Proceedings of the SPIE 16th Annual International Symposium on Aerospace/Defense
Sensing, Simulation, and Controls, Orlando, FL, April 1-5, 2002.

Road Detection and Tracking for Autonomous Mobile Robots

Tsai-Hong Hong, Christopher Rasmussen, Tommy Chang, and Michael Shneier
Intelligent Systems Division
National Institute of Standards and Technology
Gaithersburg, MD 20899

ABSTRACT

As part of the Army's Demo III project, a sensor-based system has been developed to identify roads and to enable a mobile robot to drive along them. A ladar sensor, which produces range images, and a color camera are used in conjunction to locate the road surface and its boundaries. Sensing is used to constantly update an internal world model of the road surface. The world model is used to predict the future position of the road and to focus the attention of the sensors on the relevant regions in their respective images. The world model also determines the most suitable algorithm for locating and tracking road features in the images based on the current task and sensing information. The planner uses information from the world model to determine the best path for the vehicle along the road. Several different algorithms have been developed and tested on a diverse set of road sequences. The road types include some paved roads with lanes, but most of the sequences are of unpaved roads, including dirt and gravel roads. The algorithms compute various features of the road images including smoothness in the world model map and in the range domain, and color features and texture in the color domain. Performance in road detection and tracking are described and examples are shown of the system in action.

1 INTRODUCTION

An autonomous vehicle intended for driving off-road (e.g., for military reconnaissance) should still be able to identify roads and to drive along them when conditions allow. This ability will minimize terrain-based dangers and maximize speed. Road following requires an ability to discriminate between the road and surrounding areas and is a well-studied visual task[1-5]. The work described in this paper is part of the Army's Demo III project[6]. 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) [7] architecture, which divides the system into perception, world modeling, and behavior generation subsystems.

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 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.

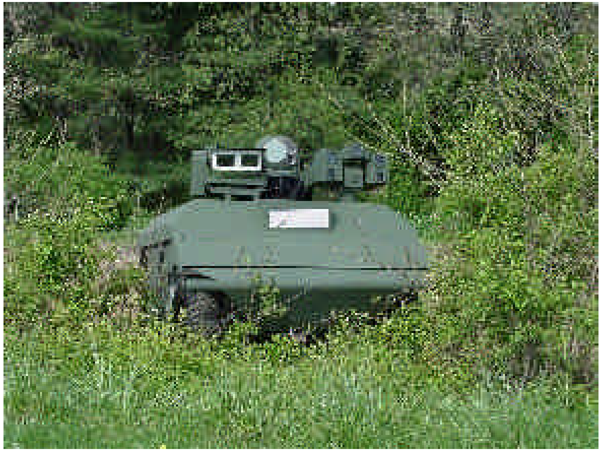

Figure 1. The Demo III XUV

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. Prediction 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[8]. 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 the regions that are of interest can be isolated, special and perhaps expensive processing can be applied to them without exceeding the available computing resources.

One way that "focus of attention" systems can work is by looking for features defined by some explicit or implicit model. The search may take many forms, from multi-resolution approaches that emulate human vision's peripheral and foveal vision, to target-recognition methods that use explicit templates for matching [9-14]. 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.

Another way focus of attention systems can work is by selecting the most appropriate sensory processing algorithm for achieving a given task. Criteria for selecting the most appropriate algorithm may include environmental factors (weather, day or night, road condition, road class, etc.) and the type of processing to be performed on the sensed information. For example, if the task is daytime driving on highways, the sensory processing system must find lane lines in daylight. A rule-based system, for example, would use these constraints to select the most appropriate algorithm to perform the task.

