Fusing Ladar and Color Image Information for Mobile Robot Feature Detection and Tracking

Tsai-Hong Hong, Christopher Rasmussen*, Tommy Chang, and Michael Shneier

{hongt, crasmuss, tchang, michael.shneier}@nist.gov
National Institute of Standards and Technology
Gaithersburg, MD 20899

Abstract

In an outdoor, off-road mobile robotics environment, it is important to identify objects that can affect the vehicle's ability to traverse its planned path, and to determine their three-dimensional characteristics. In this paper, a combination of three elements is used to accomplish this task. An imaging ladar collects range images of the scene. A color camera, whose position relative to the ladar is known, is used to gather color images. Information extracted from these sensors is used to build a world model, a representation of the current state of the world. The world model is used actively in the sensing to predict what should be visible in each of the sensors during the next imaging cycle. The paper explains how the combined use of these three types of information leads to a robust understanding of the local environment surrounding the robotic vehicle for two important tasks: puddle/pond avoidance and road sign detection. Applications of this approach to road detection are also discussed.

1 Introduction

An autonomous vehicle driving across unknown terrain must be able to detect potential obstacles and identify them well enough to determine if they can be traversed. This must be accomplished fast enough to ensure that the vehicle has enough time and space to avoid obstacles. The work described in this paper is part of the Army's Demo III project [1]. The requirements for the Experimental Unmanned Vehicle (XUV) developed for Demo III include the ability to drive autonomously at speeds of up to 60 kilometers per hour (km/h) on-road, 35 km/h off-road in daylight, and 15 km/h off-road at night or under bad weather conditions. The control system for the vehicle is designed in accordance with the 4D-Real-time Control System (RCS) architecture [2], which divides the system into perception, world modeling, and behavior generation subsystems.

Figure 1: The Demo III XUV

The XUV has two principal sets of sensors for navigation, as shown in Figure 1. On the left, outlined in white, is a ladar system that produces range images at about 20 Hz. Mounted above the ladar is a color camera (not pictured) that produces images at up to 30 Hz. On the right are a pair of stereo color cameras, and a set of stereo FLIR cameras. The work described in this paper concerns the use of the ladar sensor and its associated color camera. The way each is used in conjunction with the other and with information stored in the vehicle's internal world model is the focus of the paper.

Given the need for relatively high speed driving, the sensory processing subsystem must be able to update the world model with current information as quickly as possible. It is not practical to process all images completely in the time available, so focusing attention on important regions is required. This is done by trying to predict which regions of future images will contain the most useful information based on the current images and the current world model. Predic-

*This work was performed while the author held a National Research Council Research Associateship Award at NIST.

tion is carried out between images, across images, and between the world model and each type of image.

Prediction and focus of attention are of special interest to robotic systems because they frequently have the capability to actively control their sensors [3]. The goal of focusing attention is to reduce the amount of processing necessary to understand an image in the context of a task. Usually, large regions either contain information of no interest for the task, or contain information that is unchanged from a previous view. If regions of interest can be isolated, special and perhaps expensive processing can be applied to them without exceeding the available computing resources.

Most focus of attention systems work by looking for features usually defined by some explicit or implicit model. The search may take many forms, from multiresolution approaches that emulate human vision's peripheral and foveal vision, to target-recognition methods that use explicit templates for matching [4, 5, 6, 7]. Once a set of attention regions has been detected, a second stage of processing is often used to further process them, or to rank them. This processing may require more complex algorithms, but they are applied only to small regions of the image.

In this paper, we describe an approach to feature detection and tracking that falls within the above general description, but differs from previous approaches in using multiple sensor types that interact to locate and identify features. A world model containing the system's current best guess about the state of the world is used to predict where features should appear and how they should look to each of the sensors. The world model serves as a common coordinate system for comparing and integrating ladar and camera observations, as well as a framework for doing simple tracking of detected features. Judicious choice of which sensor to run initial perception routines on often significantly reduces computational burden, while combining the ladar and camera information tends to boost performance over either used singly.

