Obstacle Detection for Unmanned Ground Vehicles: A Progress Report *

Larry Matthies, Alonzo Kelly, Todd Litwin, and Greg Tharp
Jet Propulsion Laboratory - California Institute of Technology
Mail stop 107-102, 4800 Oak Grove Drive, Pasadena, California 91109
Phone 818-354-3722, Fax 818-354-8172, larry.matthies@jpl.nasa.gov

Abstract

To detect obstacles during off-road autonomous navigation, unmanned ground vehicles (UGV's) must sense terrain geometry and composition (ie. terrain type) under day, night, and low-visibility conditions. To sense terrain geometry, we have developed a real-time stereo vision system that uses a Datacube MV-200 and a 68040 CPU board to produce 256×240-pixel range images in about 0.6 seconds/frame. To sense terrain type, we are using the same computing hardware with red and near infrared imagery to classify 256×240-pixel frames into vegetation and non-vegetation regions at a rate of five to ten frames/second. This paper reviews the rationale behind the choice of these sensors, describes their recent evolution and on-going development, and summarizes their use in demonstrations of autonomous UGV navigation over the past five years. This work has been the first to show that stereo vision can be practical for autonomous UGV navigation, and is now the first to show a real-time terrain classification system with very low computing requirements.

1 Introduction

Unmanned ground vehicles are being developed for a variety of applications in the military, in hazardous waste remediation, and in planetary exploration. Such applications often involve limitations on communications that require UGV's to navigate autonomously for extended distances and extended periods of time. Under these circumstances, UGV's must be equipped with sensors for detecting obstacles in their path. This paper provides a progress report on work being done at the Jet Propulsion Laboratory (JPL) on obstacle detection sensors for the "Demo II" UGV Program, which is sponsored by the Advanced Research Projects Agency (ARPA) and the Office of the Secretary of Defense (OSD).

The goal of the Demo II Program is to develop technology enabling UGV's to perform autonomous scouting missions. In its full generality, this application will require operating during the day or night, in clear or obscured weather conditions, and over terrain that will include a variety of natural and man-made obstacles. Obstacle detection sensors for the full problem must be able to perceive the terrain geometry, perceive the material type of any ground cover (ie. terrain type), and do so at night and through haze.

We have been addressing the problem of perceiving terrain geometry (by day and by night) and we are beginning to address the problem of perceiving terrain type (by day). For sensing geometry, we are using stereo vision, because its properties of being non-emissive, non-scanning, and non-mechanical make it attractive for military vehicles that require a low signature sensor and well-registered range data while jostling over rough terrain. In collaboration with other labs, we have recently begun experimenting with stereo vision on thermal imagery to address night operations. For sensing terrain type, we have done preliminary investigations of discriminating soil, vegetation, and water using visible, near infrared, and polarization imagery.

Section 2 reviews the current status of the real-time stereo vision system we have developed for UGV applications. The system currently produces range imagery from a 256×45 pixel window of attention in about 0.6 seconds/frame, using a Datacube MV-200 image processing board and a 68040 CPU board as the computing engines. This vision system is installed on a roboticized HMMWV[1] that serves as a testbed UGV. Section 3 describes a number of enhancements currently in progress on this system, including recent tests with thermal imagery, simple algorithms for real-time obstacle detection, methods to support focus of attention, and approaches to terrain classification. Section 4 reviews how the vision system has evolved through three major demonstrations over the last five years and shows results from an autonomous navigation trial conducted with our HMMWV on a dirt road near the laboratory. These demonstrations have driven HMMWV's at speeds in the neighborhood of 5 to 10 kph over gentle, but not barren cross-country terrain.

This work has been the first to show that stereo vision can provide range data of sufficient quality, at sufficient speeds, and with sufficiently small computing resources to be practical for UGV navigation. Future work will attempt to increase the quality of the range data, miniaturize the computing system, and integrate terrain classification with range imaging.

