Methods for Evaluating the Performance of Unmanned Ground Vehicle Water Detection

Arturo L. RANKIN, Tonislav IVANOV and Shane BRENNAN

Abstract- **Water detection is a critical perception requirement for unmanned ground vehicle (UGV) autonomous navigation over cross-country terrain. Under the Robotics Collaborative Technology Alliances (RCTA) program, the Jet Propulsion Laboratory (JPL) developed a set of water detection algorithms that are used to detect, localize, and avoid water bodies large enough to be a hazard to a UGV. The JPL water detection software performs the detection and localization stages using a forward-looking stereo pair of color cameras. The 3D coordinates of water body surface points are then output to a UGV's autonomous mobility system, which is responsible for planning and executing safe paths. There are three primary methods for evaluating the performance of the water detection software. Evaluations can be performed in image space on the intermediate detection product, in map space on the final localized product, or during autonomous navigation to characterize the avoidance of a variety of water bodies. This paper describes a methodology for performing the first two types of water detection performance evaluations.**

Index Terms—**Water detection, stereo vision, terrain perception, and ground truth.**

1. INTRODUCTION

Detecting water hazards is a significant challenge to unmanned ground vehicle (UGV) autonomous navigation over cross country terrain. Traversing through deep water bodies could cause costly damage to the electronics of UGVs. Moreover, a UGV that either breaks down due to water damage or becomes stuck in a water body during an autonomous military mission could cause further complications. These include the shifting of critical resources away from the primary mission to a rescue mission, the placing of soldiers into harm's way to support a rescue mission, the loss of advanced technology to an enemy, and mission failure.

Under the Robotics Collaborative Technology Alliances (RCTA) program, several researchers developed methods for water detection under the advanced perception technology thrust [5, 7, 8, 12-17]. The Jet Propulsion Laboratory (JPL) participated in this effort, focusing primarily on the cues for water that can be exploited from a stereo pair of color cameras mounted to the front of a UGV. Early in the program, JPL developed an all-purpose multi-cue water detector that uses a rule base to combine water cues from color, texture, and object reflections (detectable in dense stereo range data) [16]. Subsequently, JPL developed three specialized stand-alone water detection algorithms (also using a forward-looking stereo pair of color cameras) to handle three general scenarios: water bodies in cluttered environments that are reflecting objects in the background (such as trees) [14], water bodies that are out in the open and far away (where reflections of the sky dominate) [15], and water bodies out in the open and close to the UGV (where the color coming out of the water body dominates) [13]. A summary of JPL's four water detectors is in [10].

All four water detection algorithms operate within image space during autonomous navigation and the results are fused into a single terrain classification image. Dense stereo range data is then used to localize detected water in a digital terrain map [12]. The 3D coordinates of water body map cells are then output to a UGV's autonomous mobility system, which is responsible for planning and executing safe paths.

There are three primary ways to evaluate the performance of the JPL water detection software. Evaluations can be performed in image space on the intermediate detection product, in map space on the final localized product, or during autonomous navigation to characterize how well detected, localized water bodies are avoided. The first two types of performance evaluations are unit-level since they characterize a single subsystem (i.e., the water detection subsystem). The last type of performance evaluation is system-level since it characterizes the combined water detection subsystem and a UGV's response to its output. All three evaluation methods have been utilized during the RCTA program. At the close of the RCTA program, the Army Research Laboratory (ARL) designed and executed a system-level experiment to test the avoidance of hazardous terrain (including water bodies) during autonomous navigation [3]. The portion of the experiment that included water body avoidance was performed on experimental unmanned vehicles (XUVs) at Fort Indiantown Gap (FITG), PA. Fig. 1 shows an XUV and highlights the sensors used for water detection.

ARL has already published several papers that describe how system-level hazard avoidance evaluations were performed during RCTA [2-3]. In this paper, we focus on unit-level performance evaluations. We describe a methodology for conducting water detection performance evaluations in image and map space and water localization performance evaluations in map space. These methods are used to answer three primary questions; was a water body

Manuscript received December 10, 2010; revised April 11, 2011. This work was funded by the U.S. Army Research Laboratory under the Robotics Collaborative Technology Alliances program. This paper is extended from "Evaluating the performance of unmanned ground vehicle water detection" published at Performance Metrics for Intelligent Systems Workshop, Baltimore, MD, USA, Sep., 2010.

A. Rankin, T. Ivanov, and S. Brennan are all with the Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California, USA 91109 (e-mail: Arturo.Rankin@jpl.nasa.gov, Tonislav.Ivanov@jpl.nasa.gov, and Shane.Brennan@jpl.nasa.gov).

detected, how much of it was detected, and was it detected in the correct location?

Fig. 1. JPL water detection software has been integrated onboard XUVs and evaluated at Ft. Indiantown Gap, PA. A forward-looking stereo pair of color cameras provides water cues from color, texture, and object reflections.

2. WAS IT DETECTED?

This type of performance evaluation can be performed in image or map space. The JPL water detection software detects water bodies in image space and localizes them in map space. Water detection results from each frame are merged into a world map that is a part of the water detection software. The world map is north oriented and vehicle centered, having a size of 100 meters squared and a resolution of 0.4 meters. This means that as the vehicle moves, new data is added to one side of the map as old data drops off the other side of the map. Although a history of water detection information from previous frames is maintained in the world map, a detected water body is not output by the world map to an XUV's autonomous mobility (AM) system until it has been "seen" at least twice. We answer the question "was it detected?" by evaluating the world maps being output from the water detection software to the AM system.

A forward-looking stereo pair of color cameras is used to build the water detection world map. The stereo cameras are calibrated using a planar target having a fixed pattern of 80 equally spaced dots of know diameter. Synchronized left and right camera images of the target are acquired with the target positioned to cover portions of the field of view at a range of distances from the cameras. Once calibration imagery is acquired, it is processed with JPL dot finding software that generates left and right CAHVOR camera models. CAHVOR camera models precisely describe how 2D image coordinates map to 3D space, and vice versa. They are used in converting stereo disparity to stereo range data [6, 18].

As seen in Fig. 1, the stereo cameras straddle an XUV lidar sensor, which is mounted on a pan-tilt unit (PTU). An XUV's navigation sensors are a Smiths Industries (model #9181C 2 MRADIAN) inertial reference unit (IRU), a NovAtel Propak-LB+HP-RT2 global positioning system (GPS) receiver, a precision lightweight GPS receiver (PLGR+96 GPS RECEIVER-PPS), and wheel encoders. The PLGR supplies the IRU with GPS and heading data. The NovAtel GPS receiver, used with an OmniStar subscription GPS package, yields a horizontal error of approximately 6 cm and a 99% horizontal error of less than 10 cm. The navigation system generates relative and absolute position (x, y, z) coordinates of the vehicle origin (midpoint of the two axles at ground height). The IRU provides continuous XUV orientation angles (roll, pitch, yaw). A camera-to-vehicle calibration is required to transform stereo range data from the camera frame to the vehicle frame.

As illustrated in Fig. 2, the camera-to-vehicle calibration is performed with an XUV on a surface known to be level. The (x, y, z) offsets between the vehicle origin and the intersection of the pan and tilt axes are either measured by hand, measured with the aid of a total station surveying instrument, or obtained from a computer aided design (CAD) model of the vehicle. The (x, y, z) offsets from the intersection of the pan and tilt axes to the left camera focal plane array (FPA) are estimated. The tilt offset between the left camera and the lidar is measured with a digital level accurate to within 1/100th a degree.