2 Methods

2.1 The World Model

The World Model (WM) contains a 3-D, annotated representation of the current state of the terrain surrounding the vehicle and is updated continually by the sensors. We use a modified occupancy grid representation [8], with the vehicle centered on the grid, and the grid tied to the world. The WM thus scrolls under the vehicle as the vehicle moves about in the world. The world model is the system's internal representation of the external world. It acts as a bridge between sensory processing and behavior generation by providing a central repository for storing sensory data in a unified representation, and decouples the real-time sensory updates from the rest of the system. The WM process has two primary functions:

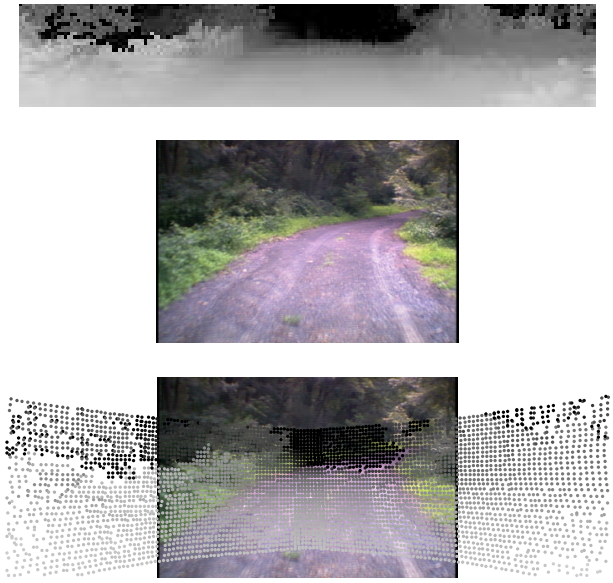


Figure 2: Camera-ladar registration. Darker laser pixels are more distant.

Create a knowledge database (map) and keep it current and consistent by updating existing data in accordance with inputs from the sensors and deleting information no longer believed to be representative of the world. The WM also assigns confidence factors to all map data and adjusts them as new data are sensed. Types of information in the map include state variables (e.g., time, position, orientation), system parameters (e.g., coordinate transforms, sensor to vehicle offsets, etc.), and lists or classes of sensed objects. The world model process also provides functions to update and fuse data and to manage the map (e.g. scrolling and grouping objects.)

Generate predictions of expected sensory input based on the current state of the world and estimated future states of the world. For the Demo III off-road autonomous driving application, very little a priori information is available to support path planning between the vehicle's position and a final goal position. The world model therefore constructs and maintains all the information necessary for intelligent path planning [9].

Prediction is used to focus attention on regions that have previously been identified as interesting. It facilitates tracking, enables confidences in features to be

updated, and allows information found in one sensor to influence processing in another. Prediction is mediated in our system by the world model. Since we use a grid representation fixed to the world, it is straightforward to project regions in the world model into each of the sensor coordinate systems. Currently, we only predict where a feature is expected to occur, not what it may look like.

2.2 Coordinate System Transformations

Since the sensors are mounted on a mobile platform, and the sensors themselves move, projections are not fixed, but must be computed each time they are needed. There are two kinds of projections: The ladar data are projected into the world model, and features identified in the ladar data are projected into the color image space.

Each sensor is at a known base position on the vehicle, and has a known sensor coordinate system. The vehicle is moving, however, and the WM maintains its representation in world coordinates, fixed on the ground. Thus, all coordinates must be converted from sensor to vehicle, and from vehicle to world. Some of the sensors also move relative to their base position. The ladar, for instance, may rotate about its horizontal axis (tilt). Finally, the sensors sample at different times, so a correction must be made for their relative positions in space when mapping between images.

The ladar-to-WM coordinate transformation includes the ladar-to-vehicle and vehicle-to-world-model transformations. The projection from the WM to the color camera image includes WM-to-ladar and ladar-to-image transformations. The ladar-to-image transformation is particularly important in order to achieve an accurate registration between ladar features and image features. This transformation is not invertible because of the lack of depth information in the camera image. In order to register the ladar and camera images, we first calibrated the camera's internal parameters using J. Bouguet's Matlab toolbox [10]. The external orientation between the camera and ladar was obtained by correlating corresponding points imaged by each device over a number of scenes and then computing a least-squares fit to the transformation according to the procedure described in [11]. Results are shown for a sample scene in Figure 2.