2 The JPL Stereo Vision System

Previous versions of JPL's real-time stereo vision system have been described in [1, 2]. Here, we will outline the current version of the algorithm, then discuss how and why it has changed. Current steps in the algorithm are:

1. Digitize fields of the stereo image pairs.
2. Rectify the fields.
3. Compute image pyramids by a difference-of-Gaussian image pyramid transformation.
4. Measure image similarity by computing the sum-squared-difference (SSD) for 7×7 windows over a fixed disparity search range.
5. Estimate disparity by finding the SSD minimum independently for each pixel.

* The work described in this paper was sponsored by the Advanced Research Projects Agency.

[1] High Mobility Multipurpose Wheeled Vehicle - the modern military jeep.

6. Filter out bad matches using the left-right-line-of-sight (LRLOS) consistency check [3, 4].

7. Estimate sub-pixel disparity by fitting parabolas to the three SSD values surrounding the SSD minimum and taking the disparity estimate to be the minimum of the parabola.

8. Smooth the disparity map with a 3×3 low-pass filter to reduce noise and artifacts from the sub-pixel estimation process.

9. Filter out small regions (likely bad matches) by applying a blob filter that uses a threshold on the disparity gradient as the connectivity criterion.

10. Triangulate to produce the X-Y-Z coordinates at each pixel and transform to the vehicle coordinate frame.

11. Detect "positive" obstacles[2] by thresholding the output of a simple slope operator applied to the range image.

Since the first version, this algorithm has evolved as follows. The original version digitized full frames instead of fields (step 1), because it was first used on a Mars rover prototype vehicle that stopped to acquire imagery. At the time, the remainder of the algorithm consisted of steps 3, 4, 5, 7, 8, and 10, plus a Bayesian posterior probability measure that was applied in place of step 6 above to filter out bad matches. Matching was done at only at the 64×60-pixel level of resolution in the image pyramid. Changes to the system and the rationale behind them are described below.

Digitizing fields. Since the vehicle now drives continuously, fields are digitized instead of frames to avoid temporal misregistration from the field-interlaced cameras used on the vehicle.

Rectification. With the 64×60 resolution and 30 degree field of view (FOV) of the original system, it was possible to align the cameras well enough for stereo matching by mounting the cameras on mechanically adjustable brackets, although the alignment procedure was painstaking. Since then, we have moved to matching at the 256×240 level of resolution and have used FOV's as wide as 85 degrees. These changes produce sufficiently high angular resolution and radial distortion to make mechanical alignment impractical. Rectification currently works well enough that the residual vertical disparity at 256×240 resolution is less than one pixel. Since the adjustable camera mounts had potential to go out of alignment, this approach is more robust, as well as much easier to use.

LRLOS consistency check. This procedure ensures that disparities obtained by choosing best matches along left camera lines of sight agree with those obtained by choosing matches along right camera lines of sight. We have found that this procedure runs faster than the posterior probability measure and works better, especially at occluding boundaries.

[2]Positive obstacles are those that extend upward from the nominal ground plane, like rocks, bushes, and fenceposts. "Negative" obstacles extend downward, like potholes, man-made ditches, and natural ravines.

Blob filter. Small islands of bad matches occasionally survive the LRLOS test. These can be contiguous with larger regions of good matches or completely disjoint from good regions. A simple blob coloring algorithm [5] can be used to reject small disjoint regions. Furthermore, it is possible to eliminate small regions of bad matches that are contiguous to larger, good regions by making the connectivity criterion in the blob coloring algorithm be a threshold on the disparity gradient.

Positive obstacle detection. As discussed elsewhere [6], obstacles can be difficult to define and expensive to compute. To enable real-time operation, we initially implemented an algorithm that detected only positive obstacles using a simple slope estimation algorithm applied to columns of the range image [2]. This algorithm assumes that obstacles are vertical step displacements of minimum height H on an otherwise flat ground plane. On each scanline, we project a line of sight out to intersect the ground plane, assume an obstacle of height H at that point, and determine on what scanline the top of the obstacle would appear in the image. A table of vertical scanline offsets is determined in this way for every pixel in the image. To detect obstacles, for every pixel p_1 in the range image, we use it and its higher partner p_2 in the same column to compute the change in height at that point in the image; if the change in height exceeds a threshold, the entire interval of the range image from p_1 to p_2 is declared to contain an obstacle. Although this algorithm is very simple, it is reasonably robust and very fast.