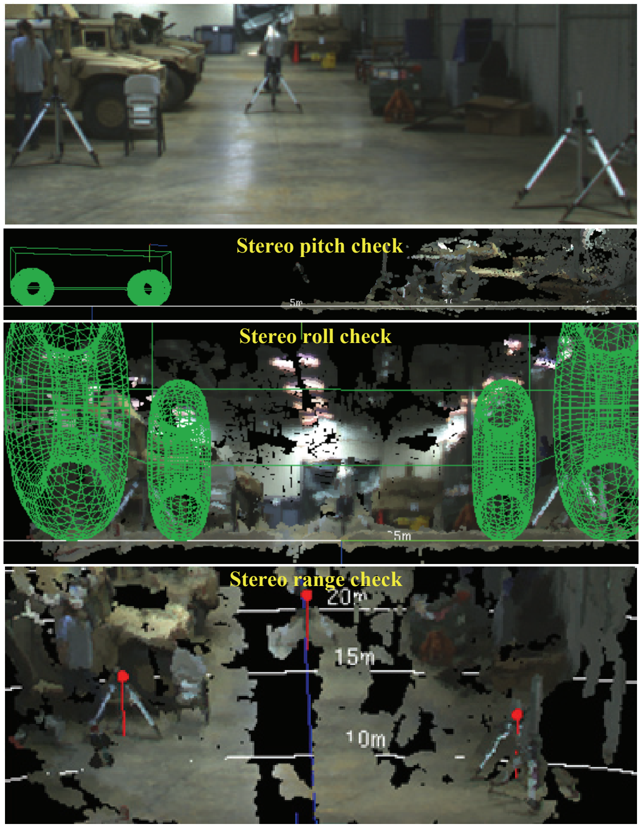


Fig. 2. A camera to vehicle calibration is performed on a surface known to be level. The vehicle is centered on a line on the floor. The camera to vehicle calibration parameters are adjusted until the pitch and roll of the floor are level in the stereo point cloud, and tripods surveyed by a total station are at the correct position in the stereo point cloud. The red markers in the bottom image correspond to the surveyed position of three tripods in the top scene.

Once the camera calibration and camera-to-vehicle calibration are complete, the camera-to-vehicle calibration parameters can fine tuned until the pitch and the roll of the floor are level in the stereo point cloud, and objects surveyed with a total station are in the correct position in the stereo point cloud. In the bottom image of x, the red markers correspond to tripod positions surveyed in the vehicle frame.

A performance evaluation has been performed on 12,265 color stereo image pairs collected during tele-operation of an XUV around a 6.9km test course at Fort Indiantown Gap (FITG) called *Forever Loop*. As illustrated in Fig. 3, there were three water bodies were on the course the day of the data collection. During the data collection, vehicle state data was recorded in the header of each stereo image.

Fig. 4 contains a plot of the XUV's relative positioning data (*TranRel*), which illustrates the shape of the *Forever Loop* course. The alternating blue and magenta colors indicate where the XUV stopped so that a new data log could be started. Note that the data logger failed to restart at one location, causing a ~100m gap in the data. The course start and end points are very close, indicating a very low drift rate in the navigation solution.

All three of the puddles on the course were consistently detected by the stereo reflection detector. The two small traversable puddles were detected in 12 consecutive world maps, starting at a distance of 7 meters. The larger traversable puddle was detected in 59 consecutive world maps, starting at a range of 13 meters. As soon as there was stereo range data on the reflections, the reflections were detected.

Fig. 3. There were three water bodies were on the *Forever Loop* course the day of the data collection; two small puddles and one large puddle.

False positive water detection occurred in 26 of the 12,265 world maps (0.21%). In 24 of these maps, the false positive water detection occurred on non-traversable log barriers lining the sides of a trail. In the other 2 world maps, the false positive water detection occurred in tall vegetation lining the side of trails. All of the false positive water detection occurred in areas that were either unsafe or less ideal than neighboring terrain.

3. HOW MUCH WAS DETECTED?

3.1 Ground Truthing Water Regions in Image Space

The JPL water detection software detects water in image space. Therefore, the ideal place to evaluate the performance of JPL water detection is in image space. An evaluation of a terrain classifier is typically performed by accurately specifying the pixels that belong to the class of interest and generating a Receiver Operating Characteristic (ROC) curve which plots the true positive classification rate against the false positive classification rate.

In the past, ground truthing water pixels for image sequences (potentially containing hundreds of images) has been a very tedious process. In each image in a sequence, the perimeter of a water body was manually segmented by moving the mouse cursor around the water perimeter and "clicking" (i.e., selecting) a limited number of vertices. The vertices were then connected with straight lines and all the pixels within these N-sided polygons were labeled water. However, since water bodies are not constrained to follow straight lines, there is some inherent ground truth error with this method. Increasing the number of vertices around the perimeter of a water body will reduce the error but will also increase the time required to manually ground truth a sequence. Since the boundary of a water body expands in image space as it is approached, new vertices needed to be selected for each subsequent frame.

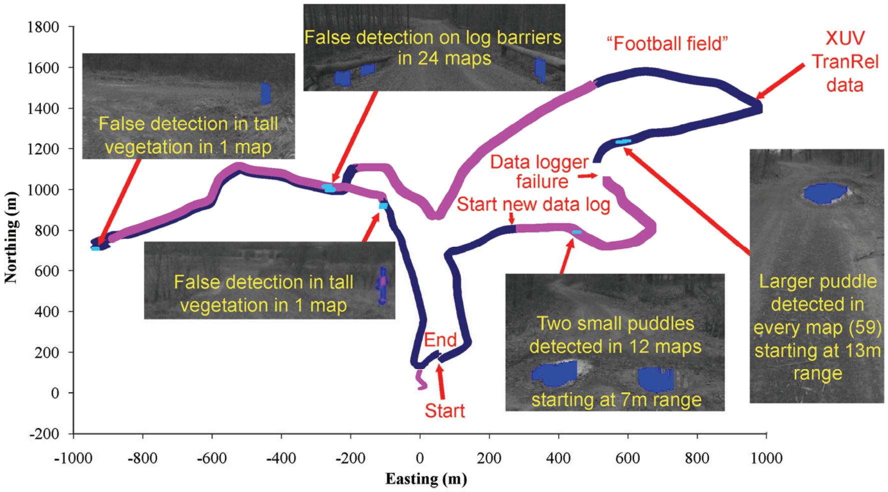


Fig. 4. Water detection software was evaluated on color stereo imagery collected during UGV tele-operation of a 6.9km test course at FITG called *Forever Loop*. Three water bodies were on the course the day of the data collection; two small puddles and one large puddle. 12,265 stereo pairs of images were processed with the terrain reflection based water detector, but only detected water seen in more than one image was reported in the world map. The two small puddles were detected in 12 consecutive maps starting at a range of 7 meters. The large puddle was detected in 59 consecutive maps starting at a range of 13 meters. The false positive water detection rate was 0.2%.

To automate the process of ground truthing water bodies in sequences of images, JPL has developed a software tool using Open Graphics Library (OpenGL). In the first image of a sequence, vertices still need to be manually selected. But in each subsequent image, their 2D image coordinates are automatically updated. In the first frame, since we know the surface of a water body is horizontal, we estimate the elevation of a water body by averaging the elevation of each selected vertex using stereo range data. Given the 2D image coordinates of a vertex, it is trivial to look up its 3D coordinates since a left rectified image is registered with the corresponding stereo range image. The stereo correlator may fail to produce disparity data for some of the vertices. For vertices that have stereo 3D coordinates, their vectors are scaled so that they terminate in the estimated water surface plane. For vertices that have no stereo 3D coordinates, a left CAHVOR camera model [6, 18] is used to modify the rays extending from the vertex pixels so that they also terminate in the estimated water surface plane. The new 3D coordinates of each vertex are recorded for use in subsequent frames. Each time a vertex is modified or added, the corresponding 3D coordinates are generated using the above procedure.

Since UGV position (x, y, z) and orientation data (roll, pitch, yaw) are recorded for each frame in a data set, the 3D coordinates of each vertex can be expressed in a gravity-based world coordinate frame. Given the motion of the UGV from one frame to the next, the left CAHVOR camera model is transformed and used to perform a linear mapping between the 3D world coordinates of the vertices and their 2D image coordinates [6, 18]. Since the perimeter of a water body typically has texture in visible imagery, it is unlikely that the stereo correlator would fail to produce disparity data for all vertex image coordinates. We currently do not attempt to filter out noisy stereo range measurements from the water body elevation estimate, but a random sample consensus (RANSAC) algorithm could be incorporated to do this.