3 Feature Types

In this section we will discuss two examples of feature types that our system detects and tracks: puddles and road signs—specifically, signs marking an endangered butterfly sanctuary. Puddles, ponds, and mud are a serious problem for off-road mobile vehicles because of the danger they pose of the vehicle getting stuck in them, as well the possibility of water damage to the engine and/or critical electrical components. For our purposes, butterfly signs indicate the borders of an ecologically protected zone where we do testing that the vehicle must not enter, but human-readable signs might also mark minefields or contain other important information [12], making the ability to find them a critical one.

Figure 3: Butterfly signs

A third task that is ongoing work, road finding, is also briefly discussed. There has been some work on following marginal rural roads using color cues [13], but road *detection*, which is a vital skill for backcountry navigation, has been less studied.

3.1 Butterfly signs

Butterfly signs are rectangular yellow placards mounted on six-foot wooden posts (painted orange on top) that delimit a "no driving zone," thus affecting the path-planning module of the XUV system. Two such signs are shown in Figure 3.

In the ladar domain, signs can be distinguished from the background because they frequently jut above surrounding foliage, with good depth contrast and nothing above them. When fixed, the limited vertical field of view of the ladar (15 degrees) tends to cut off the tops of the signs, so we use a sign-finding operator that simply searches for a vertical bar (i.e., not end-stopped) in the ladar range image at a scale corresponding to 5-10 meters distance to the sign.

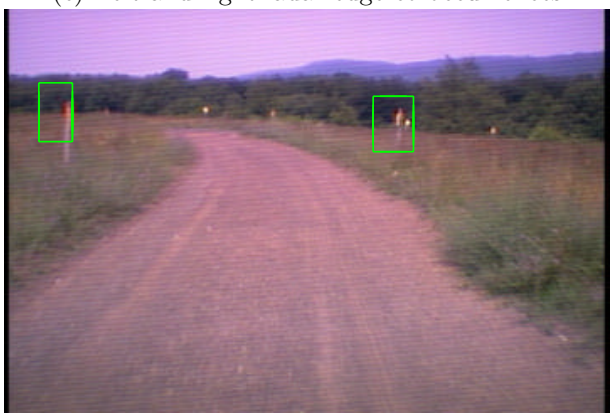
The steps of the method are illustrated for a sample range image in Figure 4: first, two odd-phase Gabor filters [14] are run over the range image in Figure 4(a) to find left and right vertical edges, respectively, and the output of the filter in Figure 4(b) is thresholded

to isolate strong edges. Second, we search for range image locations where left and right edges *co-occur*—that is, all (x, y) such that there is a left edge at $(x - \delta, y)$ and a right edge at $(x+\delta, y)$ for $\delta \leq 2$. This yields ladar-based hypotheses for sign locations, as shown in Figure 4(c).

(a) Ladar range image

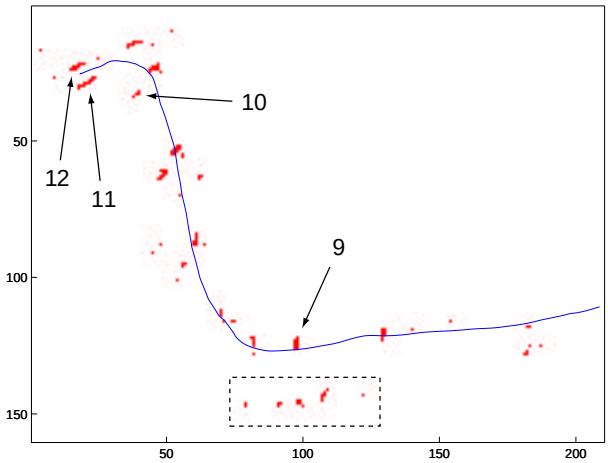
(b) Gabor-filtered range image (right edges)

