Road Following in an Unstructured Desert Environment using Monocular Color Vision as Applied to the DARPA Grand Challenge

Jaesang Lee*, Carl D. Crane III*
Sanggyum Kim **and Jungha Kim **

* Center for Intelligent Machines and Robotics, University of Florida, Gainesville, Florida
(Tel : +1-352-392-9461; E-mail: ccrane@ufl.edu)
** Graduate School of Automotive Engineering, Kookmin University Seoul, Korea
(Tel : +82-2-910-4715; E-mail: jhkim@kookmin.ac.kr)

Abstract: This paper describes the development of an autonomous ground vehicle that is being developed to participate in the October 2005 DARPA Grand Challenge. The authors of this paper are members of Team CIMAR which was one of twenty five teams selected by DARPA to participate in March 2004 in the inaugural competition to develop an autonomous vehicle that can navigate from near Los Angeles to near Las Vegas at speeds averaging twenty miles per hour. Most of the event was held on open terrain and trails in a rocky desert environment. This paper describes ongoing activities in preparation for the October 2005 event with emphasis placed on the utilization of color monocular vision to identify regions of smooth terrain that can be navigated at high speed.

Keywords: autonomous vehicle, navigation, vision system, image segmentation,

1. INTRODUCTION

1.1 2004 Event Description

The Grand Challenge was established by the Defense Advanced Research Projects Agency (DARPA) in order to encourage researchers to accelerate the development of autonomous vehicle technologies that can be applied to military requirements. The inaugural event held in March 2004 consisted of three parts, i.e. (1) application and acceptance into the event, (2) qualification, inspection, and demonstration (QID), and (3) the actual race from Barstow, CA to Primm, NV. The team that was able to complete the course first within a ten hour time frame would be awarded a prize of one million dollars.

Twenty five teams were invited to participate in the event from a total of approximately eighty technical reports that were submitted. Fifteen of the teams, including Team CIMAR, were judged to have passed the QID and were selected to participate in the actual race event. Two hours before the start of the race, each team was given a data file that contained approximately three thousand waypoints along with a corridor width for each pair of waypoints. Teams could use the two hour period to plan a path for the vehicle based upon any a priori data such as trails. The first vehicle to start on the course departed at 6:30 am. Other vehicles were started at five minute intervals. No vehicle completed the course. The furthest distance traveled was approximately seven miles.

1.2 Vehicle and System Architecture for 2005 Event

A new vehicle has been designed and built for the upcoming 2005 event. Figure 1 shows the vehicle, named the NaviGator, that was built by the company Georgia All-Terrain Monsters. Engineers at the Eigenpoint Co. added actuators to the vehicle to automate the steering, accelerator, brakes, and transmission.

The system architecture to be used on the vehicle consists of components that have been developed and documented in the U.S. Department of Defense Joint Architecture for Unmanned Systems (JAUS) Reference Architecture as well as additional components that were designed specifically for this event. A schematic of the architecture is shown in Figure 2.

Fig. 1 NaviGator vehicle

Briefly, the architecture can be divided into four primary elements. The components that are encircled by the blue dashed line specifically perform closed-loop control in order to keep the vehicle on a specified path. The components that are encircled by the red dashed line perform the sensing tasks required to locate obstacles and to evaluate the smoothness of terrain. The components encircled by the green dashed line act to determine the 'best' path segment to be driven based on the sensed information. Lastly, the components encircled by the yellow dashed line act as a repository for a priori data such as known roads, trails, or obstacles, as well the waypoint corridor information that was supplied two hours prior to the start of the race.

