

# Stereo vision based terrain mapping for off-road autonomous navigation

Arturo L. Rankin<sup>\*</sup>, Andres Huertas, Larry H. Matthies  
Jet Propulsion Laboratory, California Institute of Technology  
4800 Oak Grove Drive, Pasadena, CA, USA 91109

## ABSTRACT

Successful off-road autonomous navigation by an unmanned ground vehicle (UGV) requires reliable perception and representation of natural terrain. While perception algorithms are used to detect driving hazards, terrain mapping algorithms are used to represent the detected hazards in a world model a UGV can use to plan safe paths. There are two primary ways to detect driving hazards with perception sensors mounted to a UGV: binary obstacle detection and traversability cost analysis. Binary obstacle detectors label terrain as either traversable or non-traversable, whereas, traversability cost analysis assigns a cost to driving over a discrete patch of terrain. In uncluttered environments where the non-obstacle terrain is equally traversable, binary obstacle detection is sufficient. However, in cluttered environments, some form of traversability cost analysis is necessary. The Jet Propulsion Laboratory (JPL) has explored both approaches using stereo vision systems. A set of binary detectors has been implemented that detect positive obstacles, negative obstacles, tree trunks, tree lines, excessive slope, low overhangs, and water bodies. A compact terrain map is built from each frame of stereo images. The mapping algorithm labels cells that contain obstacles as no-go regions, and encodes terrain elevation, terrain classification, terrain roughness, traversability cost, and a confidence value. The single frame maps are merged into a world map where temporal filtering is applied. In previous papers, we have described our perception algorithms that perform binary obstacle detection. In this paper, we summarize the terrain mapping capabilities that JPL has implemented during several UGV programs over the last decade and discuss some challenges to building terrain maps with stereo range data.

**Keywords:** Terrain mapping, stereo vision, passive perception, autonomous navigation

## 1. INTRODUCTION

The complexity of cross country terrain can greatly vary, from simple scenes such as a bare plain, to difficult scenes, such as a forest with underbrush and overhangs. Successful unmanned ground vehicle (UGV) autonomous navigation over cross country terrain requires reliable perception and representation of the natural terrain. A reliable representation of the natural terrain is critical to enabling a UGV to plan safe paths and reach its commanded goal location during a military mission. One common way to represent natural terrain in the vicinity of a UGV is to construct a terrain map using ranging sensors mounted to the UGV [1]. Other imaging sensors, such as a color, multi-spectral, thermal infrared, or polarization camera, may be mounted to the UGV to assist in classifying the terrain. A Cartesian grid map containing cells of uniform size is a convenient way to represent the natural terrain in a UGV's operating environment.

There are a variety of Cartesian grid maps the robotic community has experimented with over the last two decades. Depending on the complexity of the terrain and required UGV operating speeds, a 2D, 2.5D, or 3D grid map can be used. A 2D grid map is typically an occupancy grid that stores the location of non-traversable, traversable, and unknown terrain [2,3]. In an occupancy grid, if an obstacle is detected in a cell, that cell is non-traversable. A 2D grid map could be effectively used to represent cross country terrain that is predominantly planar with discrete obstacles. However, for more complex environments, a higher dimensional map is required. A 2.5D grid map, in its simplest form, is a digital elevation map (DEM) that stores a single elevation measure for each cell. The more common 2.5D grid map used by UGVs is a terrain map, which stores an elevation measure and terrain characteristics within each cell. Alternatively, a 3D grid map could be used [4].

A 3D grid map (also called a voxel map) represents the world using adjacent stacks of cubes. If a data point from a ranging sensor lies within a cube, that cube is labeled as occupied (as opposed to empty). The advantage of using a voxel representation is that the elevation of multiple objects at the same range and azimuth is fully captured. The

disadvantage of using a voxel representation is that the maps are considerably larger, and thus more costly to search with a path planner. A voxel representation has been used to generate coarse paths for aerial and underwater robotic vehicles [5], but it is not a representation easily amenable to searches with a local path planner that considers UGV kinematic constraints. For UGVs, a voxel representation is typically used as an intermediate representation within which terrain details can be extracted, such as the terrain support surface [6,7]. The support surface elevation can then be stored in a 2.5D terrain map that is used to make local driving decisions.

Terrain maps commonly include an obstacle field that records detected obstacles [8]. Some types of natural terrain that may be detected as obstacles include non-traversable rocks, tree trunks too wide to push over, logs, stumps, ditches, holes, low overhangs (such as thick branches), large water bodies, mud bodies, and steep terrain. Instead of, or in addition to the obstacle field, terrain map cells can include the traversability cost of traveling through that cell [9,10]. Other information can be contained in each cell, for example, in [11], the terrain map cells contained surface slope, uncertainty measures, average elevation, and a list of positive obstacles and overhangs.

A composite world map, or world model, can be built by merging a sequence of range images [12], or by merging single frame terrain maps and a priori information if it is available [13]. Since a world map has limited size, to avoid driving off the map, or having to periodically recenter the vehicle in the world map, a vehicle-centered scrolling map is typically used [1]. The world map is typically global in that the cell edges are aligned with the four cardinal directions (north, south, east, and west). Broten et. al. experimented with building an egocentric world map [14], but found that turning maneuvers caused loss of fidelity, such as a smearing effect on obstacles [15]. The map representation, size, and cell resolution appropriate for a UGV depends on the perception sensor capabilities, the UGV mobility requirements, and the characteristics of the operating environment. For example, the slow moving packbot size Urban Reconnaissance Robot performed autonomous navigation using 2.5m x 2.5m single frame occupancy maps with a cell resolution of 10cm [2]. For the higher speed Demo III experimental unmanned ground vehicle (XUV), autonomous navigation was performed using a 50m x 50m world map with a cell resolution of 40cm [13].

Vehicle mounted active ranging systems (such as lidar) and passive ranging systems have been successfully used to construct terrain maps for UGVs [16]. Fixed baseline stereo ranging offers several advantages over active ranging methods. Stereo sensors have a smaller angular resolution than lidar, are available at lower cost, and with the emergence of megapixel cameras, provide higher resolution images. Stereo can provide range data at longer ranges than lidar, particularly on horizontal driving surfaces even when the grazing angle is shallow. In addition, stereo sensors do not emit an electromagnetic signature detectable by a foe. Lidar, however, continues to provide higher quality range data than current real-time stereo algorithms, but its range is limited. (For example, the Demo III lidar sensor had a maximum range of 50 meters.) Some of the issues that have constrained the quality of stereo range data in the past are 1) the range uncertainty of the stereo range data increases quadratically with range, 2) because stereo utilizes a correlation window, there can be a tendency to enlarge objects, 3) because the correlation window can overlap an object in the foreground and the background, range data can bleed between the foreground object and background, 4) stereo range data may be missing where there is little texture in the input imagery, and 5) there tends to be a lack of stereo range data on thin structure, such as tall vegetation.

The Jet Propulsion Laboratory (JPL) has been involved in addressing these issues under several U. S. Department of Defense (DoD) UGV programs during the last decade, such as the Army Research Laboratory (ARL) Demo III program [17], the Defense Advanced Research Projects Agency (DARPA) Perception for Off-Road Robots (PerceptOR) program [18], and the ARL Robotic Collaborative Technology Alliances (RCTA) program [8,19,20,21,22,23]. During Demo III and PerceptOR, JPL fielded passive perception systems with camera separation distances up to 35cm that processed stereo imagery at a resolution of 320x240 pixels. During RCTA, JPL has fielded a passive perception system with a camera separation distance of 50cm that processes stereo imagery up to a resolution of 1024x768 pixels [23]. Increasing our camera separation distance and processing higher resolution imagery has increased the accuracy of our range data. To address the oversizing of objects and range bleeding between foreground objects and the background, we implemented an edge preserving prefilter [19], intelligent use of object edge cues, a variation of the traditional shiftable correlation window approach [20], and a sum of absolute differences correlator (SAD) that utilizes five overlapping windows (SAD5). To address the issue of lack of range data where there is low texture, we are currently experimenting with an advanced prefilter under the U. S. Army Future Combat System (FCS) program. To address the issue of lack of stereo range data on thin structure, we have fielded a multi-baseline (9.5cm, 20.5cm, and 30cm) stereo system [18] and experimented with multi-resolution stereo processing.

These efforts have significantly improved the quality of our stereo range data over the last decade, enabling us to progressively generate higher quality stereo-vision based terrain maps with each new program. Our expectation is that the quality of stereo range data will continue to improve as advances in computing hardware enable advanced stereo algorithms to be performed in real-time. In this paper we summarize the Cartesian terrain mapping capabilities that JPL has implemented during the Demo III, PerceptOR, and RCTA programs as the quality of stereo range data has progressively improved. In the following sections, we discuss building single frame terrain maps, merging single frame terrain maps into a world map, building terrain maps using multi-resolution stereo range data, building long range terrain maps at coarse resolution, and some complications to mapping with stereo range data.

## 2. SINGLE FRAME TERRAIN MAPS

During the Demo III and PerceptOR, JPL developed a passive perception subsystem that generated a terrain map for each stereo image pair that was processed. Single frame terrain maps were shipped to a world model subsystem over a local area network (LAN). Each map cell contained elevation, terrain classification, object, roughness, and confidence values. During RCTA, a traversability cost field was added to the map data structure. The terrain map is a square north-oriented map, where the data within each cell is encoded into four bytes of memory. During the Demo III and RCTA programs, our map cell resolution was 40cm, which matched the cell resolution of the world map implemented by General Dynamics Robotic Systems (GDRS). During PerceptOR, our map cell resolution was 20cm, which matched the cell resolution of the world map implemented by Science Applications International Corporation (SAIC). The single frame terrain map elevation resolution was 2cm. A new origin was calculated for every single frame map in global coordinates, making sure it coincided with a world map cell corner. That way, single frame terrain maps correctly line up with a portion of the world map, facilitating the merging of map data.