(c) Left and right ladar edge co-occurrences

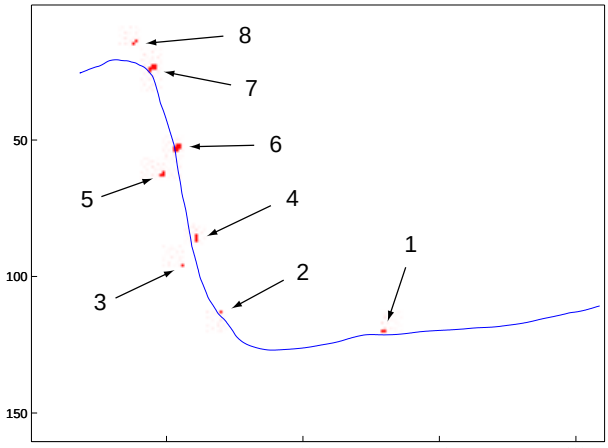


(d) Color confirmation matches

Figure 4: Butterfly sign detection



(a) Ladar only



(b) Ladar and color

Figure 5: Butterfly sign maps. Numbered arrows point to correctly-detected sign locations; the dotted box in (a) is a group of trees. Units are in meters.

We can accumulate these hypotheses in the world model as the XUV drives; a representation of this part of the WM after the XUV has followed a road with $\sim$15 signs along it for several hundred meters is given in Figure 5(a). The WM is shown at a 1-meter grid square resolution, with only squares that have had five or more sign hypotheses projected to them shown in red; the vehicle path calculated from its inertial navigation system is shown as a blue line. Nearly all of the signs are found, but at the cost of a number of false positives from vegetation. The spots inside the dotted rectangle, for example, are likely tree trunks.

To increase the accuracy of our sign-finder we add color, which has been shown to be a useful cue for sign detection [12]. A very simple method for modeling the sign color follows from sampling sign pixels over multiple images and performing principal components analysis on the pixel distribution to parametrize an ellipsoid in RGB space. Color similarity of an arbitrary pixel is then the Mahalanobis distance of the pixel color to the model ellipsoid's center. The ladar-based method above is easily extended by projecting ladar sign hypotheses (such as in Figure 4(c)) into the camera image, and then measuring the "yellowness/orangeness" of the local image neighborhood to check them. Specifically, we project each ladar sign hypothesis from the co-occurrence step to image coordinates (x, y) and compute the minimum Mahalanobis distance $d_{x,y}$ to the butterfly sign color over a $20x20$ region about (x, y); if $d_{x,y}$ is less than a threshold then

the hypothesis is *confirmed*. Bounding boxes on two clusters of confirmed ladar sign hypotheses from Figure 4(c) are shown in Figure 4(d). Color-confirmed ladar hypotheses for the road sequence are projected to the world model in Figure 5(b). This completely eliminates false positives, although a few correct detections are also deleted. A better color model would likely prevent this. Some signs were missed by both methods either because they were too far away to resolve or because foliage growing behind them eliminated depth contrast.

Focus of attention serves here to minimize computation: by searching first in the ladar domain, expensive image processing is limited to small neighborhoods around good candidates. By integrating sign detections over multiple frames, the world model throws out spurious sensor responses and betters the precision of the location estimates of the signs.

3.2 Puddles

In our standard ladar-based navigation system, we have found that puddles and other standing water appear as smooth, level surfaces. These qualities make such areas highly attractive to the motion planning system and therefore dangerous. We would like to detect puddles and flag them as "no go" or at least worthy of extra caution. Fortunately, a simple test follows from the optical properties of the ladar: laser beams hitting a puddle at an oblique angle are reflected away from the sensor, and result in no data being returned. Such points show up as *voids* in the ladar images, but are not the only source of missing data. Out-of-range depths are also recorded in any sky or otherwise distant regions in the ladar's field of view.

Our puddle detection algorithm thus looks for voids in the data, and then scans a region surrounding them. Puddle-derived voids are distinguished from sky by requiring that there be non-void pixels above every column in a connected component. Assuming that there are ground points somewhere adjacent to the puddle in the ladar image (rather than all pathological cases like overhanging limbs), we can obtain a reasonable estimate of the height of the water surface from the minimum height in the WM of over all points surrounding the puddle in the image. This allows us to solve for the missing range values in the puddle interior and thus properly place it in the map. These steps are illustrated in Figure 6. Without explicitly detecting puddles in this manner, height maps used for navigation have areas of missing data in them (see Figure 7(a)) that are similar to laser "shadows" behind obstacles and protruding objects. Puddle detection permits water hazards to be placed in the map for higher-level reasoning, as shown in Figure 7(b).