Fig. 5 illustrates the vertex selection process in the first frame of a sequence. Here, a large number of vertices were selected and neighbor vertices were connected with lines to label the water body. The "Action" and "View" drop down menus are shown to illustrate some of the options available to the user. Vertices can be added, moved, or deleted, and entire polygons can be moved or deleted. In addition, vertices and labeled water regions can be saved to file on a frame-by-frame basis and subsequently loaded from file to play back the results, or to further refine the labeling.

Vehicle motion causes the image coordinates of vertices along the perimeter of a water body to change. For example, as a water body is approached, it appears larger in image space. Fig. 6 shows a sample result of automatically updating the image coordinates of the perimeter vertices after 47 frames of XUV motion towards the water body. The top row contains the left color image, the right color image, and the stereo range image for frame 0. The middle row contains the same images for frame 46. The bottom image contains a stereo point cloud for frame 46. Stereo was performed at full image resolution (1024x768 pixels). The ground truth region is overlaid on the left images and the stereo point cloud in yellow. In the stereo point cloud, the orange and white areas indicate where there is ground truthing error.

We implemented the following three strategies to help minimize ground truthing error:

1) Allow the user to step through the sequence, pausing at every frame and verifying accurate labeling. If the model of the perimeter is poor, allow the user to move, add, or delete vertices.
2) Provide the option of processing the sequence in reverse order from the final image to the first image. In most of our RCTA sequences, a water body is approached from a distance and the final image contains the greatest perimeter detail. This may not be ideal for all sequences, however. In the final frame of some sequences, portions of a water body fall outside of the cameras field of view. When that occurs, additional vertices may need to be manually added as more of the water body comes into the field of view.
3) Perform non-linear segmentation between vertices to improve the modeling of the water body perimeter.

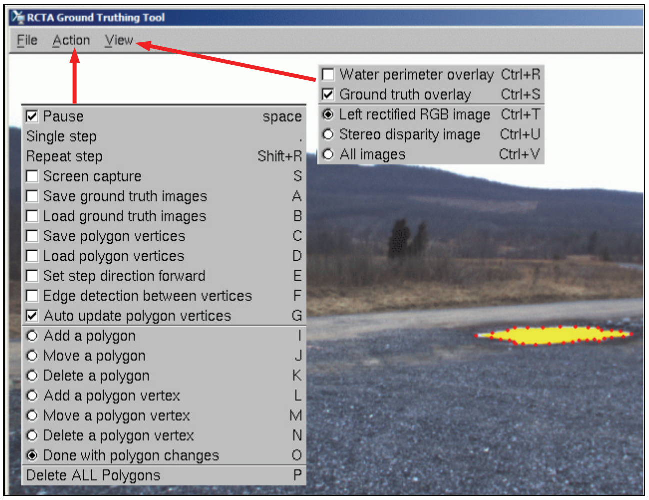


Fig. 5. A software tool has been developed to automate the process of ground truthing water bodies in sequences of images. In the first image of a sequence, vertices around the perimeter of water bodies are selected. In this example, the water body is labeled by connecting neighboring vertices with lines. Vertex selection is performed on rectified, full resolution (1024x768 pixels) imagery. Here, we are only showing a portion of the first image. The full image is shown (at a lower resolution) in the upper left image of Fig. 6.

The algorithm we selected for non-linear segmentation is called *intelligent scissors* [9]. This algorithm attempts to find the most grayscale contrast closed-loop boundary (Laplacian zero-crossing) while keeping the boundary edge smooth (gradient direction) and the texture around the boundary consistent (gradient magnitude). An optimal graph search called *live-wire boundary* is performed based on Dijkstra's [4] path finding algorithm to find a minimal cost path via dynamic programming. The open-source code for *intelligent scissors*, available under GNU Image Manipulation Program (GIMP), was adapted and integrated into the water body ground truthing tool.

For each frame in an image sequence, this code uses the automatically updated vertices from the water region segmented in the previous frame. We step around the ordered vertex list for each water body and run *intelligent scissors* between each vertex pair. *Intelligent scissors* outputs a set of connected pixels between neighbor vertices. Fig. 7 illustrates the advantage of using *intelligent scissors* to model the water perimeter between vertices. In this example, only five vertices were needed along a section of the water boundary that has a length in excess of 20 meters.

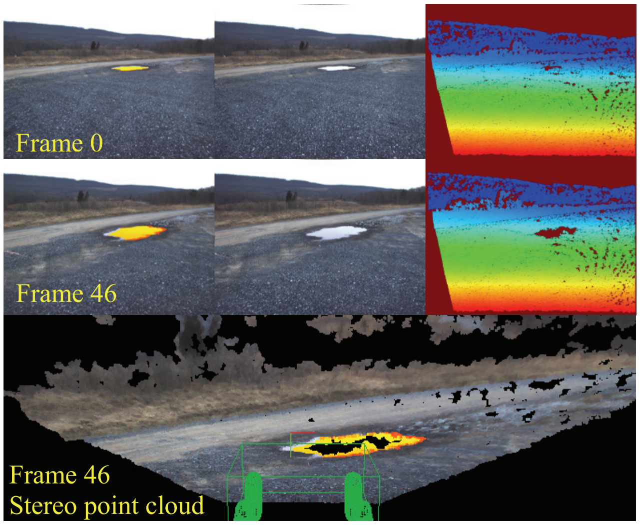


Fig. 6. Sample result of automatically updating the image coordinates of the vertices selected in Fig. 5 after 47 frames. The top row contains the left rectified image, the right color image, and the stereo range image for frame 0. The middle row contains the same images for frame 46. The bottom image contains a stereo point cloud for frame 46. The ground truth region is overlaid on the left rectified images and the stereo point cloud in yellow. In the color-coded stereo range images, red corresponds to close range, blue corresponds to far range, the colors in between correspond to an intermediate range, and maroon corresponds to no stereo data.

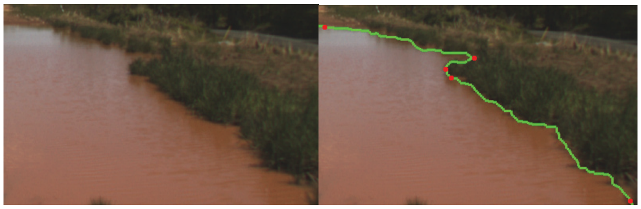


Fig. 7. The GIMP image viewer has an intelligent scissors tool to segment image regions along contours. We have extracted the intelligent scissors portion of GIMP and have integrated it into our ground truthing tool. In this example, only 5 vertices were needed along the cropped portion of the water boundary. Intelligent scissors is run for each pair of neighbor vertices to segment the water's edge.

The JPL water body ground truthing software tool has been used to label one color image sequence containing a single large water body and two color image sequences containing a single small water body. The sequences were collected at FITG (on different days) while an XUV was teleoperated towards each water body. The large water body, partly shown in Fig. 7, is at least partially visible in all 143 frames of the sequence.

Since terrain classification and stereo range images are registered with the corresponding left camera image (after rectification), only the left rectified images are ground truthed. A binary ground truth image is saved for each left rectified image in a sequence. Fig. 8 shows the results of labeling the large water body in the first and last frame in the sequence. The labeling is quite good, but it is not perfect. As illustrated in the bottom row of Fig. 8, it does not exclude interior objects extending out of a water body, such as sediment, vegetation, or rocks.

Camera pose accuracy can be affected by IMU, PTU, and camera calibration error, the resolution of the IMU and PTU, and system latency. Camera pose inaccuracy can translate into stereo range data error, particularly when operating on rough terrain or at high speed. In addition, stereo range accuracy inherently decreases with range as a function of range squared. Both of these issues can contribute to vertices "pulling away" from the boundary of a water body from frame to frame. When a water body is far away, more manual correction of vertex placement is required from frame to frame. However, since a water body is smaller in image space when it is far away, there will be less vertices to correct.