3 Ongoing Algorithm Extensions

A number of extensions to the basic stereo system are in progress. These have the goals of enabling night driving by using forward looking infrared (FLIR) cameras, detecting negative obstacles, increasing speed by using windows of attention and dual FOV camera systems, providing software exposure control, increasing the resolution by moving to full frame cameras, and augmenting obstacle detection by doing terrain classifcation as well as range imaging. These are listed in order from the most simple to the most complex. Initial implementations and results have been obtained for several of these functions; full frame imaging is not yet integrated into the real-time system. The current status for each issue is summarized below.

FLIR imagery. UGV's must be able to perform missions at night. Stereo vision with thermal imagery obtained from FLIR cameras is a possible way to detect obstacles at night. The Demo II Program has acquired several FLIR cameras operating in the 3 to 5 μm wavelength range with 256×256 detectors. In collaboration with SRI International, a pair of these cameras were mounted on the JPL HMMWV and used for a preliminary evaluation of real-time range imaging. The cameras were mounted with a 30 cm baseline and had a field of view of approximately 30 degrees; this compares with 35 cm and 57 degrees for visible imagery. Qualitatively, results at 128×120 resolution compare quite well to results with visible imagery. Results at 256×240 resolution are much sparser, for reasons yet to be determined. These initial results are encouraging. An important next step will be to characterize the quality of the range data quantitatively, in terms of the thermal sensitivity of the cameras and temperature variations of the scene, at various times of the day and night.

Negative obstacles. For cross country navigation, UGV's must detect many types of obstacles. To extend the capability of the basic system while staying within the power of the computing system, we are developing a simple algorithm for detecting negative obstacles. This algorithm is an extension of the column-oriented algorithm for positive obstacles. For each pixel in the range image, the algorithm projects a line of sight out to the ground plane and computes what would be the distance to the next range pixel above it in the same column, if that pixel were on the same ground plane. If the actual distance to the second pixel is greater than expected (with some allowance for noise), then by the geometry of the camera it must also be lower. This indicates a gap in the range data, a negative slope of the terrain, and a possible negative obstacle (eg. ravine) between the two pixels; in this case, the first pixel is labelled as a potential negative obstacle. Experimental results show that these very simple algorithms do a sufficiently good job of detecting these classes of obstacles to enable credible UGV demonstrations in these environments.

Window of attention and dual FOV. It is widely appreciated that focus of attention mechanisms may increase the performance of an autonomous system by reducing the amount of computation that must be done. UGV navigation makes this issue concrete, because the task, the computing limitations, and the criteria for focusing attention all can be well defined. Vehicle speed and reaction times determine the part of the image that must be examined next [7]; essentially, the faster the vehicle goes, the higher up the image it must look and the higher the level of resolution it must use. Here, we only note that two mechanisms are being developed to allow the stereo system to focus attention. The first is the ability to select a rectangular subregion of the image for processing; the second is that ability to switch between alternate stereo camera pairs for input ("dual FOV"). Because of restrictions in the Datacube architecture, windows of attention cannot be changed arbitrarily at run-time. However, a moderate number of fixed windows can be defined at initialization time and selected at frame rates at run-time; these can be at any level of resolution and more than one window can be processed per input image pair. The dual FOV capability involves relatively straightforward switching of inputs and calibration data at frame rate; for example, this enables changing from a wide FOV camera pair to a narrow FOV camera pair as vehicle speed increases.

Software exposure control. Even with auto-iris lenses, obtaining satisfactory dynamic range on the desired part of the image is always difficult, due to changing cloud cover, changing albedo of the terrain, and touchy adjustments on the lenses. We have implemented functions in the MV-200 that will be used in the near future to attempt to control exposure in software. Computational elements within the MV-200 memory modules do most of the work needed to compute the mean, min, max, and standard deviation of intensity over a pre-defined window of attention. We expect to filter these quantities over time and use them to drive lens aperture and/or image exposure time. Experiments with software exposure control will be conducted in the summer of 1995.