Figure 1 illustrates the structure of the single frame terrain map. In this example, the cameras happen to be pointed south. But as the vehicle or sensor platform moves, the camera's pointing direction will change. It is the pointing direction that determines the position of the cameras in the single frame terrain map. The stereo cameras will always lie somewhere on the largest circle that can fit within the map. The map center can be located with a vector originating at the midpoint of the stereo cameras and extending in the pointing direction  $\frac{1}{2}$  the length of one side of the map. If the global coordinates of the left and right cameras are  $(x_L, y_L, z_L)$  and  $(x_R, y_R, z_R)$ , respectively, the length of one side of the map is  $L$ , and the pointing direction with respect to north is  $\theta$ , then the coordinates of the map center are,

$$x_c = (x_L + x_R) / 2 + (L / 2) \cos(\theta) \quad (1)$$

$$y_c = (y_L + y_R) / 2 + (L / 2) \sin(\theta) \quad (2)$$

The lower left corner of the map is taken to be the origin. If the global coordinates of the origin of the world map are  $(X_0, Y_0, Z_0)$ , the global coordinates of the vehicle are  $(x_v, y_v, z_v)$ , and the length of one side of a map cell is  $d$ , then the coordinates of the origin of the single frame terrain map are,

$$x_0 = \left[ (\text{int}) \left( \frac{x_c - L / 2 - X_0 + d / 2}{d} \right) \right] d \quad (3)$$

$$y_0 = \left[ (\text{int}) \left( \frac{y_c - L / 2 - Y_0 + d / 2}{d} \right) \right] d \quad (4)$$

$$z_0 = z_v \quad (5)$$

Note that the coordinates of the world map are used to ensure that all single frame terrain maps will be aligned with the world map. Instead of using 32 bits to store the floating point elevation for a map cell, we assign the vehicle elevation to the map origin elevation and report the change in elevation (from the origin elevation) in terms of the number of discrete steps. At the 2cm map elevation resolution, terrain elevation from -20.5 to 20.5 meters (with respect to the vehicle's elevation) was encoded into 11 bits.

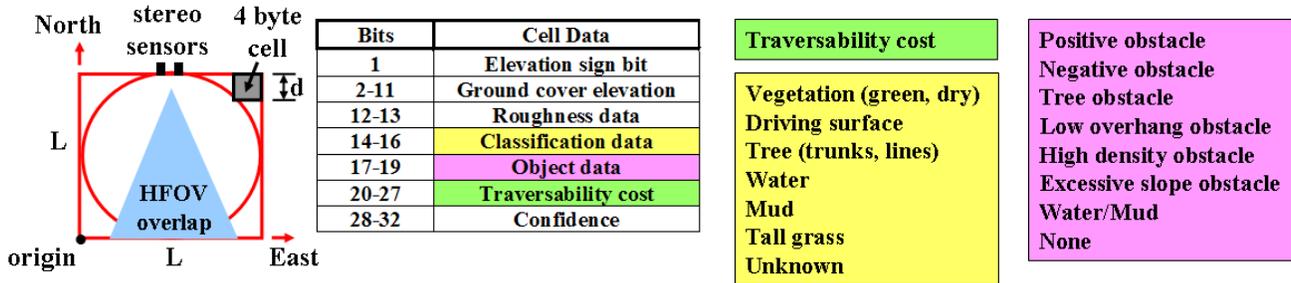


Fig. 1. The data for a single frame terrain map cell is packaged into four bytes of memory. The origin of the map is determined for every stereo pair of images based on the direction the stereo sensors are currently pointed. Only cells within the horizontal field of view (HFOV) of each stereo sensor were packaged into the compressed terrain maps sent to the world map.

For each four byte map cell, bits 1-11 are used to encode elevation, bits 12-13 are used to encode terrain roughness, bits 14-16 are used to encode terrain classification, bits 17-19 are used to encode object classification, bits 20-27 are used to encode terrain traverability cost, and bits 28-32 are used to encode a confidence value. The vehicle position and orientation, map origin, and cell resolution are all included in a header for each map. The map was compressed into a vector prior to delivery to the world map. For each row in the map, only the data between the first column with data and the last column with data are packaged into the compressed map. At the beginning of the data for each row are a start column index and an end column index.

Terrain classification can be performed in image space and input to a single frame terrain map, or it can be performed in map space. During Demo III, JPL implemented a color-based terrain classification algorithm based on Bayesian assignment that operated in image space. The class likelihoods were represented using a mixture-of-Gaussian model and parameters of the models were estimated from training data using the Expectation Maximization algorithm [17]. Figure 2 shows an example terrain map that encodes elevation and terrain classification. When image pixels with different terrain classification values fall within a single map cell, the predominant classification is assigned to that map cell. The elevation of the range data within each cell was averaged in this example. Averaging range data is only reasonable for scenes that do not contain overhangs and when obstacle detection will not be performed within the terrain map. Figure 3 illustrates that when color sensors are not available, stereo range data can be used to perform geometry-based terrain classification in map space. Here, the maximum ground cover elevation for each cell is determined, and a least-squares plane fit of 1.2m x 1.2m local patches of terrain is performed. The plane fit residual is used as a measure of terrain roughness. The terrain slope, plane fit residual, and elevation standard deviation are thresholded to classify the terrain as ground, ground clutter, or unknown.

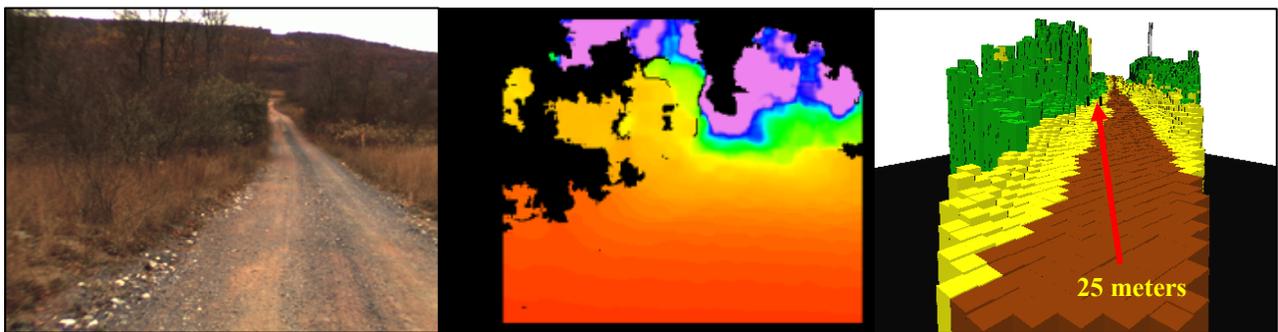


Fig. 2. A stereo pair of color images can be used to build a map that encodes elevation and color-based terrain classification. In the middle stereo range image, the red pixels correspond to close range, blue pixels correspond to far away, the colors between red and blue correspond to an intermediate range, and magenta pixels are beyond 100 meters. In the terrain classification map, brown represents soil, yellow represents dry vegetation, and green represents green vegetation. Here, the elevation of the range data within each cell was averaged. This is a 60 meter map with 40cm x 40cm cells.

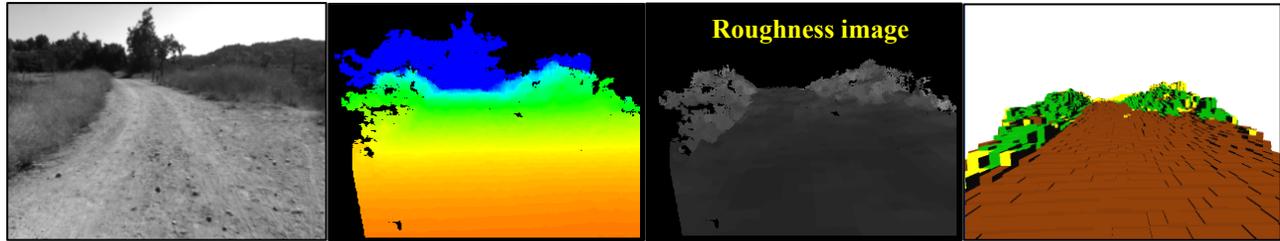


Fig. 3. A stereo pair of color or monochrome images can be used to build a map that encodes elevation and geometry-based terrain classification. The images from left to right are monochrome intensity, stereo range, roughness, and terrain classification map. In the terrain classification map, the maximum ground cover elevation is shown. Brown represents the ground, green represents ground clutter, and the yellow cells have unknown classification.

Although the range uncertainty of stereo range data increases quadratically with range, the height uncertainty increases linearly with range. This can be exploited to detect tall objects, such as tree lines, at far range. The range to the tall objects can then be refined as the UGV approaches them. As trees can provide a UGV with cover from aerial detection, a dash to a detected tree line may be a desirable tactical behavior. During Demo III, JPL used stereo ranging to detect tree lines out to 100 meters. A stereo pair of 3CCD cameras was used during the daytime and a stereo pair of mid-wave infrared cameras was used during the nighttime, both having a baseline of 35cm. Figure 4 contains two sample tree-line detection results. The upper left image shows the elevation profile of a single column of range data that extends down range and up a tree line. The upper right image shows cumulative tree-line detection in blue from thresholding the terrain height and slope on a column by column basis. The bottom row shows a left RGB image of a tree line normal to an XUV and approximately 66 meters away (in the northwest direction), and a single frame occupancy map from performing tree-line detection on stereo range data. The range to the tree line was measured with a single point laser range finder. In the single frame occupancy map, the mean tree-line detection range is 64 meters. Note that the detection swath has a depth of several meters. This is due to the natural porosity of the leading edge of tree lines and the reduction of stereo range accuracy with increasing range inherent to using a fixed baseline stereo rig. Using the wider baseline, high-resolution stereo system, such as the one developed under RCTA [23], would reduce this effect.