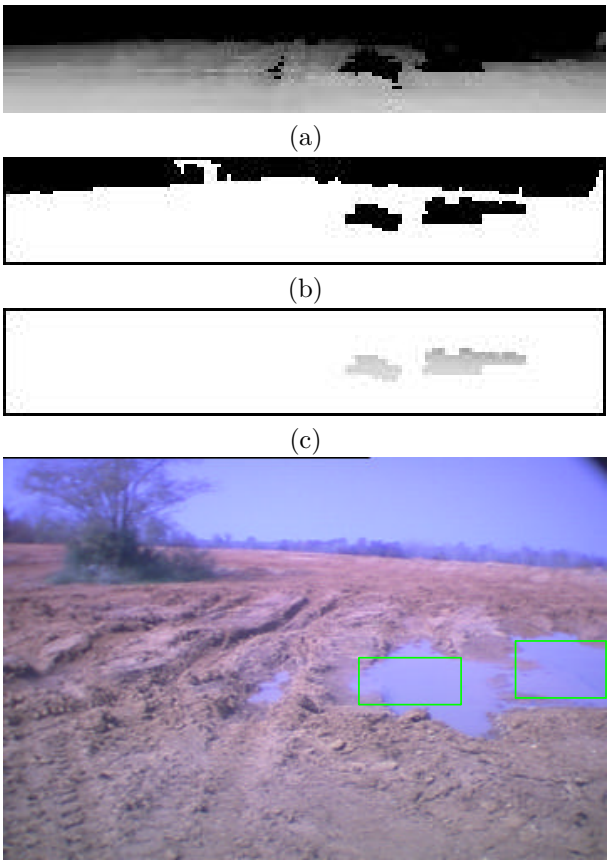


Figure 6: Puddle detection. (a) Raw ladar image containing puddles; (b) Smoothed sky and puddles after segmenting voids, morphological closing; (c) Puddles after sky removal, range calculation; (d) Simultaneous color image with bounding boxes of projected puddles.

In the case of puddles, prediction plays an important role in ladar processing. As the vehicle approaches a puddle, the angle at which a laser ray hits the water gets steeper, and at a *critical angle* the sensor starts to record a return from the puddle. Without knowing that the region had already been identified as a puddle, the sensor would start to indicate that the region was traversable and smooth, which would make it a preferred location for the planner. Marking a region as already identified in the world model prevents this behavior. Note that puddles are unusual in that the confidence in most features increases through multiple views whereas using the ladar sensor to view puddles over time reduces their confidence. The behavior of the ladar sensor in the neighborhood of other features that produce voids, such as holes and occlusions, is