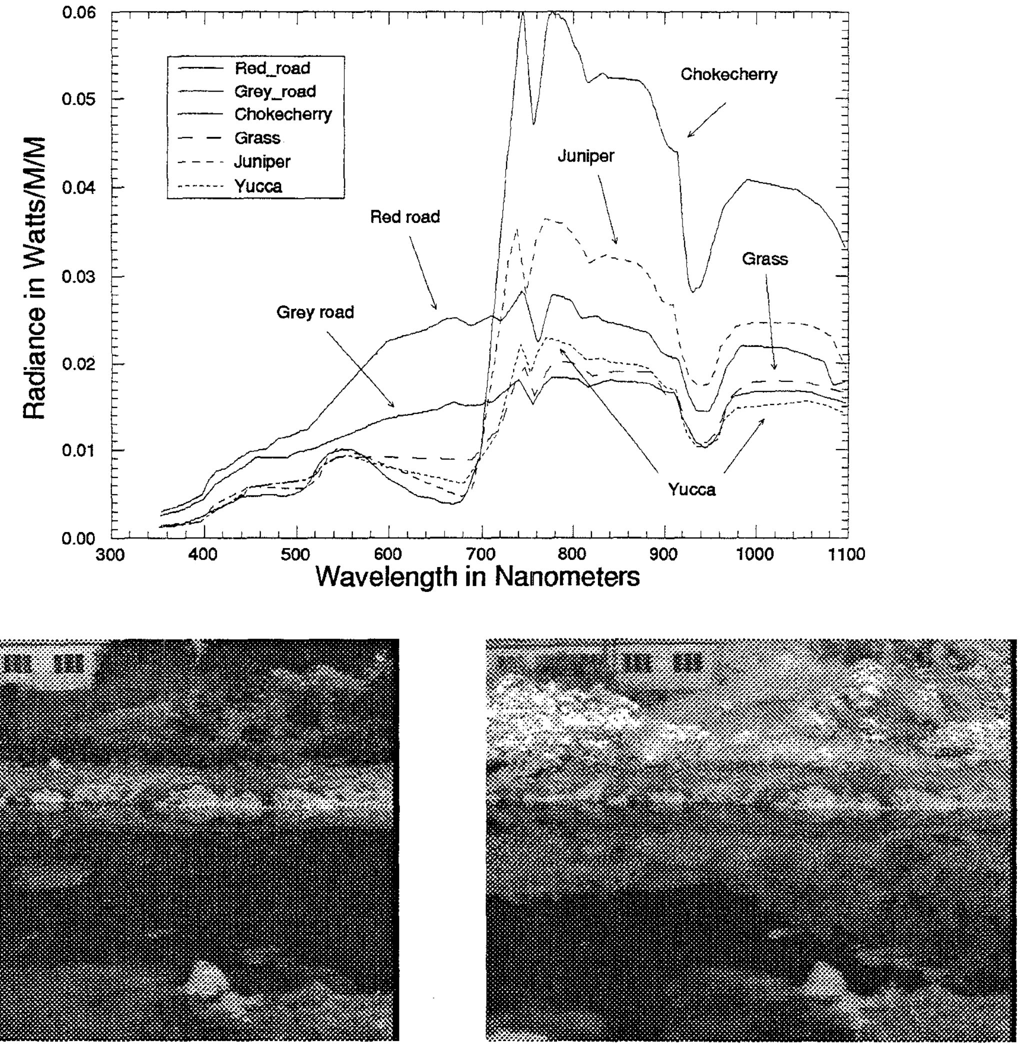
Full frame, non-interlaced range imaging. One of the least satisfactory characteristics of the stereo system is that it must discard half the available vertical resolution, due to the interlaced scanning of the cameras. However, cameras with full frame, non-interlaced scanning are now available (eg. the Pulnix TM-9700). The stereo software can be configured easily to process this imagery; however, three potential problems lurk:

- there will be more demanding requirements for image alignment because of the higher angular resolution of the images;
- there may be more matching ambiguity because of wider disparity ranges;
- there may be slower run-times because of the higher resolution.