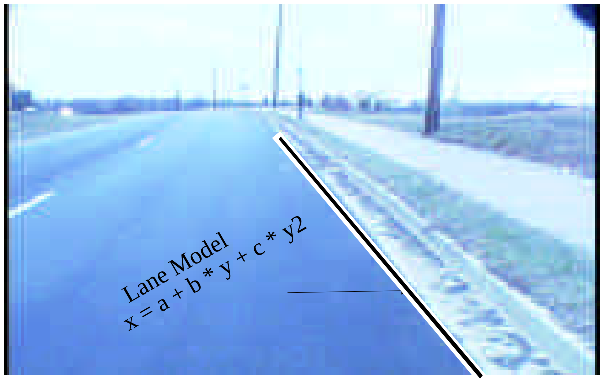

Figure 2. Lane model in a road with lanes

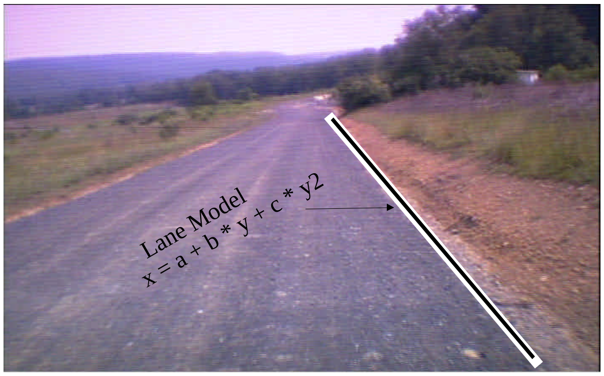

Figure 3. Boundary model of a road

In this paper, we describe a road and lane detection and tracking method that falls within the above general description, but differs from previous approaches in using multiple sensor types that interact to locate and identify roads. The way each sensor is used in conjunction with another sensor and with the functions of vehicle's internal world model is the focus of the paper. Further, a world model containing the system's current best guess about the state of the world and the task is used to predict where roads should appear and how they should look to each of the sensors. The world model also helps determine the most appropriate algorithm to apply to the sensor data.

2 THE WORLD MODEL FOR ROAD AND LANE TRACKING

We briefly introduce the world model we use for road extraction and tracking. The world model contains a representation, or map, of the current state of the world surrounding the vehicle and is updated continually by the sensors. We use a modified occupancy grid representation[15], with the vehicle centered on the grid, and the grid tied to the world. The world model 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 world model process has three primary functions in the road and lane following tasks.

1. To create models of road elevation, road boundaries, and lanes within the road, and to keep them current and consistent. In this role, it updates elevation and variance of elevation in maps and road/lane models (see Figure 2., Figure 3) in accordance with inputs from the sensors. It also assigns (multiple) confidence factors to the models and all map data, and adjusts these factors as new data are sensed.

2. To generate predictions of expected sensory input based on the current state of the world and estimated future states of the world. For the road boundary and lane marker following application, we assume that the location of the road boundary or lane marker in the image domain changes very little (less than 10 pixels) between successive images. Based on the road model, the predicted road boundary or lane is projected onto the current image. This provides the focus of attention region for the sensory processing module.

3. To determine the most suitable algorithm for the sensory processing module based on the current state of the world and the task. The system’s state is updated every cycle based on prior knowledge and the state of the world model[16], For lane marker following, the lane detection algorithm is used to find lane markers, and the edges most likely correspond with each lane marker are used to update the lane model (see section 3.1). For road following, the road detection algorithm (see section 3.2) is used to segment the road.

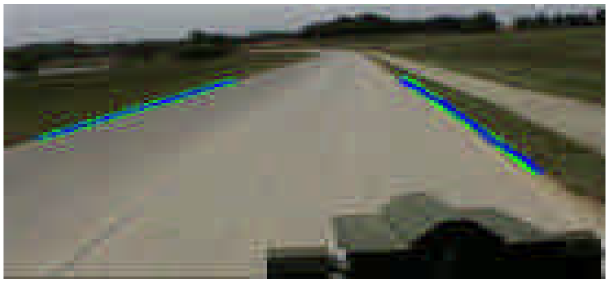

Figure 4. Initial road boundary. Blue points are road edges, green points show the model fitted to the boundary

3 SENSORY PROCESSING FOR ROAD AND LANE TRACKING

3.1 Lane tracking algorithm

The algorithm used for lane tracking is similar to that described in Schneiderman and Nashman[17]. The stages of the algorithm involve first predicting the locations of lane markers, then extracting and classifying edges, and, finally, updating the lane model. The lanes are represented by 2D models in the image plane. An initial approximate model of the lane markers is determined by fitting a second order polynomial (Equation 1) to the set of edge points labeled as lane marker points.