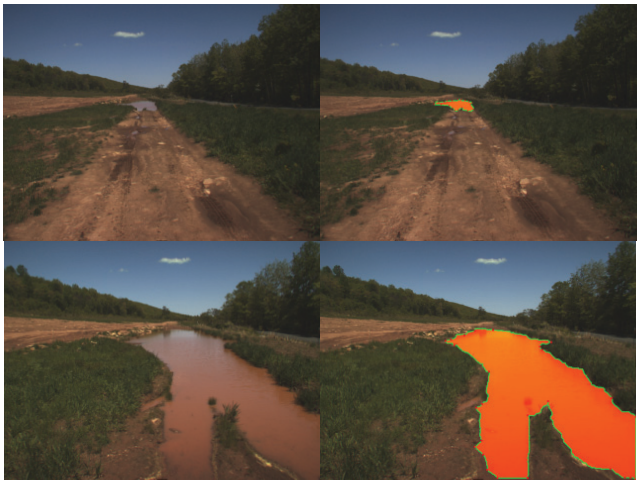


Fig. 8. The water perimeter (green) and labeled water region (orange) for the first and last frame of a 143 frame sequence. This sequence was ground truthed using our software tool. The labeling is quite good, but it is not perfect. For example, it does not exclude interior objects extending out of the water.

3.2 Characterizing Detection Performance in Image Space

The labeled sequence described in Fig. 8 and two other labeled sequences were processed off-line with the

JPL water detection software with all three specialized detectors enabled. All three sequences contained a single water body. In each sequence, the water body was detected in every frame with no false negative water detection (apart from minimal errors in water detection spilling over the ground truthed boundary).

Fig. 9 shows water detection results in the form of overlapping water cues for the first and last frame in the first sequence. The pixels labeled blue, magenta, and red indicate where the water body was detected by one, two, or three water detectors, respectively. The blue regions were detected based on the variation in color [13], the magenta regions were detected based on the variation in color [13] and sky reflections [15], and the red regions were detected based on the variation in color [13], sky reflections [15], and terrain reflections [14].

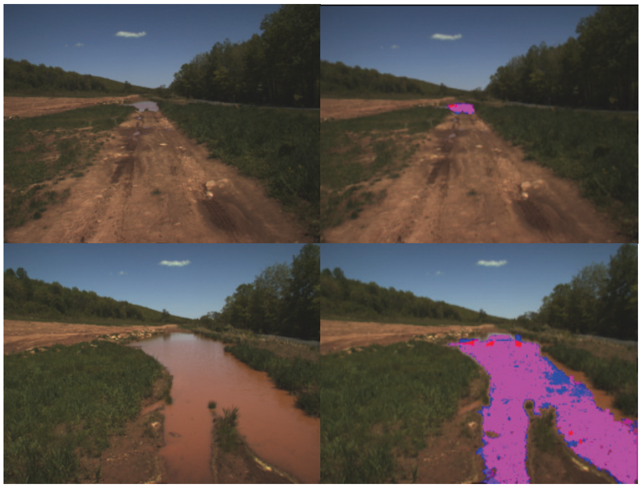


Fig. 9. Overlapping water cues for the first and last frame in the ground truthed sequence. The pixels labeled blue, magenta, and red indicate where the water body was detected by one, two, or three water detectors, respectively. The blue regions were detected based on variation in color, the magenta regions were detected based on the variation in color and sky reflections, and the red regions were detected based on the variation in color, sky reflections, and terrain reflections.

Fig. 10 shows water detection results for the first sequence overlaid on the labeled water body for the first and last frame in the sequence. The blue, red, and green pixels indicate true positive detection, false negative detection, and false positive detection, respectively. Note the detector labels almost the entire water body but tends to miss small portions of water on the perimeter. Along the right side of the water body, the missed detection is due to weak reflections of the short grass lining the water's edge.

The second sequence is a 67 frame approach to a small shallow water body on a gravel surface under an overcast sky. Fig. 11 shows water detection results overlaid on the labeled water body for the first and last frame in the second sequence. At the beginning of the sequence (where the incidence angle is higher), sky reflections dominate the color of the water body. At the end of the sequence (where the incidence angle is lower), the color of the bottom of the water body begins to dominate. As expected, the sky reflection based water detector performed well at higher incidence angles and the color variation based detector performed well at lower incidence angles. The second water body was detected in the first 50 frames by the sky reflection based water detector and the last 45 frames by the color variation based water detector, with an overlap in detection in 32 middle frames.

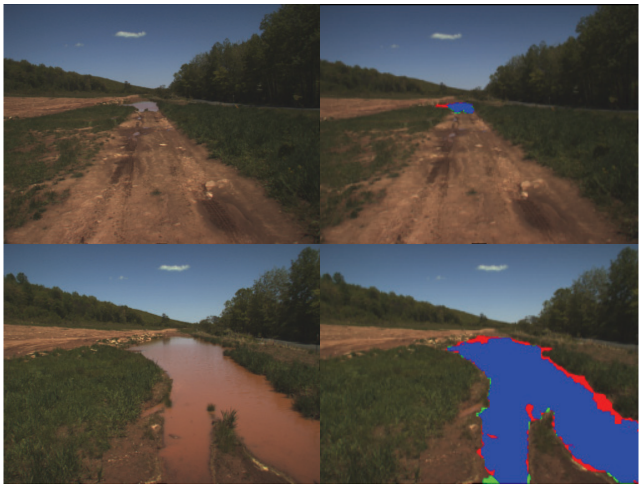


Fig. 10. Water detection results for the first and last image in the sequence. The blue, red, and green pixels indicate true positive detection, false negative detection, and false positive detection, respectively.

When the leading edge of a water body has a very mild slope, it may be difficult to precisely delineate where the surrounding terrain ends and a water body begins. For example, there may be a region that appears to be wet ground as opposed to a water body. This was the case in the final frame of the second sequence. In Fig. 11, the false negative water detection (green) at the bottom of the image occurred because a wet ground region was not ground truthed as part of the water body, but it was detected as water.

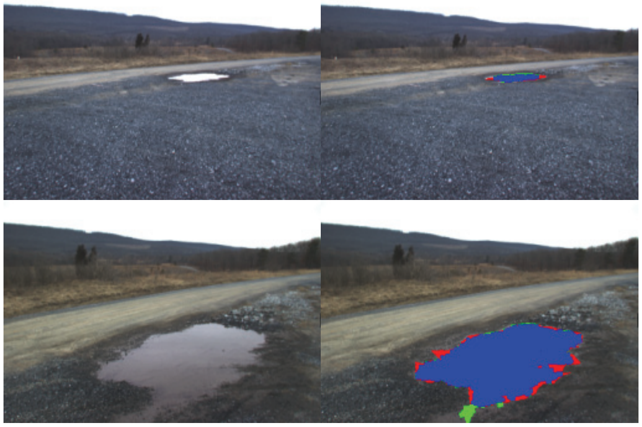


Fig. 11. Water detection results for the first and last image in a 67 frame sequence approaching a shallow water body under an overcast sky. The blue, red, and green pixels indicate true positive detection, false negative detection, and false positive detection, respectively. Water was detected in the first 50 frames by the sky reflection based water detector and the last 45 frames by the color variation based water detector.

The third sequence is a 44 frame approach to a small muddy water body containing sky and terrain reflections. Fig. 12 shows water detection results for the third sequence overlaid on the labeled water body for the first and last frame in the sequence. At far range, the terrain reflections are mostly of trees in the distance. At close range, the terrain reflections are mostly of the embankment along the trailing edge of the water body. The water body was detected in 28 frames by the sky reflection based water detector, in 23 frames by the color variation based detector, and in 43 frames by the object reflection based water detector.