We expect to address the first issue, if necessary, through improved calibration. One approach to the remaining issues is to pre-warp the imagery to produce zero disparity for the nominal ground plane; provided that no part of the scene deviates substantially from that plane, this technique reduces ambiguity and run-time by reducing the necessary disparity search range. Unfortunately, this is a fragile assumption, which will often be violated in cross-country navigation. However, by also matching at a lower resolution over a wider search range, we expect that a multiresolution algorithm can be developed that will achieve the strengths of pre-warping with only slightly greater cost. Thus, the real-time constraints of UGV navigation are likely to lead to new multiresolution stereo algorithms, because they will address different requirements than the off-line algorithms developed in the past.

Terrain classification. Clearly, geometry is not the only scene property important for obstacle detection; material or terrain type is also important. A great deal of research has been done on classifying and segmenting images based on color, texture, and other features. As with stereo, however, real-time performance is essential for terrain classification. Therefore, we looked for non-traditional image features that might improve the robustness of classification with very simple, low-level algorithms. Two such possibilities are the use of near infrared imagery to distinguish between soil and vegetation and the use of polarization to detect highlights on standing bodies of water. In this paper, we will briefly illustrate the potential of near infrared imagery.

It is well known in the remote sensing literature that live vegetation is very reflective in the near infrared [8]. During the ARPA Autonomous Land Vehicle (ALV) Program in the mid-1980's, the U.S. Army Engineer Topographic Laboratory took spectral reflectance measurements, at ground level, from several different types of soil and ground cover. Figure 1 shows such spectra for roads of grey and red dirt and for four types of vegetation. In general, we see that soil gets gradually brighter from blue through near infrared. Vegetation is dark in blue and red, somewhat brighter in green, and anywhere from a little brighter to dramatically brighter in near infrared. CCD cameras are sensitive out to about 1100 nm, so in principle this entire spectral range can be sensed by using appropriate filters with existing cameras. Figure 1 makes this point further by showing



Red band (650 nm)

Near infrared band (800 nm)

Figure 1: Top: spectral radiance for six terrain types, showing the generally brighter response of vegetation in the near infrared. Bottom: red and near infrared images of a scene containing a pond in the foreground and various soils, bushes, trees, and buildings in the background.

ground level imagery taken with a CCD camera using red and near infrared filters. These images illustrate the contrast reversal shown by the soil and vegetation between the red and the near infrared band. We have used red, near infrared, and polarization imagery of this scene to obtain good discrimination between vegetation, water, and "other" pixels, using very simple, pixel-wise ratios of the bands. We are now using a filter wheel camera and the MV-200 board to implement a real-time classification system that uses the red and near infrared bands to label pixels as vegetation or non-vegetation at five to ten frames per second.

4 System Integration and Demonstrations

The stereo vision system described in section 2 has been integrated into robotic vehicles and used in the following configurations and demonstrations:

- On the Mars rover tesbed vehicle "Robby", running at 64×60 resolution in several seconds per frame, for autonomous navigation over 100 m of sandy terrain interspersed with bushes and mounds of dirt (Sept'90) [1]. In this demonstration, the vehicle stopped at each frame.

- On the JPL HMMWV, running at 64×60 resolution in ≈ 1.5 seconds/frame, for autonomous obstacle detection and halting during the U.S. Army "Demo I" exercise at Aberdeen Proving Ground (April'92). The HMMWV drove continuously at $\approx$ 1 to 3 m/sec ($\approx$ 5 to 10 kph) while detecting half-meter obstacles in time to stop.

- On a Lockheed Martin HMMWV, integrated with the "Ganesha" and "Ranger" obstacle mapping and path planning systems developed at CMU [9, 10], running at 128×120 resolution in under one second per frame, for autonomous navigation on flat terrain with moderately tall grass and positive obstacles $\geq$ 1.0 m high ("Demo B", June'94). The vehicle velocity in this case was 2 m/sec (7.2 kph).