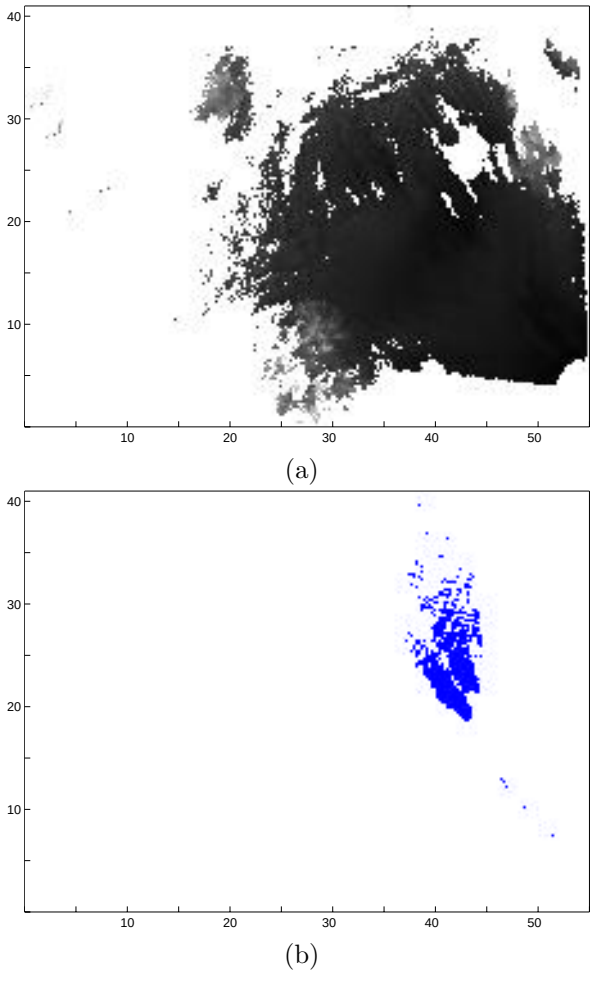


Figure 7: Mapping puddles. (a) Height map made from unprocessed ladar images; (b) Map of projected puddle regions with five or more hits per 0.25 meter grid square.

very different from that around puddles. This enables a distinction to be made over time that might not be made from a single view.

Before the critical angle is reached, using the ladar alone is generally sufficient to correctly identify a puddle. After the critical angle the world model location of the puddle serves to guide the XUV away from water, but as it is unsupported by sensory data the reliability of the map decays over time. To maintain the reliability of the WM puddle locations even after the critical angle, we look in the color image for supporting evidence to reduce false positives. When a puddle is detected in the ladar data (with sufficient confidence in the map), a window is placed about the potential puddle region, and projected into the color image as in Figure 6. In the color domain, the system tries to determine if the region has a similar color to what is above it. Often, this will be the sky, so a blue color will mean a puddle. At other times, however, the puddle may reflect trees, grass, or clouds. The algorithm searches for a match, but may fail if the puddle is reflecting something not in the image. When the puddle is verified, color information from the puddle points can then be placed into the world model. By continually updating this color information while the ladar still sees a void, the system can smoothly transition to relying on color alone to segment a puddle even after the critical angle is reached. The confidence of puddle in the map is increased by a predefined value that depends on the robustness of the color classification algorithms.

3.3 Roads

Color from captured camera images can be combined with the 3-D information returned by the ladar range-finder in the world model. We do this by simply projecting ladar points into the current image and reading off the (R, G, B) values of the image pixels they land on, and then carrying that color information along when the ladar data is projected into the world map, averaging color per grid square. Example maps with fused height and color information created from two driving sequences are shown in Figures 8(a) and (b). The grid square size is 0.25 meters and every ladar image from each sequence is mapped (the sequence shown in (a) has 1072 ladar frames and (b) has 572). Height maps are displayed with the minimum height in the map as black and the maximum as white (unmapped squares are also white).

Observing that roads' structural and visual characteristics often differ from those of bordering areas in *height* (bushes, trees, and rocks tend to "stick up"), *smoothness* (roads are locally flat, while grass, etc. are bumpier), and *color* (asphalt, dirt, and gravel roads' hues are separable from those of vegetation and sky [13]), it is possible to formulate an objective function to distinguish roads in the ladar-color domain as contiguous world model regions with small variance in height and acceptable color distributions (brown, black, etc.). In Figure 8(a), the ladar data alone is sufficient to discriminate the road via height and smoothness features. When, however, the road and non-road are differentiated mainly by color as in Figure 8(b), color image information becomes critical. The ladar data remains useful, however, as a means of focusing attention by permitting obstacle regions such as trees and foliage to be masked out, reducing the map area searched for possible roads. We are currently investigating the use of GPS information from the

XUV's navigational system and *a priori* map information to selectively cue evaluation of the objective function based on proximity.

4 Conclusion

By focusing on a few critical subtasks of the general off-road autonomous navigation problem, we have demonstrated the utility of combining information from a ladar and color camera for feature detection and tracking. The two sensors have strengths that are often complementary, and careful staging of algorithmic modules results in increased task performance without imposing the computational burden that simply analyzing or filtering both modalities and fusing them afterward would. Further, fusing data via a world model has proven a flexible way to integrate synchronized information from the two sensors while improving its quality over time, and being able to project the combined information into the sensor domain enables cheap prediction of what the sensors should see in subsequent views. By using both individual sensor characteristics and prediction, it is possible to focus attention on important features and to bring more sensor resources to bear on identifying them.