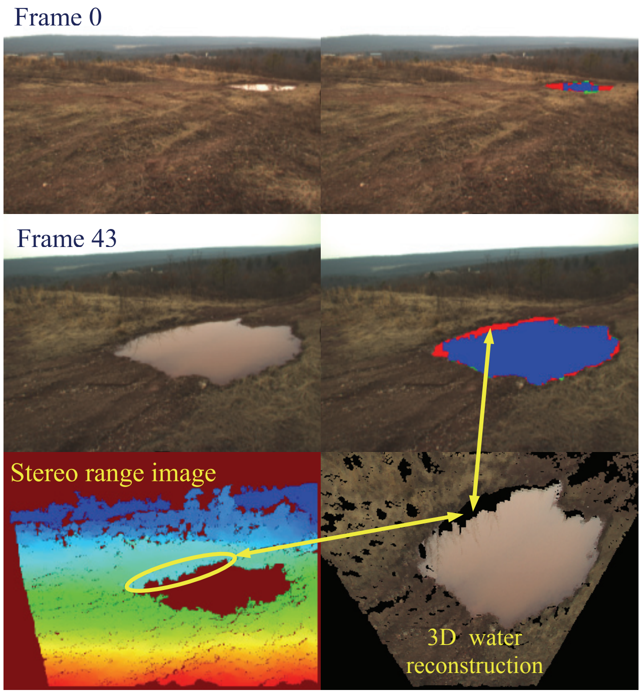


Fig. 12. Water detection results for the first frame (top row) and last frame (middle row) of a 44 frame sequence approaching a muddy water body containing sky and terrain reflections. The blue, red, and green pixels indicate true positive detection, false negative detection, and false positive detection, respectively. Water was detected in 28 frames by the sky reflection based water detector, in 23 frames by the color variation based detector, and in 43 frames by the object reflection based water detector. The bottom row shows the stereo range image and a 3D reconstruction of the water surface. Some reflections of the trailing edge embankment were undetected.

The bottom row of Fig. 12 shows a stereo range image and 3D reconstruction of the water's surface. The stereo elevation around the perimeter of the water detection region is averaged and used to construct the water's surface [12]. As we saw in Fig. 9, reflections of short terrain were largely undetected. This sequence illustrates that it can be difficult to precisely delineate where the trailing edge of a water body ends and surrounding terrain begins. The embankment and the reflections of the embankment appear equally dark.

Fig. 13 contains a graph of the true positive and false positive water detection rates for each frame in the three ground truthed water sequences as a function of the minimum range to the water's leading edge. The true positive detection rate is calculated as the number of pixels correctly classified as water divided by the number of ground truth water pixels. The false positive detection rate is calculated as the number of pixels incorrectly classified as water divided by the number of non-water pixels.

Overall, the true positive detection rate increases as the size of a water body in image space increases for each sequence. The true positive detection rate was the lowest for the third sequence. At the beginning of that sequence, the water body spanned only a small number of rows and the tree reflections split the sky reflections into two regions. Both regions were too small to definitely declare to be water. Note that the false positive detection rate was consistently low for all of the sequences. The false positive detection rate was 3.3% in one frame in the first sequence but remained below 0.8% for all the other frames. Because the false positive detection rate was fairly constant for all three sequences, we chose the plot format in Fig. 13 instead of the standard ROC format. False positive water detection only occurred around the perimeter of the labeled water body. There was substantially less false positive water detection around the perimeter of the labeled water body than false negative water detection.

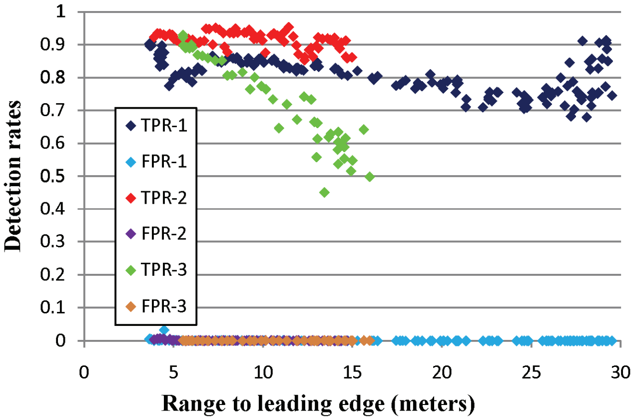


Fig. 13. The true and false positive detection rates for each frame in the three ground truthed water sequences as a function of the minimum range to the water's leading edge. Each marker corresponds to an image frame. The average XUV speed for sequence 1, 2, and 3 were 2.22, 1.86, and 3.27 m/s, respectively. The average frame rate for sequence 1, 2, and 3 were 9.33, 10.08, and 8.64 frames per second, respectively.

4. WAS IT DETECTED IN THE CORRECT LOCATION?

4.1 Ground Truthing Water Regions in Map Space

The JPL water detection software localizes detected water in a terrain map. Therefore, the ideal place to evaluate the performance of water body localization is in map space. In order to test the accuracy of water body localization, ground truth 3D coordinates of a water body's perimeter is needed in the same coordinate frame that the vehicle's position is expressed. Ground truth water body measurements can be obtained by tracing the water's perimeter with a positioning sensor, such as differential global positioning system (DGPS), or by surveying fiducials around the water's perimeter with a total station (or other surveying instrument). The location of detected water bodies can then be compared to the ground truth data to determine its accuracy.

Stereo range data is used by the JPL water detection software to localize detected water. Stereo range data around the perimeter of a detected water body is averaged to estimate the elevation of the water body. We don't use the stereo range data corresponding to the surface of detected water bodies for two reasons. First, there may be little or no stereo range data on a water body since the surface of water bodies tend to lack texture, particularly when they are stationary. Secondly, stereo range data on reflections of objects in water has a range that corresponds to the range to the reflected object, not the surface of the water body.

Once the elevation of detected water bodies is estimated, the 3D coordinates (in the stereo range image) of the pixels classified as water are modified to correspond to the surface of water bodies. This modified 3D data are used to label cells in a terrain map as water cells for comparison against the ground truth measurements. As a water body is approached, the estimate of its elevation improves. Temporal filtering is performed in the terrain map to relocate previously detected water [12].

Water localization experiments can be performed on a ground truthed test site with or without the vehicle constrained to a predefined path. During RCTA, one of the test sites used to evaluate water localization constrained the vehicle path by using a General Dynamics Robotic Systems (GDRS) instrumented train. The instrumented train contained a General Electric 24 volt DC motor, an Ogura Fail-Safe brake, an inertial measurement unit (IMU), a GPS receiver, three 3CCD color cameras that provided narrow baseline (9.5cm), mid baseline (20.5cm), and wide baseline (30cm) stereo ranging, a GDRS ladar, and the same autonomous mobility computing hardware used on XUVs. Data from the IMU and GPS were combined with a Kalman filter to provide continuous, smoothed absolute positioning data accurate to within 0.5% of the distance traveled [11].

The instrument train was constructed to enable multiple RCTA researchers to evaluate their terrain classification algorithms in a controlled environment. The use of train tracks ensures the perception sensors always follow the same route for each test run. This type of experiment enables one to directly compare the results of multiple test runs where single or multiple factors may be varied each run, such as train speed, time of day (to evaluate the effects of lighting), and the day of year (to evaluate the effects of different environmental conditions).

Perception sensor and navigation data can be logged during test runs for offline processing with terrain classification algorithms, or terrain classification can be performed in real-time and its results logged. Fig. 14 contains a picture of the instrumented train and two man-made water bodies (each approximately 1.5m x 2m) constructed adjacent to train tracks at GDRS, Westminster, Maryland. The length of the train tracks at this test site was 171 meters. Fig. 15 illustrates the train's relative positioning solution is continuous with low drift.

Two wide baseline stereo data collections were performed on the test course, the first at a command speed of 1 m/s and a frame rate of 1 frame per second (fps), and the second at a command speed of 3 m/s and a frame rate of 3 fps. The average speed during the two sequences was 0.94 and 2.62 m/s, respectively. The average speed for the first sequence during the frames where the water bodies were detected was 0.85 and 1.08 m/s, respectively. The average speed for the second sequence during the frames where the water bodies were detected was 3.50 and 3.42 m/s, respectively. For both sequences, interlaced left camera color images and right camera monochrome images were logged at a resolution of 640x480 pixels. The images processed at resolution of 320x240 pixels.