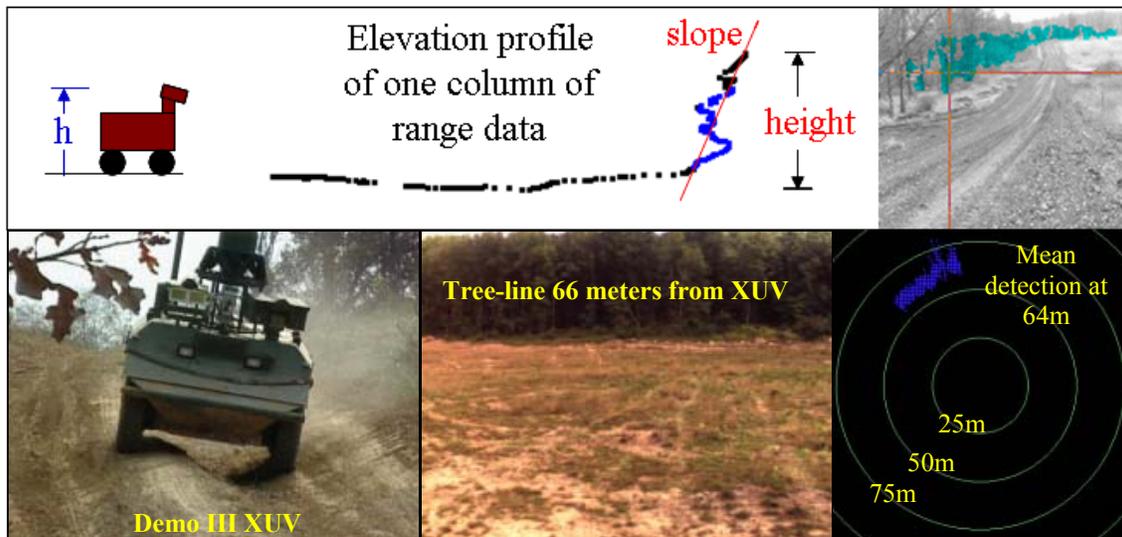


Fig. 4. Tree-line detection was performed on XUVs during the Demo III program out to a range of 100 meters using a stereo pair of color cameras. The upper left image shows the elevation profile of the column in the cross hairs in the upper right image. Tree lines are detected by thresholding the slope and height of stereo range data. The bottom right single frame occupancy map shows tree-line detection for the bottom middle RGB image.

During RCTA, JPL implemented and evaluated seven discrete obstacle detection algorithms using stereo imagery collected on a surveyed obstacle course [8]. These included a positive obstacle detector, a negative obstacle detector, a non-traversable tree trunk detector, an excessive slope detector, a range density based obstacle detector, a multi-cue water detector, and a low-overhang detector. Figure 5 illustrates a terrain map generated during the daytime from a

stereo pair color images that encodes elevation and detected obstacles. In this example, detected tree trunks, positive obstacles, excessive slope, and a low overhang are represented. Stereo ranging is not limited to the daytime. Figure 6 illustrates a terrain map generated during the nighttime from a stereo pair of thermal infrared cameras that also encodes elevation and detected obstacles. In this example, a detected positive obstacle (stack of hay bails) and negative obstacle (0.5 meter wide trench) is represented. The trench was detected by combining geometric and thermal cues [18].

Obstacle cells are treated as prohibited or “no-go” regions. While encoding maps with elevation and detected obstacles enables you to represent unsafe (no-go) regions, a disadvantage of this approach is that it does not indicate how traversable non-obstacle regions are with respect to other non-obstacle regions. An example is a rut running down the length of one side of a road that is mild enough to avoid causing a UGV to high center. Ideally, a UGV should drive down the smooth side of the road and avoid driving with its side wheels in the rut. This can’t be accomplished with a binary obstacle map. To address this limitation, our single frame terrain map representation was modified during RCTA to also include an 8-bit traversability value for each cell. Cells containing a traversability cost of 255 are no-go regions. Cells that have no range data to analyze are labeled with a traversability cost of 0. Traversable cells are assigned a cost of 1-254, depending on their level of traversability.

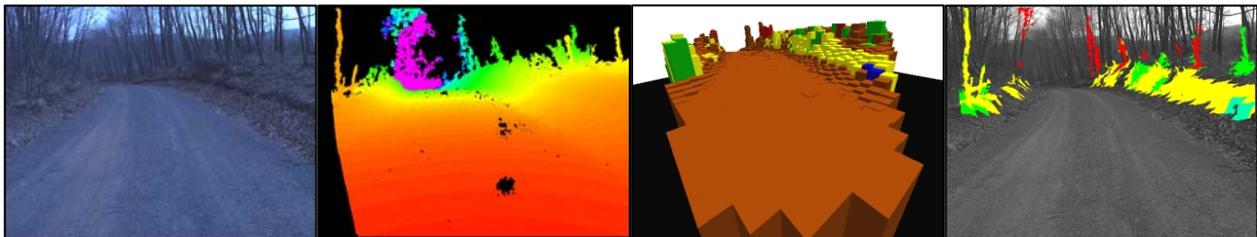


Fig. 5. A stereo pair of color or monochrome images can be used to build a terrain map that encodes elevation and detected obstacles. The images from left to right are a color image of a scene from Ft. Indiantown Gap, PA, stereo range, a 50 meter terrain map with a cell resolution of 40cm, and obstacles transferred from map space to image space. In the terrain obstacle map, non-brown cells are no-go regions. Green indicates a positive obstacle, red indicates a tree trunk, yellow indicates excessive slope, and blue indicates a low overhang exists above that cell.

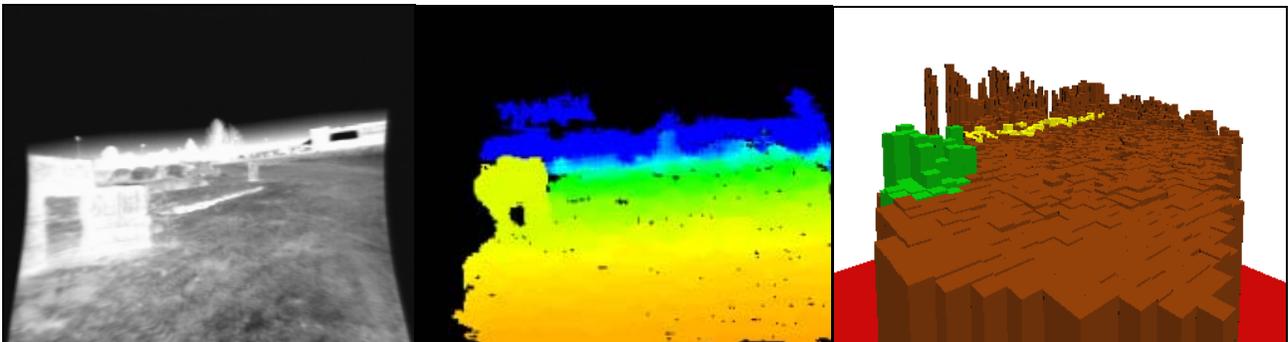


Fig. 6. A stereo pair of FLIR images can be used to build a terrain map during the nighttime that encodes elevation and detected obstacles. The images from left to right are a mid-wave infrared image of a scene containing hay bails and a trench, stereo range, and a 25 meter terrain map with a cell resolution of 20cm. In the terrain map, non-brown cells are no-go regions. Green indicates a positive obstacle and yellow indicates a negative obstacle.

As with [9], we generate a traversability cost for each map cell by evaluating the terrain elevation for step and slope hazards. A step hazard is a change in elevation that a UGV cannot safely achieve. Terrain with a slope that is unclimbable, or a slope that could cause a UGV to tip over, is a slope hazard. Slope hazard analysis is performed in the single frame terrain map on 1.2m x 1.2m patches of terrain using the minimum elevation in each cell. Slope traversability cost is a linear function of terrain slope. A slope exceeding the maximum safe slope for a UGV is considered non-traversable. Step hazard analysis is performed by differencing the maximum and minimum ground cover elevation for each cell. Step traversability cost is a linear function of the change in elevation in each cell. Map cells having a change in elevation exceeding the maximum safe step are considered non-traversable. On 40cm resolution terrain maps, slope analysis is performed out to 30 meters and step analysis is performed out to 50 meters. Slope and step cost are merged into a single traversability cost map by using the maximum cost at each map cell. Figure

7 shows an example stereo range image, slope cost image, step cost image, fused cost image, and single frame traversability cost map for a scene at Ft. Indiantown Gap.

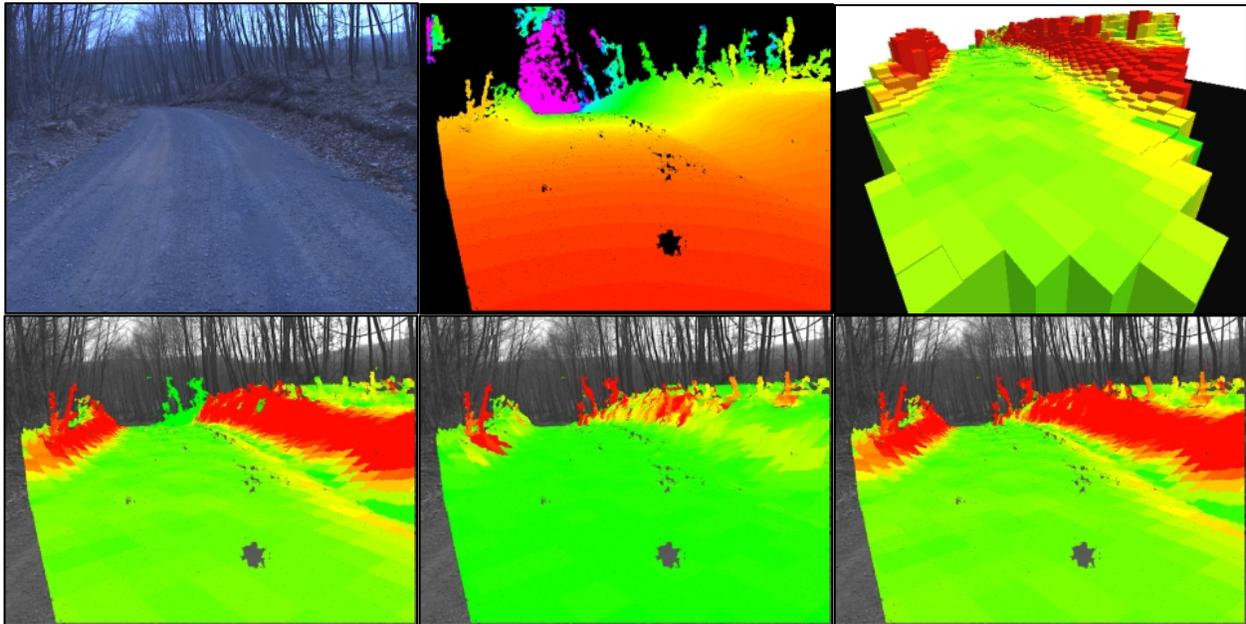


Fig. 7. A stereo pair of color or monochrome images can be used to build a map that encodes elevation and terrain traversability cost. The top row images from left to right are a scene from Ft. Indiantown Gap, stereo range, and a 50 meter terrain map with a cell resolution of 40cm. The bottom row images from left to right are traversability cost based on slope hazards, traversability cost based on step hazards, and traversability cost based on combined slope and step hazards. Green corresponds to low cost, red corresponds to high cost, and the colors in between correspond to an intermediate cost.