(1) $$x = c_1 + c_2 y + c_3 y^2$$

The parameters c_1, c_2, c_3, determine the shape and position of the lane marker model. We assume that the vehicle is located on the road and that some lane markers appear in the image. First, we project the vehicle's heading direction into the image, as a line through the center of the vehicle and up through the image. For each projected point in the vehicle heading line, we search to the right in each row to find the pixel with the highest intensity. This pixel is labeled as a lane marker. Once the initial model is constructed, it is used to predict where the lane marker will appear in the next video image.

In each new image, edge extraction is performed only in the focus area, a square box 10 _ 10 pixels in size that is predicted to contain lane markers. For every point in the focus area, the edge magnitude and gradient are computed using 3 _ 3 Sobel operator. The edge in each image row that has approximately the same gradient direction as predicted by the lane marker model and has the greatest magnitude is selected and classified as a lane marker. The blue data points in Figure 4. are classified as lane markers.

Using the points classified as lane markers, a second order polynomial model of each lane is updated recursively, using the square root information filter (SRIF)[18]. The algorithm is outlined below.

Given an over-determined linear equation:

(2) $$A\vec{x} = \vec{b} + \vec{e}$$

where $\vec{e}$ is the error in measurement. QR decompose A (using householder transformation):

(3) $$A = QR$$

Q is an orthonormal matrix and R is an upper triangular matrix. Since Q is orthonormal, its inverse is simply itself:

(4) $$Q^{-1} = Q$$

Multiply both sides of (3) by Q^{-1}:

(5) $$\begin{aligned} Q^{-1}A\vec{x} &= Q^{-1}(\vec{b} + \vec{e}) \\ R\vec{x} &= Q(\vec{b} + \vec{e}) \end{aligned}$$

For fitting a quadratic polynomial, the dimension of R is $(n \times 3)$ where n is the number of data points. Since R is upper triangular, only the first 3 rows contain non-zero elements. Letting $R_{3\times 3}$ denote the upper matrix and letting $\vec{z} = Q\vec{b}$, rewrite equation (5):

(6) $$R\vec{x} = \begin{bmatrix} R_{3\times3} \\ 0_{m\times3} \end{bmatrix}\vec{x} = \begin{bmatrix} z_{3\times1} \\ z_{m\times1} \end{bmatrix} + \vec{e}_2$$

Since Q is orthonormal:

(7) $$\begin{aligned} \|\vec{e}_2\| &= \|Q\vec{e}\| = \|\vec{e}\| \\ \|\vec{e}_2\| &= \|R\vec{x} - \vec{z}\| = \vec{e}_2^T\vec{e}_2 \\ &= (\vec{z}_{3\times1} - R_{3\times3}\vec{x})^T(\vec{z}_{3\times1} - R_{3\times3}\vec{x}) + \vec{z}_{m\times1}^T\vec{z}_{m\times1} \end{aligned}$$

Only the first term above depends on $\vec{x}$, which can be solved from the system by backsubstitution directly -R_{3x3} is already triangular. The error of the fit is:

(8) $$\|\vec{e}_2\| = \vec{z}_{m\times1}^T\vec{z}_{m\times1}$$

The least squares solution is updated with next batch of data, $A_2, \vec{b}_2$, by forming the system:

(9) $$A = \begin{bmatrix} \sqrt{(\lambda)}R_{3\times3} \\ A_2 \end{bmatrix}, \vec{b} = \begin{bmatrix} \sqrt{(\lambda)}z_{3\times1} \\ b_2 \end{bmatrix}$$

where $\lambda \in [0,\ldots,1]$ is the exponential weight[19]. The total error is:

(10) $$e_{total} = e_{total} + \|\vec{e}_2\|$$

3.2 Road tracking algorithm

Our algorithm for road detection uses a statistical classification method applied to data from both the ladar sensor and the color camera. The algorithm has three steps:

1. Classification of the road surface based on the variance of elevation in the world model map. When ladar data are inserted into the map, each ladar point recursively updates the elevation and variance of elevation[20] in the map cell to which it projects. In each map cell, variances either within a single cell or across a small region of neighboring cells are computed. If the variance of the elevation is small, the map cell is classified as possibly being road and the ladar point data is classified as a possible road point.