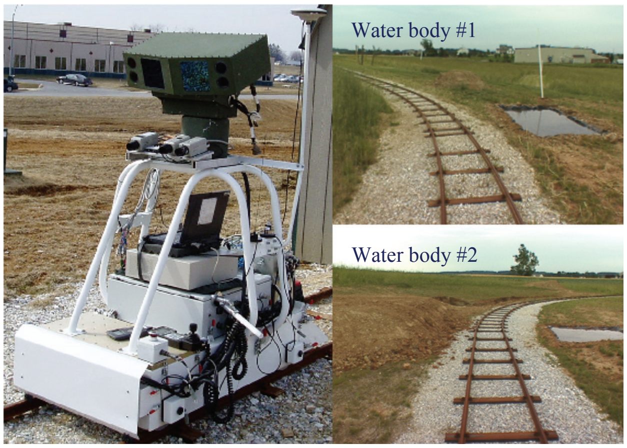


Fig. 14. A JPL passive perception system was mounted to a GDRS instrumented train. Color stereo imagery was collected on a surveyed obstacle course containing two rectangular man-made water bodies at speeds of 1 and 3 m/s.

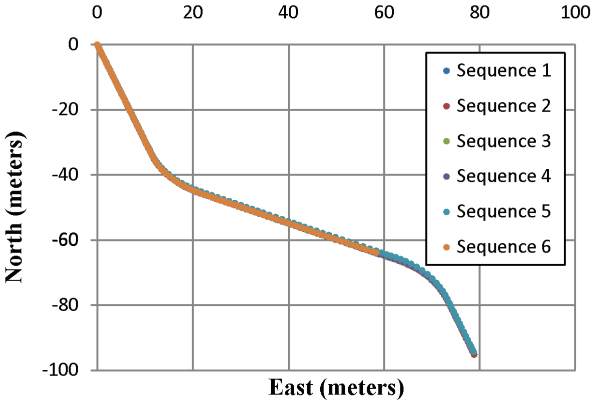


Fig. 15. The train's relative position data was logged during six sequences at 1, 1, 1, 3, 1, 3 m/s, respectively. The relative positioning solution is continuous with low drift. The last sequence ended early.

The corners of the rectangular water bodies were surveyed using a NovAtel OEM4-G2 GPS receiver operating in RT-2 differential mode (with a base station less than 30 meters from the train starting point). The accuracy of this DGPS is 1cm +1 part per million (ppm). At the farthest end of the test course, the DPGS error due to the distance from the GPS base station was less than 2mm. At each corner, DGPS data was averaged for 10 minutes, yielding an accuracy of 1cm circular error probable (CEP), i.e., half of all the data were within 1cm of the ground truth. The standard deviation in both latitude and longitude directions was less than 1cm. DGPS land survey accuracy was verified with a Leica TCR307 total station on six key points on the train rails. Fig. 16 illustrates water detection for one frame during a 1m/s run. Note in the stereo range image that there is no range data on much of the water body, except where there is a reflection of the pole directly behind the water body.

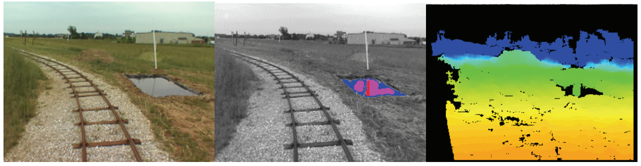


Fig. 16. A frame showing the first man-made water body on the test course described in Fig. 14 during a 1m/s run (left), water detection results overlaid on a grayscale intensity image (middle), and a wide-baseline stereo range image (right). In the color coded water detection overlay, blue, magenta, and red correspond to one, two, and three cues for water, respectively. In the stereo range image, black corresponds to no range data.

4.2 Characterizing Detection and Localization Performance in Map Space

The 3D coordinates of a water body localized with JPL's water detection software can be compared with the ground truth water body perimeter measurements to produce several measures of accuracy:

1) Difference in the detected and ground truth water body centroid (units: meters).
2) Percentage of the detected water body within the ground truth water body.
3) Percentage of the detected water body outside of the ground truth water body.
4) Percentage of the ground truth water body detected as water.
5) Maximum distance the detected water body perimeter strays from the ground truth water body perimeter (units: meters).

During the RCTA program, the ground truth measurement of a single corner of the second man-made water body was released to JPL. The rest of the water ground truth data was sequestered. At the end of the program, the sequestered ground truth data was inadvertently deleted. GDRS has been unable to recover the ground truth data from their back up disks. JPL has processed the wide-baseline data from the 1 and 3 m/s sequences off-line at an image resolution of 320x240 pixels using the all-purpose multi-cue based water detection algorithm [16]. (This analysis was performed before the specialized water detectors were implemented.) In Fig. 17, the upper graph shows the detection and localization of both water bodies for both sequences and the lower graph shows a zoom-in of the second water body.

Using a 20cm resolution terrain map, the JPL water detection software localized the corner of the second man-made water body within 20cm of the ground truth position for both sequences [11]. During the slower sequence, more water cells were detected on the leading edge. During the faster sequence, more water cells were detection along the training edge.

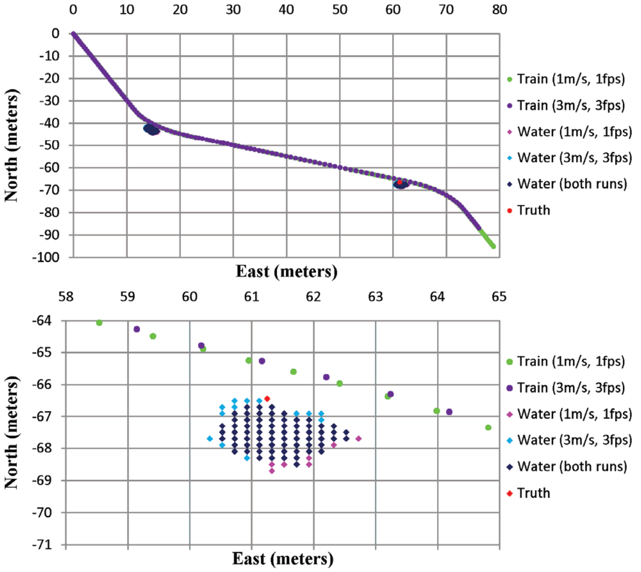


Fig. 17. Detection and localization of the two rectangular man-made water bodies shown in Fig. 16 for the two wide-baseline stereo sequences. The lower graph shows a zoom-in of the second water body. The water markers are placed at the centers of terrain map cells classified as water. Blue markers correspond to water cells detected during both sequences, cyan markers correspond to water cells only detected during the 1 m/s sequence, and magenta markers correspond to water cells only detected during the 3 m/s sequence. The red marker is the ground truth corner.

Fig. 18 contains graphs of the detection range and strength of detection for both water bodies for the 1 m/s sequence. The first water body was detected at a maximum range of 13.6 meters and the second water body was detected at a maximum range of 10.2 meters. The maximum range of detection for the second water body was slightly lower because elevated terrain near the leading edge occluded a portion of the water.

5. APPLICABILITY TO OTHER TERRAIN TYPES

This section discusses the applicability of the methods described in this paper for evaluating the classification performance of other terrain and obstacle types. Other terrain types commonly encountered on cross-country terrain includes vegetation of varying height and density, trees (both upright and fallen), rocks of varying size, telephone poles, signs, ditches, and a variety of ground surfaces such as gravel, asphalt, dry soil, mud, snow, and ice. All of these terrain types are stationary. In addition, there are obstacles types that may be stationary (e.g. ditches) or moving (e.g. vehicles and people). Some terrain types can be driving hazards for UGVs (e.g. a large mud puddle) and some are harmless (e.g. an asphalt road surface). Obstacle types should always be detected and avoided.