### 3. WORLD MAP

Using the current single frame terrain map to plan safe UGV paths is of limited value. First, planning would always be restricted to the terrain currently within the sensor’s FOV, and second, single frame terrain maps are inherently noisy, since they are constructed from stereo range data that may contain artifacts. By fusing single frame terrain maps into a composite world map, planning can be performed using all the perceived terrain and noisy map data can be suppressed. During RCTA, JPL implemented a scrolling, north-oriented, vehicle-centered world map (built on map base class software provided by GDRS) that encodes elevation, terrain classification, terrain traversability, and detected obstacles. The world map is 100m x 100m, with each cell having a 40cm resolution. Temporal filtering is performed in the world map to smooth out noisy elevation and traversability cost data from the single frame terrain maps. An example world map is illustrated in Figure 8. In this example, it is apparent from the RGB image that the road surface is smooth. But a water spot on the lens of one of the stereo cameras caused a spike in the stereo range data that resulted in high traversability cost in one of the road cells in the single frame terrain map. In the world map, the noise is smoothed out. Here, the last three values for ground cover elevation and traversability cost added to each world map cell was averaged.

In Figure 3, we illustrated geometry-based terrain classification using a single frame terrain map. Similarly, terrain classification and obstacle detection can be performed using the world map. In Figure 9, we illustrate ground detection using world map data. Ground detection was performed in the world map by thresholding the standard deviation of the terrain elevation, the local tilt of the terrain elevation (estimated with a least squares plane fit), and the plane fit residual. In the overlays, red corresponds to low values, blue corresponds to high values, and the colors in between correspond to an intermediate value. The stereo point cloud after ground clutter removal in the bottom right image of Figure 9 extends down range approximately 50 meters. Since we only expect to find some terrain types (like mud) on the ground surface, the ability to separate ground clutter from the ground is particularly useful [22]. The benefit of using world map data for ground detection is that elevation noise has already been filtered.

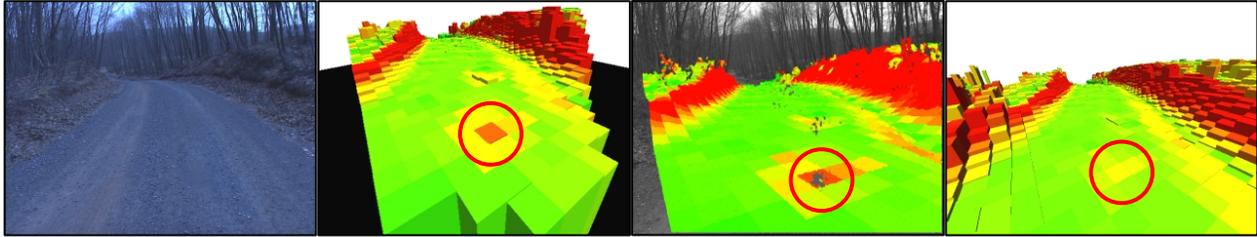


Fig. 8. Temporal filtering is performed in the world map to smooth out noise in single frame terrain maps. The images from left to right are a scene from Ft. Indiantown Gap, a single frame terrain map that encodes elevation and terrain traversability cost, traversability cost transferred from map space to image space, and a 100m x 100m temporal filtered world map with a cell resolution of 40cm. The noisy map cell in the single frame terrain map was caused by a water spot on the lens of one of the stereo cameras.

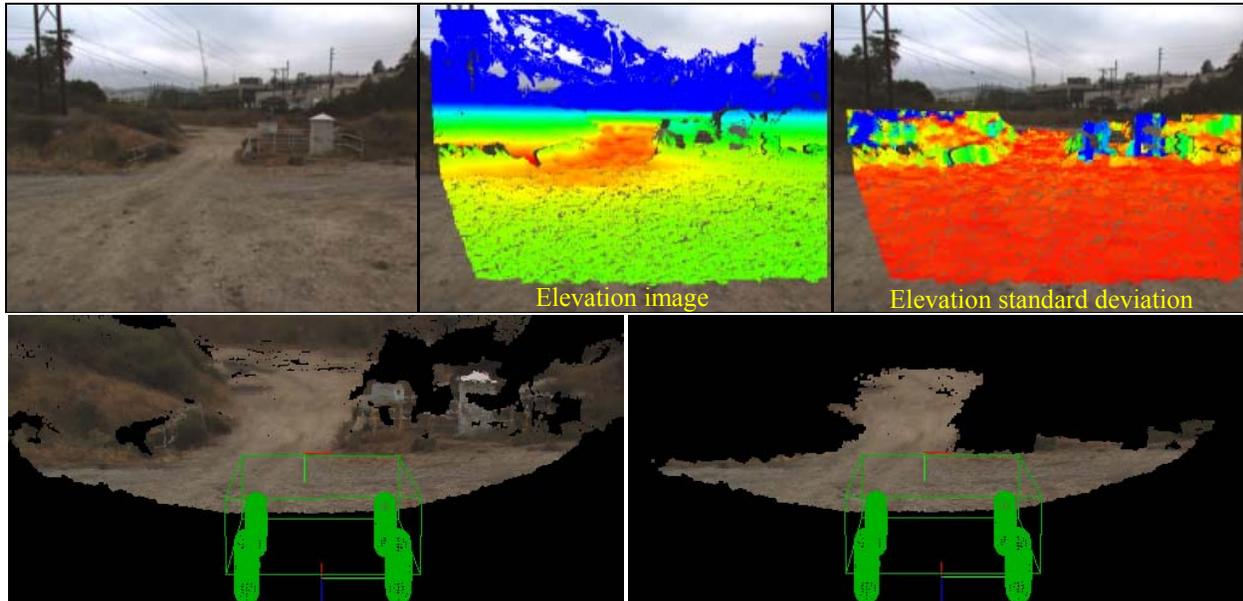


Fig. 9. Ground detection is performed in the world map by thresholding the standard deviation of the terrain elevation, the local tilt of the terrain estimated with a least squares plane fit, and the plane fit residual. In the elevation and standard deviation images, red corresponds to low values, blue corresponds to high values, and the colors in between correspond to an intermediate value in the overlays on images. The images in the bottom row contain a stereo point cloud before and after ground clutter removal.

#### 4. MAPPING WITH MULTI-RESOLUTION STEREO DATA

Demands for UGV's to operate at higher speed translate into the need for high quality terrain map data at longer ranges. This can be accomplished by exploiting the availability of high resolution cameras to process higher resolution stereo imagery. Stereo correlation, however, is computationally expensive. Under RCTA and FCS, JPL has worked several methods in parallel to increase the resolution of stereo imagery we can process in real time: 1) implement the SAD5 stereo correlator in FPGA, 2) further optimize our C++ version of the SAD5 stereo correlator, and 3) implement multi-resolution stereo processing. Our multi-resolution approach was to process the stereo imagery in four horizontal swaths of different resolutions (from the bottom of the image to the top: 256x192, 512x384, 1024x768, 256x192) so as to achieve a maximum horizontal pixel footprint of 3cm in the bottom three swaths. Of course, beyond some range, a horizontal pixel footprint of 3cm is no longer possible. That point defines the start of the top swath. In this section, we compare the quality of terrain maps generated using multi-resolution and full resolution SAD5 stereo ranging with an example.

The top row of Figure 10 shows a scene with trees on the left and center, and bushes on the right (located 10-30 meters away), and the four-level multi-resolution stereo range image for the scene. The images in the middle row of Figure 10 contain bird's-eye view 50m x 50m single frame traversability terrain maps from full and multi-resolution stereo range

images. The circular rings are at 10 meter intervals. The white regions are either out of the sensor's FOV, or there was not enough range data in that cell to perform the traversability analysis. Here, at least two range measurements in each cell were required. The images in the bottom row of Figure 10 contain 100m x 100m world maps, each generated from the single frame traversability terrain map above it and the preceding two single frame traversability terrain maps in the sequence. Here, the range data that was over 2 meters above the minimum elevation in each cell was clipped to eliminate overhangs and canopy above the UGV height from influencing the ground cover elevation.