References

[1] C. Shoemaker and J. Bornstein, "The Demo3 UGV program: A testbed for autonomous navigation research," in *Proc. IEEE International Symposium on Intelligent Control*, 1998.

[2] J. Albus, "4-D/RCS: A reference model architecture for demo III," Tech. Rep. NISTIR 5994. 3-1-1997, National Institute of Standards and Technology, 1997.

[3] J. Clark and N. Ferrier, "Attentive visual servoing," in *Active Vision*, A. Blake and A. Yuille, Eds., pp. 137–154. MIT Press, 1992.

[4] K. Toyama and G. Hager, "Incremental focus of attention for robust vision-based tracking," *Int. J. Computer Vision*, vol. 35, no. 1, pp. 45–63, 1999.

[5] J. Eklundh, P. Nordlund, and T. Uhlin, "Issues in active vision: attention and cue integration/selection," in *Proc. British Machine Vision Conference*, 1996, pp. 1–12.

[6] C. Westin, C. Westelius, H. Knutsson, and G. Granlund, "Attention control for robot vision," in *Proc. Computer Vision and Pattern Recognition*, 1996, pp. 726–733.

[7] F. Ennesser and G. Medioni, "Finding Waldo, or focus of attention using local color information,," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 805–809, 1995.

[8] D. Oskard, T. Hong, and C. Shaffer, "Real-time algorithms and data structures for under water mapping," in *SPIE Cambridge Symposium on Optical and Optoelectronic Engineering*, 1988.

[9] A. Lacaze, J. Albus, and A. Meystel, "Planning in the hierarchy of NIST-RCS for manufacturing," in *Proc. Int. Conf. on Intelligent Systems: A Semiotic Perspective*, 1996.

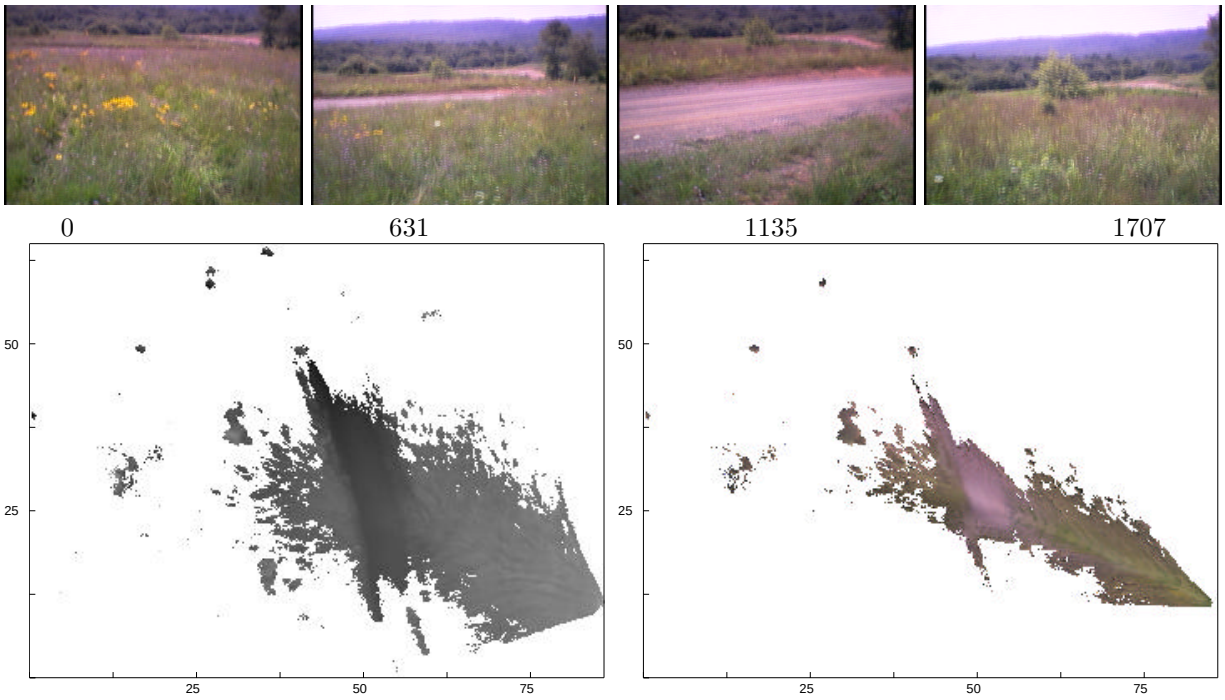
[10] J. Bouguet, "Camera Calibration Toolbox for Matlab," Available at `www.vision.caltech.edu/bouguetj/calib_doc`. Accessed May 11, 2001.