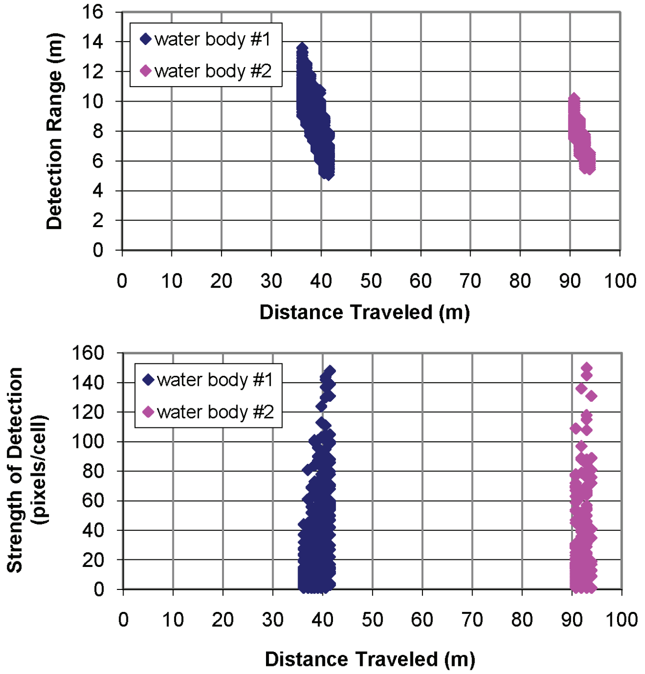


Fig. 18. Detection range and strength of detection as a function of distance traveled for the two rectangular man-made water bodies shown in Fig. 14 during the 1 m/s sequence.

5.1 Ground Truthing Other Terrain and Obstacle Types in Image Space

Terrain classification algorithms are useful for distinguishing between safe and unsafe terrain types. Pedestrian and vehicle detection algorithms are useful in detecting those classes of obstacles [10]. In order to fully characterize the performance of a terrain classifier, terrain types need to be ground truthed and the output of the terrain classifier should be compared to the ground truth regions.

The ground truthing tool described in this paper can be used to segment any type of terrain that has distinct edges. Thus far, we have used it to manually select vertices on the perimeter of water bodies in images and propagate those 2D points into future image frames using 3D coordinates from dense stereo. This technique should work equally well on other ground surfaces with distinct edges that are completely or mostly visible within an image. Some examples of ground surfaces with completely or mostly visible distinct edges are road surfaces, mud puddles bounded by dry soil, and soil bounded by vegetation.

This technique, however, would not work equally well on image sequences of ubiquitous ground surfaces (e.g. snow, desert sand). In images, ubiquitous ground surfaces typically only have a distinct edge where it meets the horizon. An option that could be added to the current tool to ground truth ubiquitous ground surfaces would be to detect the horizon and label all pixels below it. One complication is when a region to be ground truthed contains interior regions having a different terrain type. An example of this would be trees extending out of a snow covered ground. Currently, the ground truthing tool does not provide an option to exclude regions interior to a ground truthed perimeter. But it is fairly straight forward to add one.

Since the surface of a water body lies in a horizontal plane, we require that at least one selected vertex has stereo data (to determine the equation of the plane). If more than one vertex has stereo data, the *z* coordinates of the vertices are averaged. 3D coordinates are generated for vertices that do not have stereo data by finding the intersection of the ray extending from a vertex pixel and the horizontal plane. But other ground surface terrain types are not constrained to a horizontal plane. Therefore, for other ground surface terrain types, stereo data is required for each vertex. Selecting vertices around the perimeter of non-water ground surface terrain types where there are stereo data would likely only require a marginal amount of additional effort.

Fig. 19 contains examples of ground-truthed gravel road and grass regions. Grass regions are particularly difficult to ground truth. The boundary between soil and grass is usually ragged, requiring a substantial number of vertices to accurately model. Road perimeter vertices will likely propagate to future frames much better than grass perimeter vertices. This is because stereo data on grass tends to be more variable due to the prevalence of mixed foreground and background pixels. Shadows can also complicate ground truthing. Edges between terrain types tend to be less distinct in shadows, requiring more vertices.

The ground truthing tool will likely work fine on solid terrain or obstacle types that extend up out of the ground surface (e.g. rocks, tree trunks, poles) when the shapes are simple and the objects are approached head on. But if the UGV makes heading changes during the approach, or the object is in the periphery of the imagery, new portions of the object will become visible as the viewpoint relative to the object is changed. In subsequent frames, vertices may no longer lie on the edge of an object. When this occurs, the vertices will need to be manually repositioned to the new object edge.

The ground truthing tool can also be used to ground truth negative obstacles (e.g. ditches, holes) whose perimeter is not occluded by vegetation or other terrain types. Negative obstacles, however, do not always have distinct edges in color space. Therefore, for some negative obstacles, the intelligent scissors algorithm may produce unreliable results. For cases where the object being ground truthed does not have distinct edges, the tool allows you to model the object perimeter with line segments.

When the object shape is complex (e.g. a vehicle), new portions of the object will rapidly come into view even if the approach to the object is head on, making it difficult to use this tool in its current form. This tool was designed to work with stationary planar terrain. Another extension to the tool would be to perform feature tracking on each vertex to track edge points of moving objects. But this extension would likely fail at close range where new portions of an object rapidly come into view from frame to frame. For moving objects and objects with complex shapes, a bounding box is typically used to ground truth the object in image space.

Tall vegetation and tree foliage present a further challenge to the ground truthing tool. Because of its porous and non-homogeneous characteristics, ground truthing results would be unpredictable. As with pedestrians and moving obstacles, a bounding box is currently more appropriate for ground truthing tall vegetation and tree foliage.

The ground truthing tool can be further automated by enabling the tool to suggest initial seed points for the intelligent scissors algorithm. For example, a procedure like Canny edge detection can generate some very strong edge locations as candidate seed points.

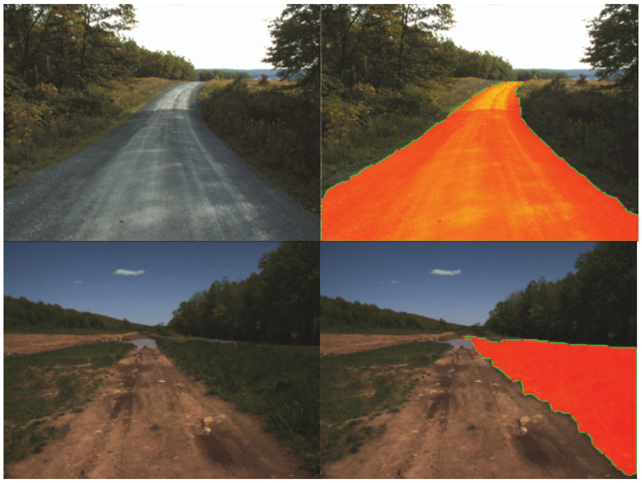


Fig. 19. Examples of ground truthing ground surface terrain types in single frames. In the top row, a gravel road is ground truthed. In the bottom row, a section of short grass is ground truthed.

5.2 Ground Truthing Other Terrain and Obstacle Types in Map Space

To ground truth water bodies in map space we used DGPS (accurate to 1cm CEP), placing the remote GPS antenna at the corners of man-made rectangular shaped water bodies. Naturally occurring water bodies, however, have irregular shaped perimeters. DGPS may be used to ground truth the perimeter of naturally occurring water bodies and other ground surface terrain types by placing the remote GPS antenna at the end of a pole and tracing the perimeter with the antenna close to the ground surface. This technique may also be used on negative obstacles. However, there may be some ambiguity on the exact location of the perimeter for irregular shaped negative obstacles with eroded edges.

Pedestrians have been successfully ground truthed using an ultra wideband (UWB) wireless tracking system [1]. This technique may be extendable to moving vehicles by placing a transmitter on poles attached to each of the four corners. For stationary positive obstacles with simple shapes, one could survey the corners with a commercial-off-the-shelf (COTS) surveying instrument such as a total station. The trunks of trees can be modeled as cylinders. Centerline points could be surveyed with a total station and the circumference measured with a measuring tape. Irregular shaped positive obstacles, such as large rocks, would be problematic to fully ground truth in 3D. One could partially ground truth them by tracing out the perimeter of the base using DGPS and measuring the maximum height.