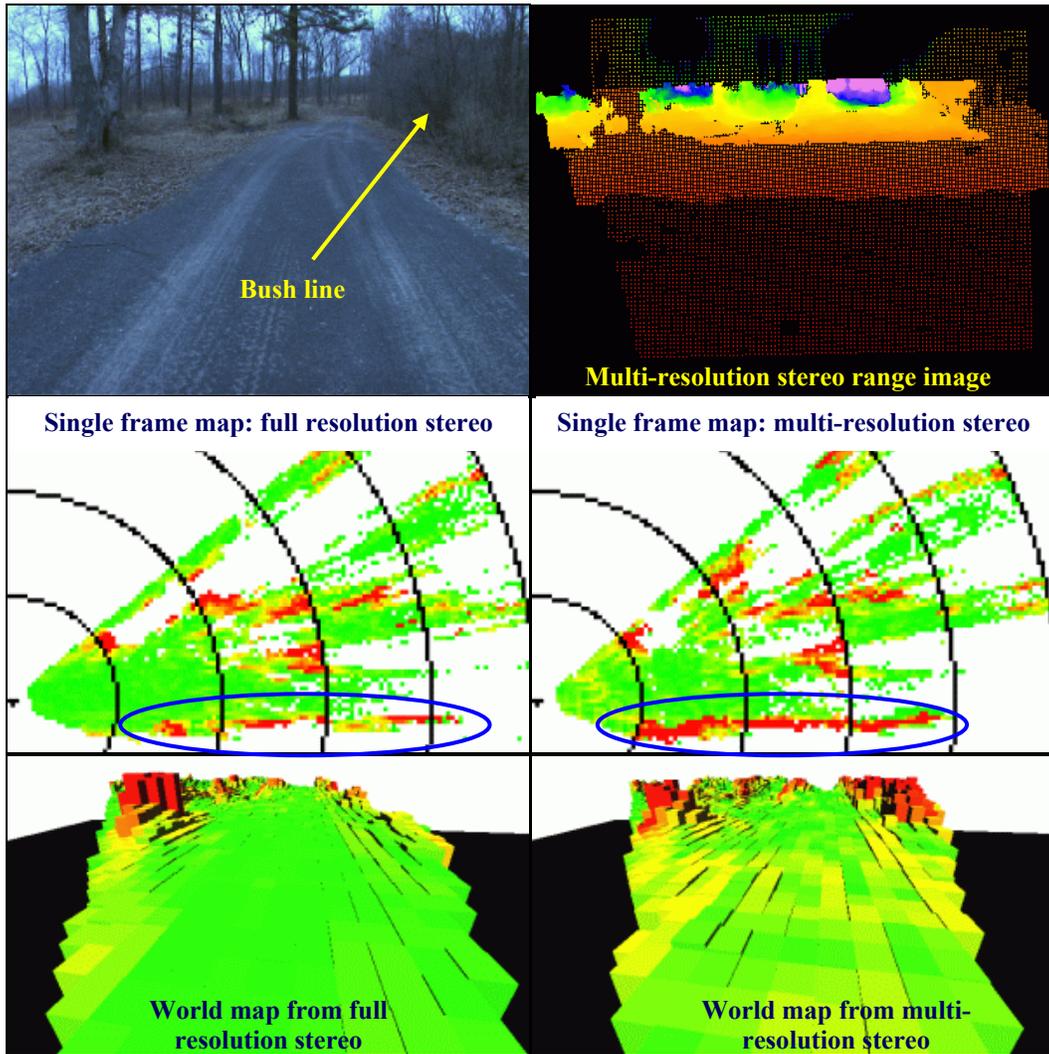


Fig. 10. Comparison of maps constructed from full resolution (1024x768) stereo range data and from four level multi-resolution stereo range data. The middle row contains 50m x 50m single frame traversability terrain maps with circular rings at 10 meter intervals. The bottom row contains 100m x 100m world maps, each generated from the single frame traversability terrain map above it and the preceding two single frame traversability terrain maps in the sequence.

The world maps generated from full and multi-resolution stereo range images both have advantages. The world map generated from full resolution stereo range data did a better job of modeling the road elevation in the near field and had more map cells with data on the road. This is simply because there was more range data available to model. The world map generated with multi-resolution stereo range data did a better job of modeling the bushes on the right side of the scene. Thin structure, such as tall grass and bushes, are typically difficult to correlate with a high-resolution wide-baseline stereo system. At lower resolution, however, these objects have their noisy characteristics smoothed, making them easier to correlate. The bushes fall into lower resolution swaths of the multi-resolution stereo range data and are thus better represented in the world maps generated from multi-resolution stereo range data.

In summary, when it is computationally too expensive to process full resolution stereo imagery, multi-resolution stereo processing can be used to process a portion of the terrain that is far away at full resolution, to increase the pixel footprint there by sacrificing some resolution close to the UGV. Generating world maps with multi-resolution stereo range data instead of full resolution stereo range data may result in a slight decrease in elevation fidelity in the near field and map data, but improved modeling of thin structure in the scene.

## 5. LONG RANGE MAPPING

At 100 meters, one sigma of stereo range error is approximately 3.27 meters and the downrange pixel footprint (on a horizontal ground surface) is approximately 7.67 meters with the JPL RCTA stereo system mounted to an XUV (1.04mrad x 1.09mrad pixel resolution, 50cm stereo baseline, 1.53m sensor height). The height error and pixel footprint, however, are significantly smaller than the range values. At 100 meters for the same stereo setup, one sigma of height error is approximately 5cm and the vertical pixel footprint is approximately 11cm. The height accuracy of stereo range data can be exploited to predict where the terrain does not have tall step hazards at far range. This may be particularly useful in perceiving long stretches of straight roads. Higher UGV speeds can be achieved when long straight roads are detected during autonomous navigation.

Because the downrange pixel footprint of stereo range data increases with range, high-resolution terrain maps tend to have large gaps between cells with data at far ranges. To reduce this effect, low-resolution terrain maps can be generated. Figure 11 contains maps generated at two different resolutions for a scene that contains a long straight road lined with tall trees and vegetation. The middle row contains a 100m x 100m, 50cm resolution single frame traversability terrain map, and a 150m x 150m, 1m resolution single frame traversability terrain map, both with circular rings at 10 meter intervals. The bottom row contains world maps, each generated from the single frame traversability terrain map above it and the preceding two single frame traversability terrain maps in the sequence. The upper left image contains traversability cost converted from the middle left map to image space. In this example, the range data that was over 6 meters above the minimum elevation in each cell was clipped.

The low volume of range pixels at long range makes long range traversability analysis at the map cell level difficult. For example, more than one range pixel is required in a cell to detect a step hazard. At long range, it only makes sense to detect tall obstacles, such as tree lines and buildings, in a low-resolution terrain map. When low-resolution terrain map cells contain only a single range pixel, that implies there is no vertical structure in that cell. On this basis, map cells with only a single range pixel are labeled traversable in a low-resolution terrain map. As the UGV approaches that terrain, a better assessment of traversability can be performed prior to it being reached.

## 6. COMPLICATIONS TO MAPPING STEREO DATA

There are several scenarios that are particularly challenging to represent in a map using stereo range data. These include scenes that contain low texture (and thus little stereo range data), such as where there is concrete, blacktop, and over-exposed terrain, scenes that contain thin structure such as tall grass, bushes, and undergrowth, scenes that contain water bodies that are reflecting trees and objects in the background, scenes with an abundance of terrain occlusions, scenes with low overhangs, and extremely cluttered environments such as a forest with heavy undergrowth. In this section, we highlight several techniques for building terrain maps with stereo range data to adequately represent complicated scenes.

### 6.1 Terrain mapping and low overhangs

Low overhangs are common when operating in forested areas. Low overhangs are particularly worrisome as they can damage UGV perception sensors, disabling autonomous navigation. The single frame terrain map is structured to record several elevation values for each map cell as intermediate data products; the maximum, minimum, ground cover, and canopy elevation. The minimum elevation in each cell is taken to be the load bearing surface elevation. To generate the canopy elevation map, we difference the maximum and minimum elevation maps. Where the difference exceeds the height of the UGV, we assume a bimodal distribution and cluster the range pixels closest to the minimum (ground) and those closest to the maximum (canopy). For each cell where there is a significant gap between the ground cluster and the canopy cluster, the minimum canopy elevation is saved in the intermediate canopy elevation map. A cell containing a significant number of canopy range measurements and a canopy elevation low enough to hit the UGV is labeled a low overhang obstacle. Figure 12 illustrates detecting low overhangs in a cluttered environment. Brown represents the load bearing surface, green indicates a positive obstacle, blue represents the minimum height of the canopy, and white indicates the cells where the canopy is low enough to be a hazard to the vehicle.