2. Road verification based on color information. Each laser point that was classified as a possible road point in step 1 is projected into color image domain. If the projected color region is classified as road in the color domain (Figure 5) the confidence of "roadness" of the map cell is increased. The algorithms for classification based on color will be described in section 3.3.

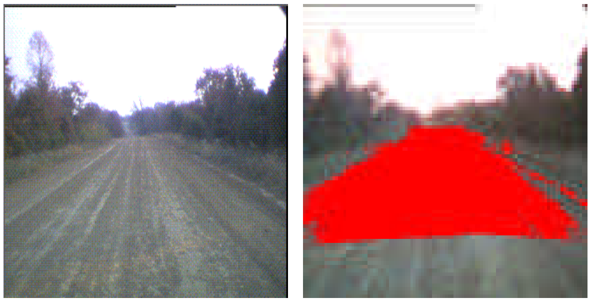

Figure 5. Color image and projected area of both ladar and color label as road

3. Update the road confidence of each map cell. Road segments whose classifications as road were supported in the color image have their confidence increased. Those that are not verified have their confidence reduced. When the road confidence rises above a threshold, the cell is confirmed as road (see Figure 6). If the confidence drops to 0, the road label is removed.

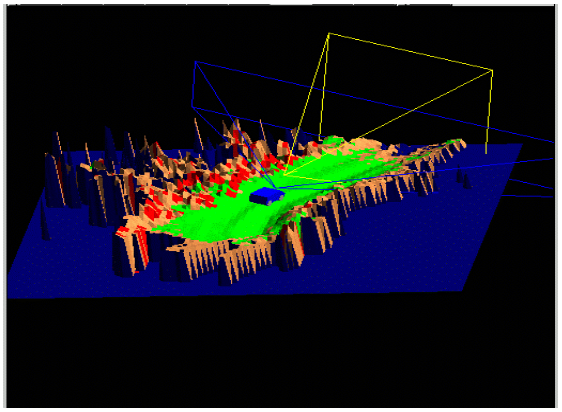

Figure 6. The 3D display of the map of the world model. Green is for road. Red is for obstacle and yellow is unknown.

This sequence of operations is repeated for each new ladar image and the corresponding color image. Since the ladar and color camera are mounted on the vehicle, some of the projections used in steps one to three are fixed, and can be computed offline.

There are two kinds of projections: Ladar data are projected into the world model map, and color images are projected into the map. 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 world model 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 navigation sensor, color camera and ladar sensor sample at different times. We use linear interpolation of readings from the vehicle’s navigation system to obtain the positions of the color camera and ladar sensor each time the sensor data are read.

The ladar-to-world-model coordinate transformation includes the ladar-to-vehicle and vehicle-to-world-model transformations. The projection from the ladar to the color image includes the world-model-to-ladar and world-model-to-image transformations. The ladar to color image transformation is particularly important for achieving 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. Bouguets's Matlab toolbox[21]. 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{Elstrom, Smith, et al. 1998 ELSTROM1998A /id}. Results are shown for a sample scene in Figure 7.