6. CONCLUSIONS

In this paper, we described a methodology for conducting water detection performance evaluations in image space, and water localization performance evaluations in map space. JPL has developed a software tool for ground truthing water bodies in stereo image sequences. The ground truthing tool enables users to step though a sequence of images and select a limited number of vertices (in left rectified images) around the perimeter of water bodies. We have extracted the *intelligent scissors* portion of GIMP and have integrated it into our ground truthing tool. *Intelligent scissors* is run for each pair of neighbor vertices to segment the water's edge. Stereo vision is used to update the 2D image coordinates of the vertices as the user steps from one frame to the next. At each frame, the user can add, move, or delete vertices, or move or delete entire polygons.

Typically, a water body is ground truthed in image space starting with the final image (where the water body is the largest) and ending with the first image (where the water body is the smallest). Thus far, we have used the JPL water body ground truthing software to label three color stereo sequences of images. In the longest sequence (143 frames), the water body was detected in every frame. The true positive detection rate ranged from 68% (at a range of 28 meters to the leading edge) to 90% (at a range of 4 meters to the leading edge). The false positive detection rate was 3.3% in one frame but remained below 0.8% in the rest of the frames.

This evaluation indicated that weak reflections of vegetation lining the edge of a water body are consistently undetected by the three specialized water detectors. In addition, the true positive water detection rate is currently underestimated when there are objects extending out of the interior of a water body, such as sediment, rocks, and vegetation. The reason for this is the water body ground truthing tool does not currently exclude these regions. Additional work is needed to enable the ground truthing tool to exclude these regions, and to filter noisy stereo range measurement during water body elevation estimation.

We also outlined several measures of accuracy for comparing the 3D coordinates of a water body localized with JPL's water detection software with 3D ground truth water body perimeter measurements. We have not experimented with any of these measures yet due to a lack of 3D ground truth water body perimeter measurements to date. More work is needed to determine the usefulness of these measures.

ACKNOWLEDGEMENT

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the U.S. Army under the Robotics Collaborative Technology Alliances and Future Combat System Autonomous Navigation System 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] B. Bodt, R. Camdens, H. Scott, A. Jacoff, T. Hong, T. Chang, R. Norcross, T. Downs, and A. Virts, "Performance measurements for evaluating static and dynamic multiple human detection and tracking systems in unstructured environments", *Performance Metrics for Intelligent Systems Workshop*, Gaithersburg, MD, 2009.

[2] B. Bodt, M. Childers, and R. Camden, "A capstone experiment to assess unmanned ground vehicle tactical behaviors developed under the robotics collaborative technology alliance", *Proceedings of AUVSI*, Denver, CO, Aug. 2010.

[3] M. Childers, B. Bodt, and R. Camden, "Assessing unmanned ground vehicle tactical behaviors performance", in this issue of *International Journal of Intelligent Control and Systems: Quantifying the Performance of Intelligent Systems*, 2011.

[4] E. W. Dijkstra, "A note on two problems in connection with graphs", *Numerische Mathematik*, Vol. 1, 1959, pp. 269-271.

[5] M. Fischler, C. Connolly, and R. Bolles, Unpublished work performed during the U.S. Army Research Laboratory funded Robotics Collaborative Technology Alliances program by SRI International, Menlo Park, CA, 2004.

[6] D. B. Gennery, "Least-squares camera calibration including lens distortion and automatic editing of calibration points", in *Calibration and Orientation of Cameras in Computer Vision*, A. Grun and T. Huang (Ed.), Springer-Verlag, 2001, pp. 123–136.

[7] T. Hong, C. Rasmussen, T. Chang, M. Shneier, "Fusing ladar and color image information for mobile robot feature detection and tracking", *7th International Conference on Intelligent Autonomous Systems*, Marina Del Ray, CA, Mar. 2002.

[8] L. Matthies, P. Bellutta, and M. McHenry, "Detecting water hazards for autonomous off-road navigation", *Proceedings of SPIE*, Vol. 5083, Orlando, FL, Apr. 2003, pp. 231-242.

[9] E. N. Mortensen, and W. A. Barrett, "Intelligent scissors for image composition", *International Conference on Computer Graphics and Interactive Techniques*, Los Angeles, CA, Aug. 1995, pp. 191-198.

[10] A. Rankin, M. Bajracharya, A. Huertas, A. Howard, B. Moghaddam, S. Brennan, A. Ansar, B. Tang, M. Turmon, and L. Matthies, "Stereo-vision based perception capabilities developed during the Robotics Collaborative Technology Alliances program", *Proceedings of SPIE*, Vol. 7692, Orlando, FL, Apr. 2010.

[11] A. Rankin, A. Huertas, and L. Matthies, "Evaluation of stereo vision obstacle detection algorithms for off-road autonomous navigation", *Proceedings of AUVSI*, Baltimore, MD, Jun. 2005.

[12] A. Rankin and L. Matthies, "Daytime water detection and localization for unmanned vehicle autonomous navigation", *Proceedings of the 25th Army Science Conference*, Orlando, FL, Nov. 2006.

[13] A. Rankin and L. Matthies, "Daytime water detection based on color variation", *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, Taiwan, Oct. 2010.

[14] A. Rankin and L. Matthies, "Water detection based on object reflections", Submitted to: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Francisco, Sep. 2011.

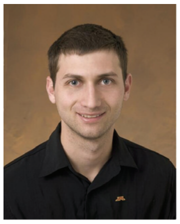
[15] A. Rankin, L. Matthies, and P. Bellutta, "Daytime water detection based on sky reflections", *Proceedings of the IEEE International Conference on Robotics and Automation*, Shanghai, China, May 2011.

[16] A. Rankin, L. Matthies, and A. Huertas, "Daytime water detection by fusing multiple cues for autonomous off-road navigation", *Proceedings of the 24th Army Science Conference*, Orlando, FL, Nov. 2004.

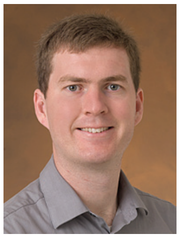
[17] A. Sarwal, J. Nett, and D. Simon, "Detection of small water bodies", *PercepTek Robotics Technical Report*, Littleton, CO, (http://www.dtic.mil, AD# ADA433004, 2004.

[18] Y. Yakimovsky and R. Cunningham, "A system for extracting three-dimensional measurements from a stereo pair of TV cameras", *Computer Graphics and Image Processing*, 7, 1978, pp. 195-210.

Arturo Rankin is a senior member of Surface Systems Perception Group at the Jet Propulsion Laboratory, California Institute of Technology, USA. He received a B.S. degree in Mechanical Engineering in 1987 from the Catholic University of America, Washington D.C., and M.S. and Ph.D. degrees in Mechanical Engineering in 1993 and 1997 from the University of Florida, USA. He has been a researcher at JPL since 1997. At JPL, he has developed obstacle detection, terrain perception, and terrain mapping software for a variety of UGV programs.



Tonislav Ivanov is an associate member of the Aerial System Perception Group at the Jet Propulsion Laboratory, California Institute of Technology, USA. He received a M.S. degree in Electrical and Computer Engineering from Cornell University in 2007. He has done research on obstacle avoidance for autonomous robots, facial recognition, head pose recognition, and localization of objects. He currently works on LIDAR data processing and the development of hazard detection algorithms for the Autonomous Landing and Hazard Avoidance (ALHAT) project. He is also working on hazard mapping from Lunar Reconnaissance Orbiter (LRO) images and developing lunar digital elevation maps. His expertise includes computer vision, artificial intelligence, machine learning, biomedical imaging, and digital signal processing.



Shane Brennan is an associate member of the Aerial System Perception Group at the Jet Propulsion Laboratory, California Institute of Technology, USA. He received B.S. (2006) and M.S. (2008) degrees in Computer Engineering from the University of California at Santa Cruz. He has been a researcher at JPL since 2008. Prior to working at JPL he held internships at HP Labs in Palo Alto, California and Vidient Inc. in Sunnyvale, California. His work focuses on aspects of image processing, computer vision, and geometrical reasoning.