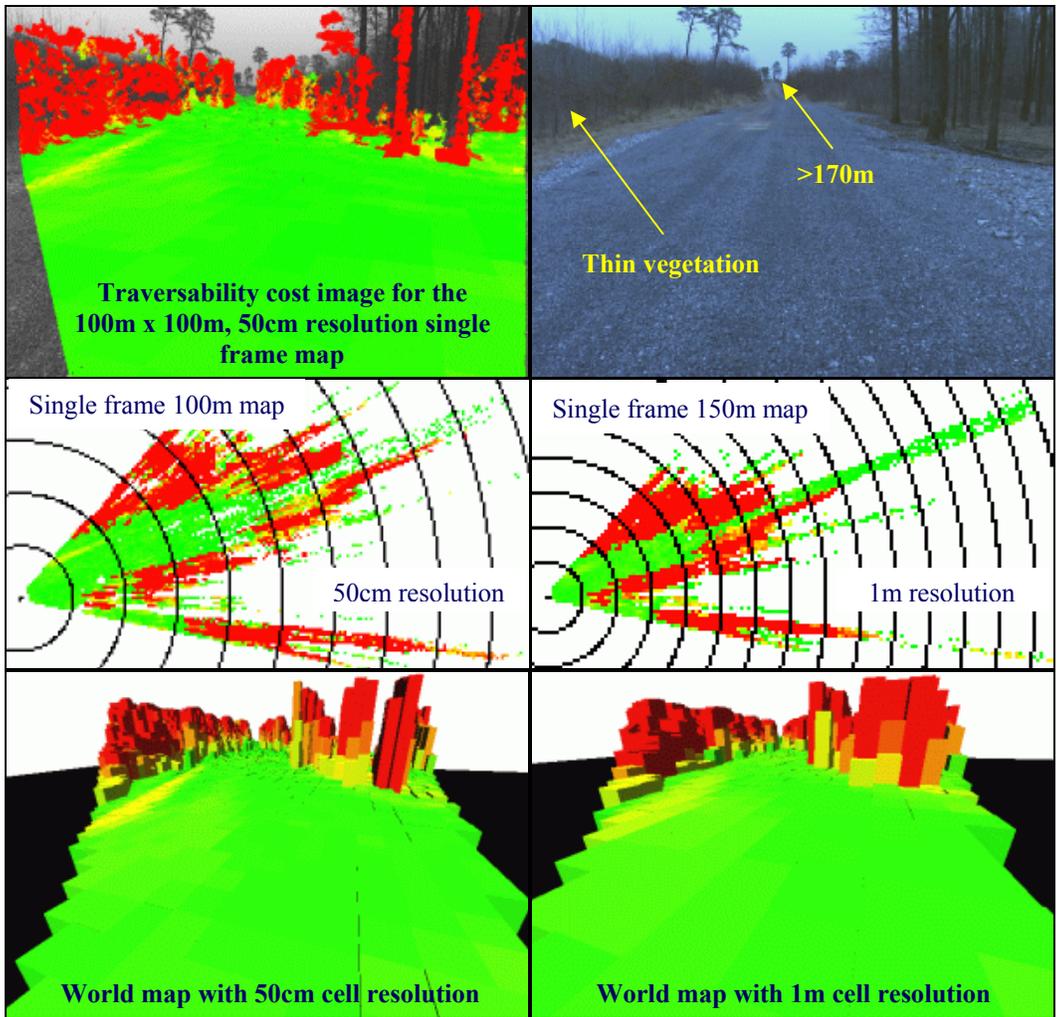


Fig. 11. Long range single frame traversability maps and world maps can be generated by using a coarser map resolution. The middle row contains a 100m x 100m, 50cm resolution single frame traversability terrain map, and a 150m x 150m, 1m resolution single frame traversability terrain map, both with circular rings at 10 meter intervals. The bottom row contains world maps, each generated from the single frame traversability terrain map above it and the preceding two single frame traversability terrain maps in the sequence. The upper left image contains traversability cost converted from the middle left map to image space.

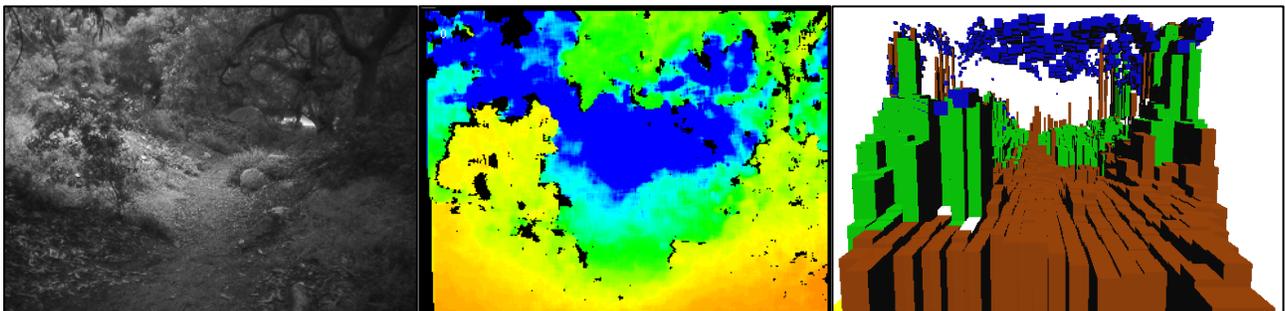


Fig. 12. Off-road terrain in cluttered environments, such as forested areas, can be challenging to represent. A stereo pair of color or monochrome images can be used to build a map that encodes the load bearing surface and detected obstacles (including low overhangs). The images from left to right are a scene from Angeles National Forest, stereo range, and a single frame terrain map. Brown represents the load bearing surface, green indicates a positive obstacle, blue represents the minimum height of the canopy, and white indicates the cells where the canopy is low enough to be a hazard to the vehicle.

## 6.2 Terrain mapping and reflections in water bodies

Water bodies are challenging terrain hazards for several reasons. Traversing through deep water bodies could cause costly damage to the electronics of UGVs. Additionally, a UGV that is either broken down or stuck in a water body during an autonomous military mission may require rescue, potentially drawing critical resources away from the primary mission and soldiers into harms way. Thus, robust water detection is a critical perception requirement for UGV autonomous navigation. JPL is currently working on a UGV water detection task under the RCTA program [21].

Water bodies reflecting terrain in the background are more challenging to represent in a terrain map than the surrounding ground surface. This is illustrated in Figure 13. The stereo range to a reflection of a tree in a water body will be farther than the range to the surface of the water body. As a result, stereo range data on reflections plot below the ground surface. Stereo range data that lie below the detected ground surface are a strong cue for water. This stereo range data can be modified to correspond to the surface of the water body in a terrain map. The elevation of the surface of the water body can be estimated by averaging the elevation of the terrain around the perimeter of the detected terrain reflection. In the lower left world map in Figure 13, the blue map cells indicate the location of the detected water. Stereo reflection range data can corrupt the ground cover and canopy elevation maps if it is not filtered or modified to correspond to the surface of the water body. Figure 14 illustrates what a single frame terrain map can look like when it is corrupted by stereo reflection range data. The elevation of the water is incorrect, a portion of the classified ground cover is below the ground (out of view in the map), and a portion of the ground is classified as canopy. In the single frame terrain map generated with the modified stereo reflection range data, the two water bodies are correctly located.

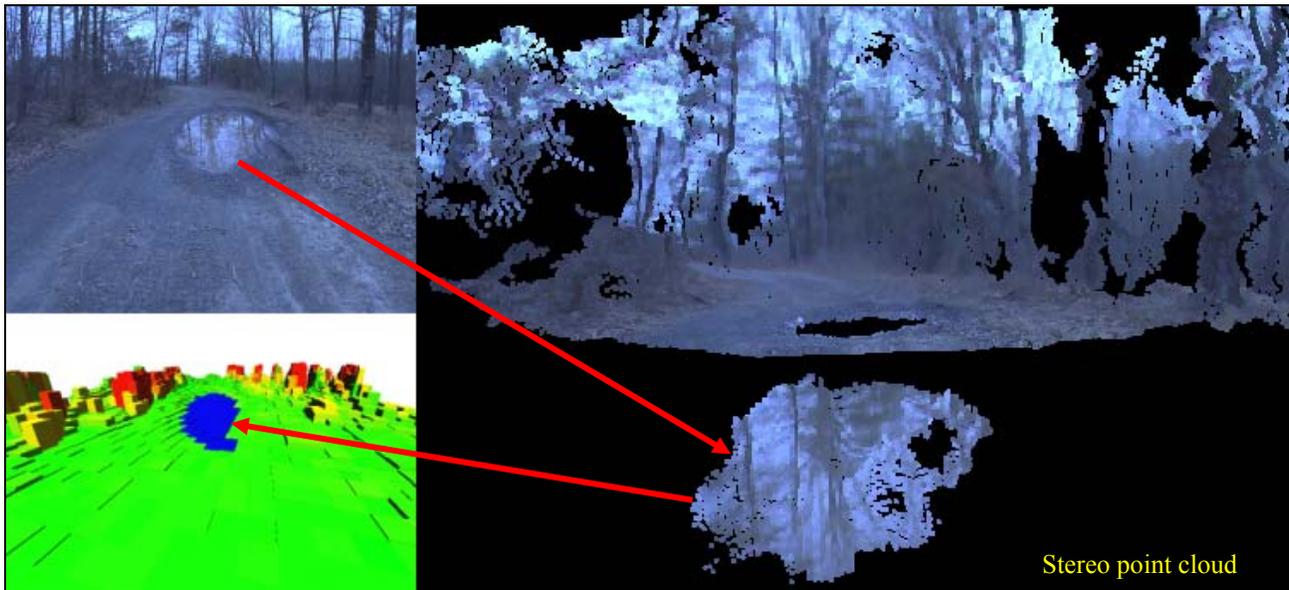


Fig. 13. Representing water bodies reflecting objects in the background poses a larger challenge than representing the ground surface. The stereo range to a reflection of a tree in a water body will be farther than the range to the surface of the water body. As a result, stereo range data on reflections plot below the ground surface. Stereo range data that lie below the detected ground surface are a strong cue for water. This range data can be corrected to correspond to the surface of the water body in a terrain map. In the lower left world map, blue represents the detected water.

## 6.3 Terrain mapping and sky pixels

Our stereo range images are registered with the rectified images from the left camera. Thus, it is trivial to transfer color information from image space to map space. Each world map cell maintains the average RGB values from the last frame of data that contributed to that cell. The upper left image in Figure 15 shows the average terrain color in a world map for a scene containing a long trail, white rocks, vegetation, and a tree line. Color adds a tremendous amount of context to the vehicle's surroundings in a terrain map. Note, however, that some blue sky pixels in the distance were placed into the world map in the lower left image of Figure 15. It was after we added color fields to our world map that we observed sky pixels were corrupting the world map. As seen at the top of the stereo point cloud in Figure 15, the stereo

correlator can match on sky pixels, mostly where the stereo correlation window (nominally 7x7 pixels) spans both the sky and the foreground object.

Under RCTA, JPL has implemented a sky detection algorithm that is used, in part, to suppress placing sky stereo range and color data into the world map. The sky detector starts at the top of the left rectified RGB image and searches down each column until the horizon is reached. Pixels with colors consistent with the sky are selected and a flood fill algorithm grows these detected sky pixels until an edge boundary is found. A sample sky detection result is shown in the middle image of the bottom row in Figure 15. A binary sky detection image is passed to the world map as an “exempt” image. Stereo range and color information are not placed in the world map for exempt pixels. The lower right image in Figure 15 contains the world map generated for the same sequence when the sky detection algorithm is enabled. Much of the sky data is eliminated. There are some sky pixels in between the trees on the right side of the scene that escaped detection by the sky detector. More work is required to improve the detection of sky pixels in between closely spaced trees.