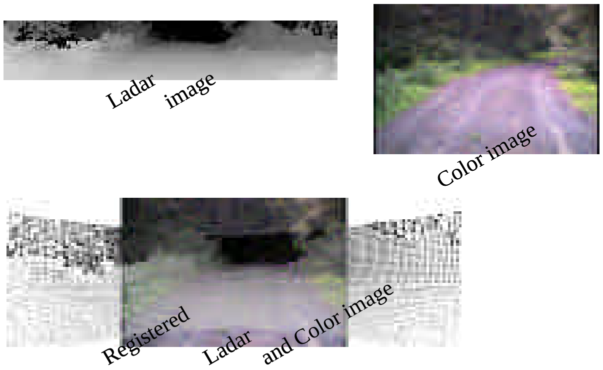

Figure 7. The mapping between Ladar data and color data

3.3 Road detection using color and ladar

Our road segmentation system works by combining depth information from a laser range-finder and color and texture image cues to segment ill-structured dirt, gravel, and asphalt roads as input to shape analysis module. We frame road segmentation as a classical classification problem in which we wish to identify small patches over the field of view as either road or non-road on the basis of a number of properties, or *features*, that we compute from them. These features are non-geometric: image location is not considered for segmentation, only local image properties. For data, a large number of registered laser and camera images were captured at frame-rate while driving 8-24 km/h on a variety of rural roads over the course of 73 minutes. A random sample consisting of approximately 0.1% of the road images (downsampled to 360 x 240) were manually segmented.

Features were computed at 10-pixel intervals horizontally and vertically over small (31 x 31) subimages. Feature types were as follows:

(1) *Color*: an "independent" color histogram[22] consisting of 8 bins per channel. This feature was chosen on the assumption that roads would be more-or-less consistent in their mix of colors—generally brown or gray—while the background is expected to exhibit more green and blue colors to allow discrimination.

(2) *Texture*: odd- and even-phase responses of a bank of Gabor filters[23] histogrammed over three wavelengths and eight equally-spaced orientations. This feature was selected in order to characterize the magnitude and dominant direction of texturedness at different scales, with the expectation that the road is more homogeneous or anisotropic (e.g., it might contain tracks and ruts) than bordering plants.

(3) *Height/bumpiness*: the mean and variance along the vertical axis relative to the vehicle support surface of laser points projecting to the local camera subimage. Height should allow protruding objects such as bushes and trees to be eliminated regardless of their visual appearance, and smoothness is included because roads should be locally flat, while tall grass and loose rocks are bumpier.

Two approaches to modeling road appearance were used: in one, a single neural network was trained on all combinations of the features (color alone, color plus texture, texture plus height/bumpiness, etc.) for the sample images. A combination of color, texture, and height/bumpiness was found to exhibit the best mix of accuracy (approximate 93%) and consistency. Our other approach was to generate a small set of road models by training separate neural networks on labeled feature vectors clustered by road "type." By first classifying the type of a novel road image, an appropriate second-stage classifier was selected to segment individual pixels. This improved segmentation accuracy over the single-model approach while still retaining good generality. More information is available in reference[24].

4 CONCLUSIONS

A system has been described that detects and tracks lane markers and road boundaries using color vision and ladar. The system makes use of a continually updated world model to predict where the sensors should focus their attention to ensure that the vehicle can drive fast enough. By predicting which regions of future images will contain the most useful information based on the current world model, the complexity of sensor processing is reduced and the signal to noise ratio of the sensor data is increased. By dynamically selecting the most appropriate sensory processing algorithms for achieving a given task, the system is able to track lanes and different types of road boundaries as appropriate. Therefore, the system can track lane markers on paved roads and switch to tracking road boundaries on unpaved dirt and gravel roads. The system is also currently being implemented on an indoor robot for a lane marker tracking and wall following task[25].

ACKNOWLEDGMENT

The research is a jointly sponsored effort which leverages technology from the NIST Intelligent Machines Initiative program and receives financial support from the Army Research Lab DemoIII UGV program and the AUTONAV research project agreement between the German Ministry of Defense and the US Department of Defence.

REFERENCES

1. Dickmanns, E. D. Vehicles capable of dynamic vision. International Joint Conference on Artificial Intelligence, 1577-1592. 1997.

2. Pomerleau, D. RALPH: Rapidly adapting lateral position handler. Proceedings IEEE Intelligent Vehicles Symposium, 506-511. 1995.