We expect that the system will operate on a Lockheed Martin HMMWV at "Demo C" in July'95, running a 256×45 window of attention in about 0.5 seconds per frame. Other parameters of the demo remain to be determined. In preparation for this demo, the most recent versions of the stereo vision system and the Ranger path planning system were integrated on the JPL HMMWV and field tested in December'94. Figure 2 shows results from a test run that travelled about 200 m down a dirt road. The lower left side of the figure shows 256×45 windows of attention from eight positions along the run; the lower right side of the figure shows the corresponding range images. A composite elevation map of the entire run is shown in the middle; a wireframe rendering of the composite map is shown in the upper right. The vehicle was instructed to follow its current heading whenever it could, but to swerve to a new heading as necessary to avoid obstacles. Two significant swerves are seen in this run. The first was to avoid a tree on the right side of the road; the second was to avoid a bush on the left side of the road, just beyond the tree. This was a very successful run, and typical of results obtained in this kind of terrain.

5 Summary and Discussion

In this paper, we reviewed the current state of development of JPL's real-time stereo vision system, described ongoing development of new capabilities, and summarized significant field demonstrations that have used this system to perform autonomous navigation. The system currently produces range data from 256×45 windows of attention in approximately 0.6 seconds/frame. Field demonstrations using this system have driven HMMWV's at speeds on the order of 5 to 10 kph over relatively flat, cross-country terrain covered with calf-high grass and sprinkled with positive obstacles 0.5 m high and larger. Ongoing development will extend the system to process imagery at 512×480 resolution using full frame, non-interlaced cameras and a variety of techniques to mitigate the computational burden. For night operations, experiments with stereo pairs of FLIR cameras have shown good range estimation results with imagery processed at 128×120 resolution. For sensing terrain type in addition to terrain geometry, the use of red and near infrared imagery has enabled us to distinguish vegetation from non-vegetation with very simple, robust algorithms that run at five to ten frames per second on the same computing hardware as the stereo vision algorithms (ie. Datacube MV-200).

This work has been undertaken with a pragmatic view toward demonstrating reliable performance with simple, fast algorithms. This strategy has paid off: it has shown that stereo vision can provide usable range data for UGV's in a timely, relatively cost-effective manner and that stereo is a promising alternative to emissive range sensors, including laser scanners. For example, optimized, software-only versions of the algorithms described here are now planned to be used on prototype microrovers for Mars exploration; until recently, stereo was still considered too computationally expensive for such vehicles. Future work will improve the quality of the range data, integrate terrain classification capabilities with range imaging, and use new imager and computing technologies to reduce the mass, volume, and power consumption of the vision system.

6 Acknowledgements

Thanks to Bob Bolles and Harlyn Baker of SRI International for bringing the FLIR cameras to JPL, as well as assisting in mounting them on the HMMWV, interfacing them to the vision system, calibrating them, and evaluating the results. Thanks to Terry Boult for helping to collect the multispectral imagery.

References

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

[2] L. H. 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.

[3] M. Drumheller and T. Poggio. On parallel stereo. In *Proc. IEEE Conf. on Robotics and Automation*, pages 1439–1448. IEEE, 1986.

[4] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. Technical Report No. 1369, INRIA, January 1991.

[5] D. H. Ballard and C. M. Brown. *Computer Vision*. Prentice-Hall, Englewood Cliffs, NJ, 1982.

[6] M. Daily et al. Autonomous cross-country navigation with the ALV. In *Proc. IEEE Int'l Conf on Robotics and Automation*, pages 718–726. IEEE Computer Society, April 1988.

[7] P. Grandjean and L. Matthies. Perception control for obstacle detection by a cross-country rover. In *Proc. IEEE Int'l Conf. Robotics and Automation*, May 1993.

[8] C. Elachi. *Introduction to the Physics and Techniques of Remote Sensing*. John Wiley and Sons, New York, NY, 1987.

[9] A. Kelly, A. Stentz, and M. Hebert. Terrain map building for fast navigation on rugged outdoor terrain. In *Proceedings SPIE Conference on Mobile Robots*. SPIE, November 1992.

[10] D. Langer, J. K. Rosenblatt, and M. Hebert. An integrated system for autonomous off-road navigation. In *Proc. 1994 IEEE Int'l Conf. on Robotics and Automation*, pages 414–419, May 1994.

