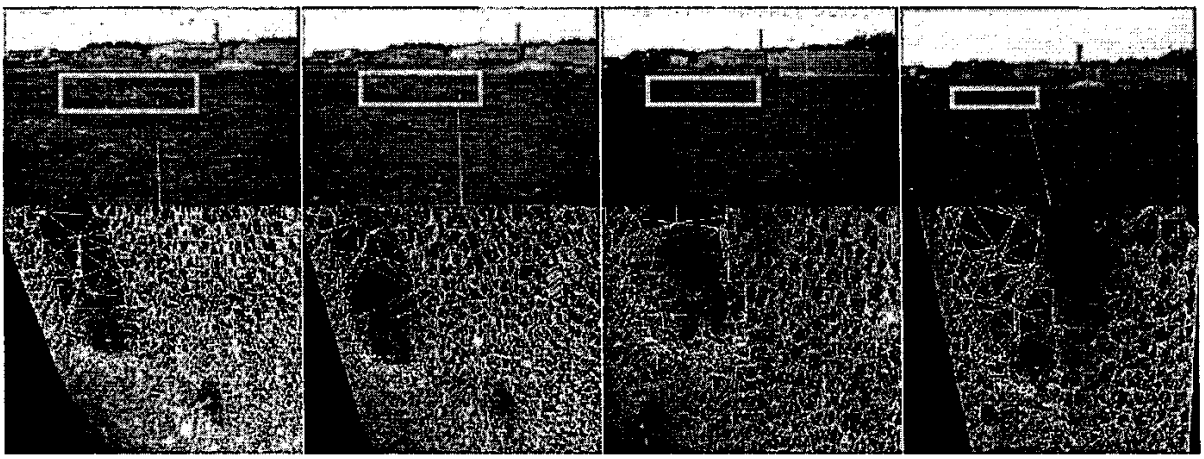


Fig. 7. Negative Obstacles at different distances (from left to right: 9, 10.5, 12 and 14.5 m respectively). Top row: reference image; Bottom row: top-down view of a false-colored (red = deeper) wire-frame rendering of the 3D shape recovered from stereo for the area corresponding to the green rectangle above.

completed in 5 minutes. Average vehicle speed is therefore 0.5 meters/sec. There are 3 discrete obstacles peppered along the "slalom" part of the path and waypoints are chosen to force an encounter with each of them. The "transition" is a short section connecting the slalom to the dirt road. The dirt road itself was too narrow to introduce obstacles but the system correctly identifies and avoids the vegetation on both sides of the road, and follows the road, based only on perception data and the location of the waypoint at the end of the road.

VI. Conclusions

The unmanned ground vehicle presented in this paper uses two different sensor modalities, namely stereo and ladar, for autonomous navigation in off-road scenarios. High resolution in a wide field of view allows the vehicle to perceive obstacles early and effectively plan maneuvers to avoid them.

Although basic perception and maneuvering capabilities have already been shown, the system cannot yet tackle all possible outdoor scenarios. To approach this goal the sensor data has to become more reliable. For the individual sensors, time integration based on frame to frame alignment, another of Acadia capabilities, will help to improve signal to noise ratio. Even more important is the integration of different sensing modalities such as IR or radar. These sensors will be closely tied together by data fusion making optimal use of each sensor's strengths and compensating their individual weaknesses.

In parallel, efforts will be devoted to improve vehicle speed. This will only be possible if the sensor data can be processed at higher rates. We are therefore pushing more and more sensor processing steps onto Acadia, which is best suited to do these tasks most efficiently. This approach reduces the required bandwidth for the PCI bus. It also frees resources on the general purpose processors which can be used to improve terrain classification and path and mission planning - tasks that are less suitable for specialized hardware.

References

[1] DARPA: *Program Solicitation Perception for Off-Road Mobility.* DARPA PS01-02, Washington, DC, 2000

[2] R.S. Inderieden, et al.: *Overview of the Mobile Detection Assessment and Response System.* In Proceedings of the DND/CSA Robotics and KBS Workshop, St. Hubert, Quebec, October 1995.

[3] P. Belluta, R. Manduchi, L. Matthies, K. Owens and A. Rankin *Terrain Perception for DEMO III.* In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Dearborn, Michigan, U.S.A., October 2000.

[4] I. Schwartz: *PRIMUS Realization aspects of an autonomous unmanned robot.* In Proceedings of SPIE, Enhanced and Synthetic Vision, 1998

[5] I. Schwartz: *PRIMUS: autonomous driving robot for military applications.* Proc. SPIE Vol. 4024, p. 313-323, Unmanned Ground Vehicle Technology II, 2000

[6] K.-H. Siedersberger et aliter: *Combining EMS-Vision and Horopter Stereo for Obstacle Avoidance of Autonomous Vehicles.* In Proceedings of the IEEE Workshop on Intelligent Computer Vision Systems (ICVS), Vancouver, Canada, 2001

[7] Ka C. Cheok, G.E. Smid, D.J. McCune: *A Multisensor-Based Collision Avoidance System with Application to a Military HMMWV* In Proceedings of IEEE Intelligent Transportation Systems, Dearborn, Mi, October 2000

[8] T. A. Heath-Pastore, H. R. Everett: *Coordinated Control of Interior and Exterior Autonomous Platforms.* Fifth Int. Symposium on Robotics and Manufacturing, Maui, HI, August 1994.

[9] G. van der Wal, J.O. Sinniger: *Real time pyramid transform architecture.*, In Proceedings of SPIE, Cambridge, MA, September 1985

[10] G. van der Wal, M. Hanson, M. Piacentino: *The Acadia Vision Processor.*, IEEE Int. Workshop on Computer Architecture for Machine Perception, Italy 2000

[11] P. Burt, L. Wixson, G. Salgian: *Electronically Directed 'Focal' Stereo.* In Proceedings of the International Conference on Computer Vision, 1995.

[12] A Kelly and A Stentz: *An Approach to Rough Terrain Autonomous Mobility*, Int. Conf. Mobile Planetary Robots, 1997.

[13] A Stentz: *Optimal and Efficient Path Planning for Partially Known Environments.* In Proc. IEEE Int. Conf. on Robotics and Automation (ICRA), 1994.