3. Lin, X. and Chen, S. Color image segmentation using modified HIS system for road following. IEEE Conference on Robotics and Automation, 1998-2003. 1991.

4. Davis, L. and Kushner, T. Road boundary detection for autonomous vehicle navigation. SPIE Vol. 579[Intelligent Robots and Computer Vision]. 1985.

5. Crisman, J. and Thorpe, C. UNSCARF, a color vision system for the detection of unstructured roads. Proceedings International Conference on Robotics and Automation, 2496-2501. 1991.

6. Shoemaker, C. M. and Bornstein, J. A. The Demo3 UGV Program: A Testbed for Autonomous Navigation Research. Proceedings of the IEEE International Symposium on Intelligent Control. 1998. Gaithersburg, MD.

7. Albus, J. S. and Meystel, A., *Engineering of Mind: an introduction to the science of intelligent systems* New York: John Wiley and Sons, 2001.

8. Clark, J. J. and Ferrier, N. J., "Attentive Visual Servoing," in Blake, A. and Yuille, A. (eds.) *Active Vision* Cambridge, MA: MIT Press, 1992, pp. 137-154.

9. Ullman, S., "Visual Routines," *Cognition*, vol. 18 pp. 97-159, 1984.

10. Toyama, K. and Hager, G., "Incremental focus of attention for robust vision-based tracking," *International Journal of Computer Vision*, vol. 35, no. 1, pp. 45-63, 1999.

11. Eklundh, J.-O., Nordlund, P., and Uhlin, T. Issues in active vision: attention and cue integration/selection. Proc. British Machine Vision Conference, 1-12. 1996.

12. Stough, T. M. and Brodley, C. E., "Focusing attention on objects of interest using multiple matched filters," *IEEE Transactions on Image Processing*, vol. 10, no. 3, pp. 419-426, 2001.

13. Westin, C.-F., Westelius, C.-J., Knutsson, H., and Granlund, G. Attention control for robot vision. Proc. Conference on Computer Vision and Pattern Recognition, 726-733. 1996. IEEE.

14. Ennesser, F. and Medioni, G., "Finding Waldo, or focus of attention using local color information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 805-809, Aug.1995.

15. Oskard, D. N., Hong, T., and Shaffer, C. A. Real-Time Algorithms and Data Structures for Underwater Mapping. Proceedings of the SPIE Advances in Intelligent Robotics Systems Conference. 1988. Boston, MA.

16. Albus, J. S. 4-D/RCS: A Reference Model Architecture for Demo III. NISTIR 5994. 3-1-1997.

17. Schneiderman, H. and Nashman, M., "A Discriminating Feature Tracker for Vision-Based Autonomous Driving," *IEEE Transactions on Robotics and Automation*, vol. 10, no. 6, 1994.

18. Bieman, G., *Factorization Methods for Discrete Sequential Estimation* New York: Academic Press, 1977.

19. Goodwin, G. and Sin, K., *Adaptive Filtering, Prediction and Control* Englewood Cliffs, New Jersey: 1984.

20. Hong, T., Abrams, M., Chang, T., and Shneier, M. O., "An Intelligent World Model for Autonomous Off-Road Driving," *Computer Vision and Image Understanding*, 2000.

21. Bouguest, J. Camera Calibration Toolbox for Matlab. www.vision.caltech.edu/bouguestj/calib.doc . 2001.

22. Swain, M. J. and Ballard, D. H., "Color Indexing," *International Journal of Computer Vision,* vol. 7, no. 1, pp. 11-32, 1991.

23. Lee, T. S., "Image representation using 2D Gabor wavelets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 959-971, 1996.

24. Rasmussen, C. “Combining Laser Range, Color, and Texture Cues for Autonomous Road Following.” Proceedings of the International Conference on Robotics and Automation. 2002.

25. Bostelman, R., Juberts, M., Szabo, S., Bunch, B., Evans, J., “Industrial Autonomous Vehicle Project Report,” NISTIR 6751, June 2001.