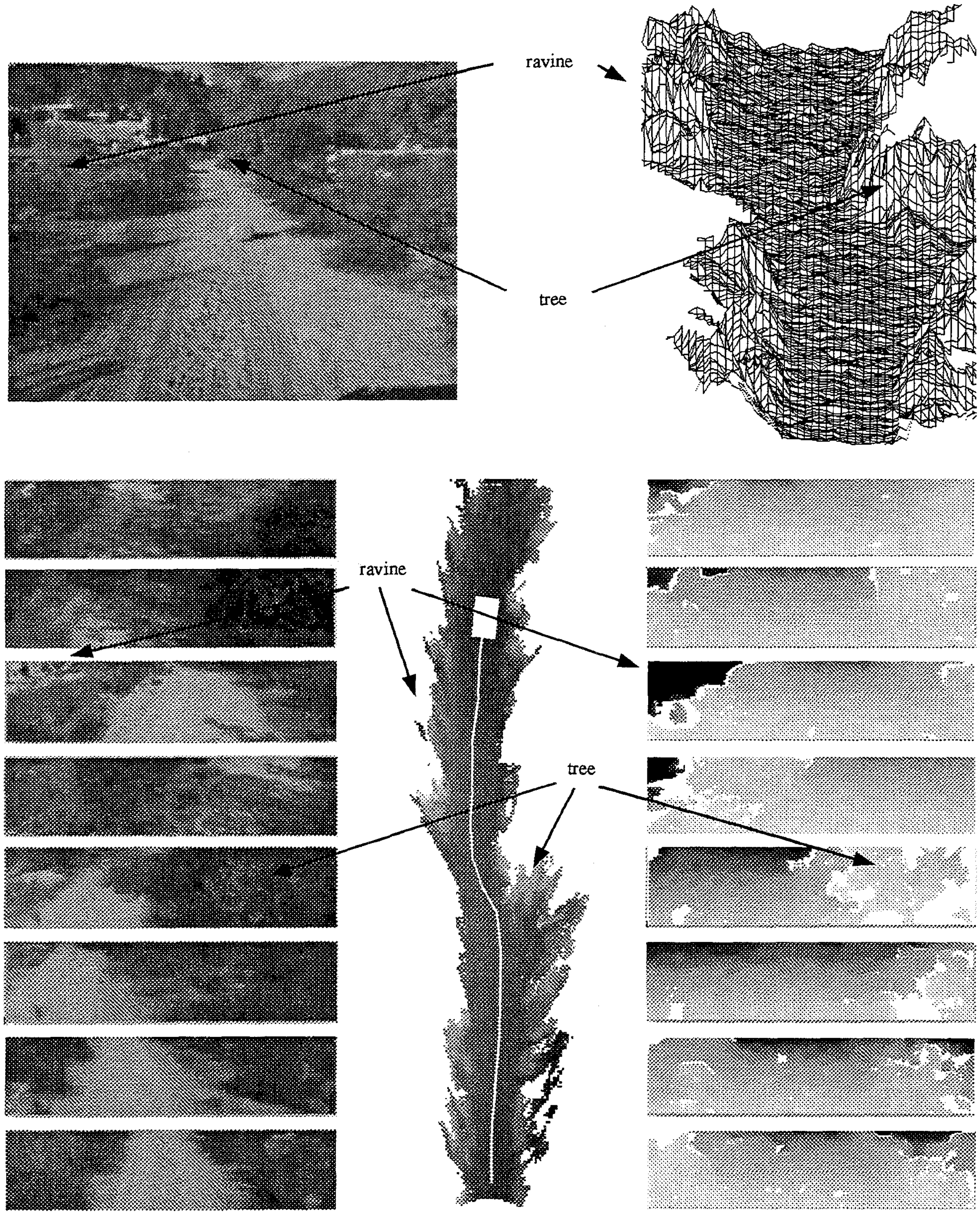


Figure 2: Road run with Ranger, covering about 200 meters. Upper left: image of the road as seen from the start of the run (the rock on the road was not present during the run). The vehicle turned left to avoid the marked tree, then turned right to avoid a bush on the far side of the road. Upper right: wireframe rendering of a composite elevation map built from all range images computed during the run. Bottom: data from the run. Shown are windows of attention from representative intensity images (left side), corresponding range images (right side), and the composite elevation map from the whole run (center). The white curve and rectangle on the elevation map represent the vehicle path and the vehicle itself at the end of the run.