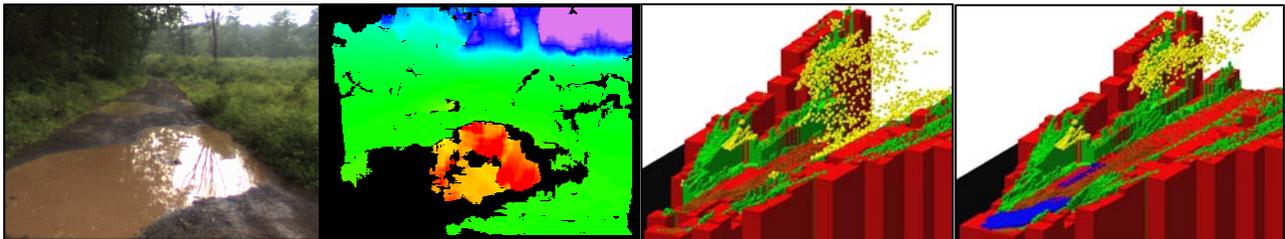


Fig. 14. Reflection range data corrupts load-bearing surface elevation maps if it is not filtered or corrected. The images from left to right are scene from Ft. Indiantown Gap containing a water body, a false color height image from stereo range data, and two single frame terrain maps encoded with elevation and terrain classification. In the elevation image, red corresponds to low elevation, blue corresponds to high elevation, and the colors in between correspond to an intermediate elevation. The left map was generated without filtering or modifying the reflection range data. The right map was generated with modified reflection range data. In the maps, blue represents water, green represents vegetation, yellow represents canopy, brown represents soil, and the red cells have an unknown classification.

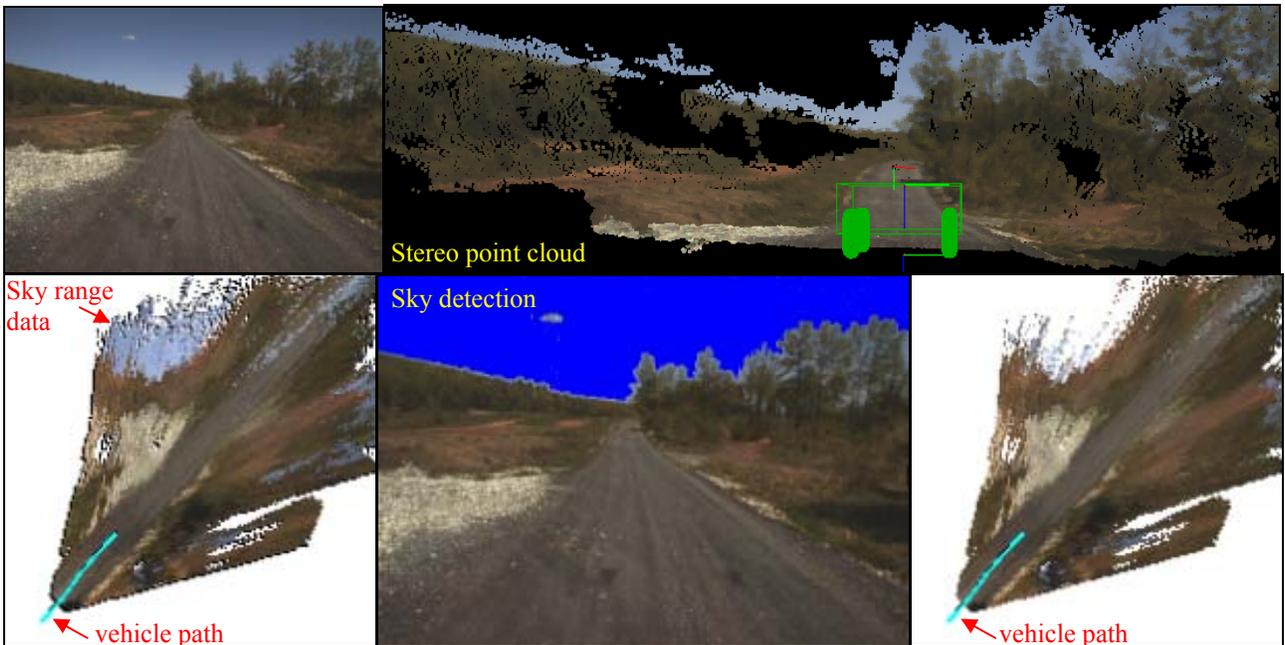


Fig. 15. Sky detection is used to minimize placing sky pixels into the world map. The upper right image contains a stereo point cloud for the scene in the upper left image. The lower left image is bird’s-eye view of a 100m x 100m world that contains RGB color fields. The lower middle image illustrates the results of our sky detection algorithm. The world map in the lower right image was generated with the sky detection algorithm enabled.

## 6.4 Terrain mapping in cluttered environments

Forested scenes with undergrowth are challenging to represent in a terrain map, particularly when there is close undergrowth that partially or fully occludes driving hazards, such as fallen trees, and traversable terrain. Figure 16 illustrates a cluttered scene from a site where the PerceptOR Raptor UGV was evaluated [18]. This scene contains close undergrowth which is difficult to correlate with a high-resolution wide-baseline stereo system. In section 4, we discussed processing scenes containing thin structure at lower resolution to increase the density of a stereo range image. An alternative is to use a multi-baseline stereo system. A daytime multi-baseline stereo system was implemented by JPL during PerceptOR with three color cameras that included 9.5cm, 20.5cm, and 30cm stereo baselines. Example narrow and wide baseline stereo range images, single frame terrain maps, and world maps are shown in Figure 17 for the scene in Figure 16. The maps encode elevation and discrete obstacles. In the maps, orange represents high density obstacles, green represents positive obstacles, red represents tree trunk detection, and blue represents low overhangs. The log is detected in the maps generated from narrow baseline stereo, but not the ones generated from wide baseline stereo.



Fig. 16. The PerceptOR Raptor UGV was operated on a variety of terrain types, including forested terrain at Ft. A. P. Hill. A multi-baseline stereo system was used that could provide wide, mid, or narrow baseline stereo range data, depending upon how cluttered the scene was.

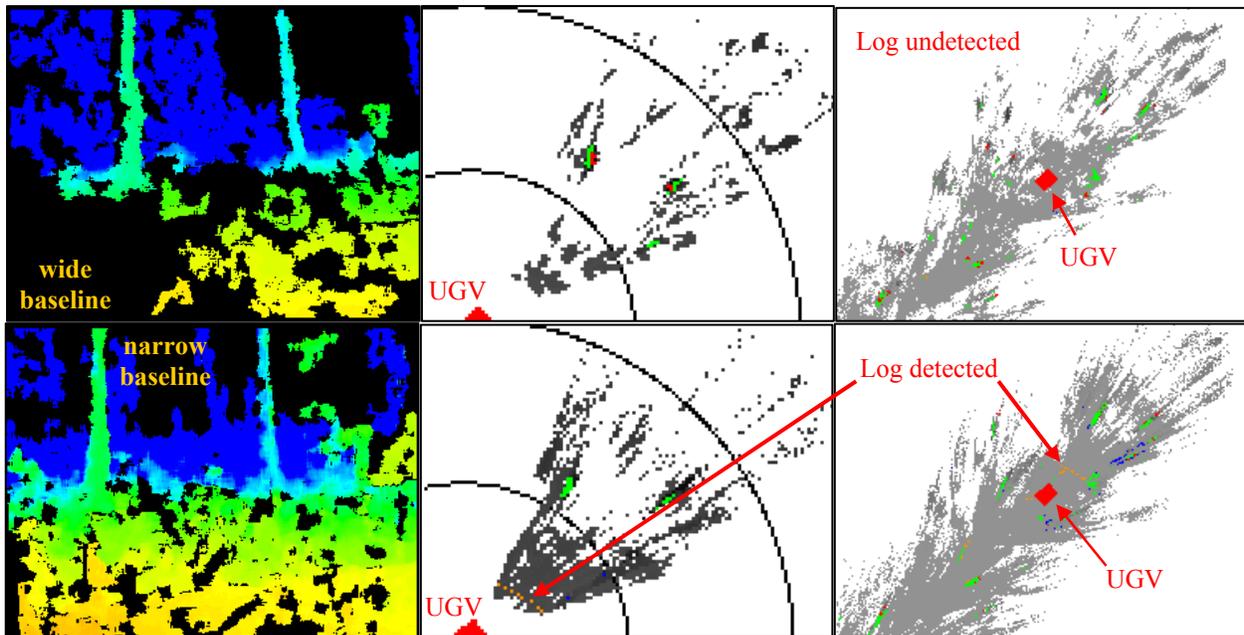


Fig. 17. Reducing the stereo baseline can increase the quantity of stereo range data for a cluttered scene, enabling the detection of partially occluded drive hazards. In the top row from left to right are a wide-baseline stereo range image and the corresponding single frame terrain map and world map. In the bottom row from left to right are a narrow baseline stereo range image and the corresponding single frame terrain map and world map. The maps encode elevation and discrete obstacles. In the maps, orange represents high density obstacles, green represents positive obstacles, red represents tree trunk detection, and blue represents low overhangs. The log is detected in the maps generated from narrow baseline stereo, but not the ones generated from wide baseline stereo.

## 7. CONCLUSIONS

The ability to generate a reliable representation of natural terrain in real time using perception sensors mounted to a UGV is critical to successful UGV autonomous navigation on off-road terrain. Over the last two decades, active and passive ranging systems have been successfully used to construct a variety of terrain maps for UGV autonomous navigation. Ladar, however, continues to provide higher quality range data than fixed baseline stereo, but the maximum range of ladar, its relatively coarse angular resolution, and its electromagnetic signature are all limiting factors. JPL has been involved in research to improve the quality of stereo range data under several DoD programs during the last decade including Demo III, PerceptOR, and RCTA. Research efforts under these programs with high-resolution wide-baseline stereo, multi-baseline stereo, multi-resolution stereo, edge preserving prefilters, and advanced correlators have progressively led to higher quality stereo range data being available for terrain mapping. Color cameras have been used to perform stereo ranging during the daytime and thermal infrared cameras have been used to perform stereo ranging during the day and nighttime.