[11] M. Elstrom, P. Smith, and M. Abidi, "Stereo-based registration of LADAR and color imagery," in *SPIE Conf. on Intelligent Robots and Computer Vision*, 1998, pp. 343–354.

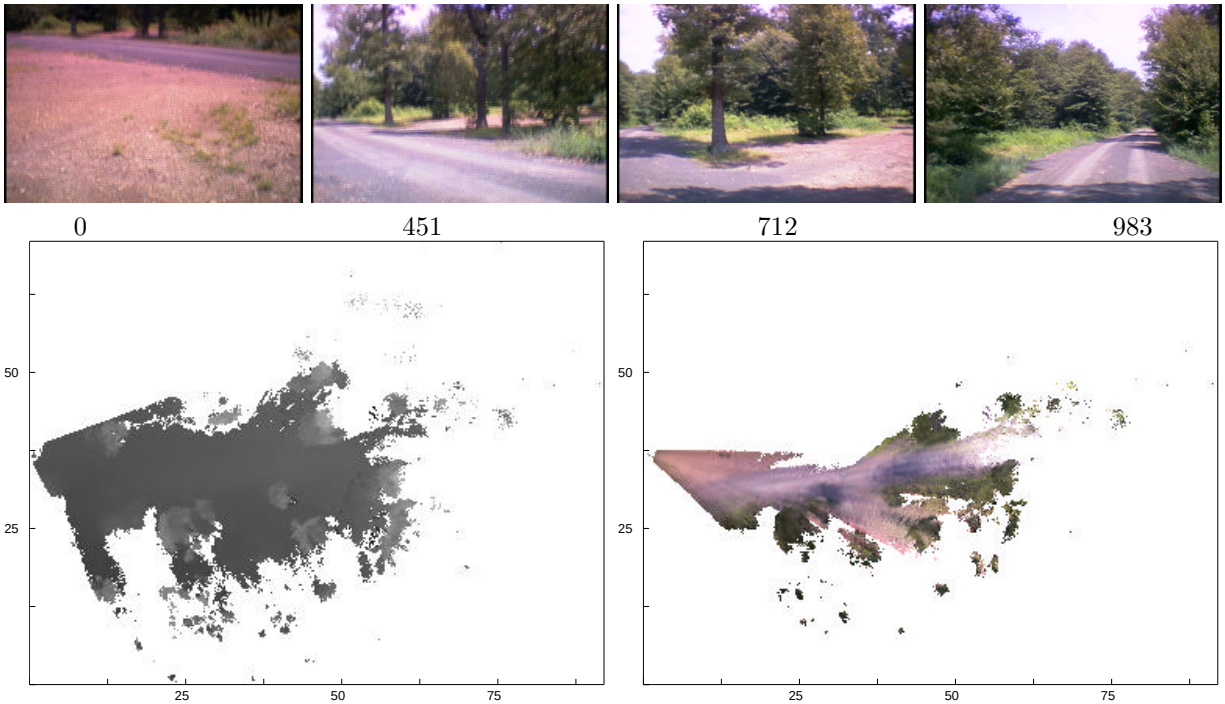
[12] G. Piccioli, De Micheli, and M. Campani, "A robust method for road sign detection and recognition," in *Proc. European Conf. Computer Vision*, 1994, pp. 495–500.

[13] J. Fernandez and A. Casals, "Autonomous navigation in ill-structured outdoor environments," in *Proc. Int. Conf. Intelligent Robots and Systems*, 1997.

[14] T. Lee, "Image representation using 2D Gabor wavelets," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 959–971, 1996.



(a)



(b)

Figure 8: Road detection: (a) Field height and color map; (b) Trees height and color map.