In this paper, we have summarized the Cartesian terrain mapping capabilities developed by JPL during the Demo III, PerceptOR, and RCTA programs as the quality of stereo range data has progressively improved. During Demo III and PerceptOR, passive perception subsystems were fielded that generated a compact north-oriented 2.5D terrain map for each pair of stereo images that were processed. These single frame terrain maps were shipped over a LAN to a world model subsystem. Each map cell contained bit fields that encoded terrain elevation, terrain classification, object classification, terrain roughness, and confidence values within four bytes of memory. The map origin was selected so that the single frame terrain maps correctly lined up with a portion of the world map, facilitating the merging of map data into a single composite map. While encoding maps cells with terrain elevation, terrain classification, and detected obstacles is a reasonable way to represent unsafe regions, it does not indicate how traversable non-obstacle regions are. In addition, the four state terrain roughness measure did not have enough resolution to adequately represent traversability. Therefore, during RCTA we added an 8 bit traversability cost field to the single frame terrain map data structure, maintaining the four byte per cell assignment by reducing the resolution of other bit fields.

Stereo range data has different noise characteristics than ladar. To trap stereo related noise before it reaches the vehicle-level world model, we also implemented a scrolling, north-oriented, vehicle-centered world map (using a map base class provided by GDRS). The world map is 100m x 100m, with each cell having a 40cm resolution. Temporal filtering was performed in the world map to smooth out noisy elevation and traversability cost data from the single frame terrain maps prior to delivery to the world model. In addition, several algorithms were implemented to extract terrain characteristics from the world map, such as the ground surface. On a 2.1GHz Pentium M processor, end-to-end SAD5 stereo ranging and terrain mapping was performed on 512x384 and 1024x768 resolution images at 2.5fps and 0.54fps, respectively.

One way to increase the speed of terrain mapping is to reduce the quantity of stereo range data. When it is computationally too expensive to perform stereo processing on full resolution imagery, portions of the terrain that is far away can be processed at high resolution and portions of the terrain close to the UGV can be processed at lower resolution. On a 2.1GHz Pentium M processor, end-to-end SAD5 stereo ranging was performed on 1024x768 resolution images at 0.85fps. On a slightly slower processor (1.4GHz Pentium M), end-to-end SAD5 four-level multi-resolution stereo ranging was performed on the same images at 16.9fps. On a 3.6GHz Pentium 4 processor, end-to-end SAD5 four-level multi-resolution stereo ranging was performed on the same images at 24.2fps. (These benchmarks date back to the 4<sup>th</sup> quarter of FY06, when this work was completed. On currently available hardware, there has been a significant speed up in stereo and map processing. For example, end-to-end SAD1 and SAD5 stereo ranging is currently performed on 1024x768 resolution images at 8fps and 3fps, respectively, on an Intel core 2 quad processor.) Generating world maps with four-level multi-resolution stereo range data resulted in a slight decrease in elevation fidelity in the near field, a slight decrease in the number of map cells with data, but improved modeling of thin structure in some scenes. We also experimented with generating terrain maps at lower resolution. Since the height accuracy of stereo range data is significantly better than the range accuracy at far range, low resolution terrain maps can be generated to perceive long stretches of straight roads.

In the section 6 of this paper, we addressed several complications to mapping with stereo range data. Low overhangs are particularly worrisome since an undetected overhang, such as a tree branch, can potentially damage critical UGV components. We assume a bimodal distribution in each cell and cluster the data to the ground and hypothetical canopy, clipping any data above the UGV height and labeling map cells that have a detected canopy lower than the UGV height as low-overhang obstacles. Stereo range data on reflections of terrain in water bodies plot below the ground surface and

can corrupt a terrain map if not modified or removed. We have implemented a stereo reflection detector that modifies the reflection range data to correspond to the surface of the detected water body. Another complication to terrain mapping occurs at the border of terrain and the sky. Here, the stereo correlator may match on sky pixels, producing erroneous range measurements that can corrupt a terrain map. To suppress placing erroneous sky range data within a terrain map, a sky detector is run on input color imagery and stereo range data for a detected sky pixel is ignored. For cluttered environments such as forests, the range data from a wide-baseline stereo system (and terrain maps generated from it) can be rather sparse. To address this issue, we have experimented with a multi-baseline stereo system that can be switched to process narrow-baseline stereo imagery when a UGV is in a cluttered environment.

Our expectation is that the quality of stereo range data will continue to improve as advances in computing hardware enable advanced stereo algorithms to be performed in real-time, further closing the quality gap between stereo and lidar. With current programs such as FCS having a requirement to be capable of performing UGV autonomous navigation independently with active and passive sensors, stereo ranging and terrain mapping continue to be critical research areas. JPL is currently expanding its UGV terrain mapping capability under the FCS program.

### ACKNOWLEDGEMENTS

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by SAIC under the DARPA PerceptOR program and by ARL under the Demo III and RCTA programs, through agreements with the National Aeronautics and Space Administration (NASA). Reference herein to any specific commercial product, process, or service by trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

### REFERENCES

- [1] A. Kelly, "An intelligent, predictive control approach to the high-speed cross-country autonomous navigation problem", Ph.D. Thesis, Carnegie Mellon University, (1995).
- [2] L. Matthies et. al., "A portable autonomous urban reconnaissance robot," *Proceedings of the Sixth International Conference on Intelligent Autonomous Systems*, Venice, Italy, (2000).
- [3] M. Hebert and E. Krotkov, "Local perception for mobile robot navigation in natural terrain: two approaches", *Workshop on Computer Vision for Space Applications*, Antibes, 24-31, (1993).
- [4] H. Moravec, "Robot spatial perception by stereoscopic vision and 3D evidence grids", Technical Report CMU-RI-TR-96-34, CMU Robotics Institute, (1996).
- [5] J. Carsten, D. Ferguson, and A. Stentz, "3D Field D\*: Improved path planning and replanning through three dimensions", *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China, 3381-3386, (2006).
- [6] A. Lacaze, K. Murphy, and M. DelGiorno, "Autonomous mobility for the demo III experimental unmanned vehicles", *Proceedings of the AUVSI Conference*, Orlando, (2002).
- [7] A. Kelly, et. al., "Toward reliable off road autonomous vehicles operating in challenging environments," *The International Journal of Robotics Research*, 25(5-6), 449-483, (2006).
- [8] A. Rankin, A. Huertas, and L. Matthies, "Evaluation of stereo vision obstacle detection algorithms for off-road autonomous navigation", *Proceedings of the 32<sup>nd</sup> AUVSI Symposium on Unmanned Systems*, Baltimore, (2005).
- [9] J. Collier, G. Broten, and J. Giesbrecht, "Traversability analysis for unmanned ground vehicles", Defence Research and Development Canada, Suffield, Technical Memorandum 2006-175, (2006).
- [10] J. Gu, Q. Cao, and Y. Huang, "Rapid traversability assessment in 2.5D grid-based map on rough terrain", *International Journal of Advanced Robotic Systems*, 5(4), 389-394, (2008).
- [11] T. Chang, T. Hong, S. Legowik, and M. Abrams, "Concealment and obstacle detection for autonomous driving", *Proceedings of the IASTED Conference on Robotics and Applications*, Santa Barbara, CA, (1999).
- [12] I. S. Kweon and T. Kanade, "High-resolution terrain map from multiple sensor data", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), 278-292, (1992).
- [13] T. Hong, S. Balakirsky, E. Messina, T. Chang, and M. Shneier, "A hierarchical world model for an autonomous scout vehicle", *Proceedings of the SPIE 16<sup>th</sup> Annual International Symposium on Aerospace/Defense Sensing, Simulation, and Controls*, 343-354, (2002).

- [14] G. Broten, J. Giesbrecht, and S. Monckton, "World representation using terrain maps", Defence Research and Development Canada, Suffield, Technical Memorandum 2005-248, (2005).
- [15] G. Broten, S. Monckton, J. Collier, and D. Mackay, "Terrain maps: lessons learned and new approaches", Defence Research and Development Canada, Suffield, Technical Memorandum 2006-189, (2006).
- [16] M. Hebert, "Active and Passive Range Sensing for Robotics", *Proceedings of the IEEE International Conference on Robotics and Automation*, San Francisco, 102-110, (2000).
- [17] P. Bellutta, R. Manduchi, L. Matthies, K. Owens, and A. Rankin, "Terrain perception for demo III", *Proceedings of the IEEE Intelligent Vehicle's Symposium*, Dearborn, MI, 326-331, (2000).
- [18] A. Rankin, C. Bergh, S. Goldberg, P. Bellutta, A. Huertas, and L. Matthies, "Passive perception system for day/night autonomous off-road navigation", *SPIE Defense and Security Symposium: Unmanned Ground Vehicle Technology VI Conference*, Orlando, 343-358, (2005).
- [19] A. Ansar, A. Castano, and L. Matthies, "Enhanced real-time stereo using bilateral filtering", *Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission*, Thessalonica, Greece, (2004).
- [20] A. Ansar, A. Huertas, L. Matthies and S. Goldberg, "Enhancement of stereo at range discontinuities", *Proceedings of the SPIE Defense and Security Symposium: UGV Technology VI Conference*, Orlando, (2004).
- [21] A. Rankin and L. Matthies, "Daytime water detection and localization for unmanned ground vehicle autonomous navigation", *Proceedings of the 25th Army Science Conference*, Orlando, (2006).
- [22] A. Rankin and L. Matthies, "Daytime mud detection for unmanned ground vehicle autonomous navigation", *Proceedings of the 26th Army Science Conference*, Orlando, (2008).
- [23] M. Bajracharya, B. Moghaddam, A. Howard, and L. Matthies, "Detecting personnel around UGVs using stereo vision", *Proceedings of the SPIE Defense and Security Symposium: Unmanned Systems Technology X Conference*, Vol. 6962, Orlando, (2008).