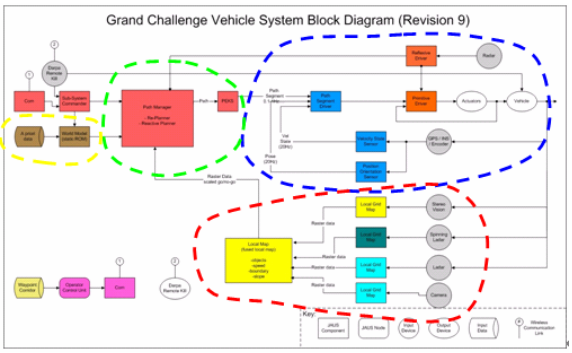


Fig. 2 System architecture

2. COLOR MONOCULAR VISION SYSTEM

A critical element of the system design is the selection and incorporation of sensors to detect obstacles and to evaluate the smoothness of the terrain. A total of five sensor types are being installed on the vehicle, i.e. a spinning ladar, stereo vision, stationary ladar, radar, and monocular vision.

The focus of this paper is on the development of the monocular vision system to identify regions of smooth terrain that can be navigated at high speeds.

2.1 Pre-processing

The algorithm first works by preprocessing an image to eliminate the portion of the image that corresponds to the sky. Prior to the segmentation of the image, a proper selection of the most suitable color space becomes an important part of the process. In this project, the sky portion of the image hinders the classification procedure. Considering the sky portion not only slows down the image processing speed unnecessarily, but also presents problems with the texture-based classification results as sometimes parts of the sky image may be classified as road. (Figure 3 shows some examples from desert images)

Fig. 3 Sample road image

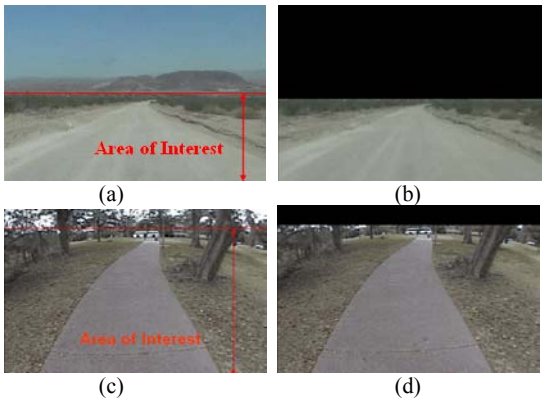
A HSV (Hue, Saturation, Value) color model is used for detecting the sky part of the image as the Hue channel is an excellent discriminator between the sky and ground (see Figure 4). Otsu's algorithm [1], which uses a histogram approach, is used to segment the sky portion of the image because of its simplicity [2].

Even if this method finds an optimal threshold value for dichotomizing pixels into two classes, i.e. road and sky, in the Hue histogram it is possible that road area pixels are classified as sky pixels since many scenes contain noise caused by road and lighting conditions. To resolve this problem, every scene is searched to find the maximum sky value starting from the top of the dichotomized image until the road part of the image is met. This procedure prevents misclassification individual pixels of the road area of the image as being part of the sky region.

Figure 5 shows the result where the pixels associated with the sky portion of Figure 5(b) are assigned the value of zero (black in RGB). This process helps to remove long distance background objects such as mountains and trees.

(a) (b)

Fig. 4 (a) Oiginal RGB image (b) Hue channel image



(a) (b)

(c) (d)

Fig. 5 (a)(c) Area of interest (b)(d) Application to image

2.2 Selecting relevant features

The RGB (Red, Green, and Blue) color space is the standard representation in the world of computers and digital cameras and is therefore often a natural choice for color representation. Also RGB is the standard output from a CCD-camera. Since roads typically have a different color than non-drivable terrain, color is a highly relevant feature for segmentation.

2.3 Training set generation

In this project, to reduce the computational expense of processing large images, the dimensions of the scene are reduced from the original digital input of 720 × 480 pixels to a 320 × 240 reduced image. Next, a sub-image which has 20% of the reduced image height and width is used to define the drivable area and two sub-images which have 10% of the height and 20% of the width of reduced image are used to define the background. Figure 6 shows the selected drivable-area by a blue box and the two selected non-drivable background areas by red boxes.

Fig. 6 Drivable area and non-drivable area

2.4 Road Following Algorithm

A Bayesian decision theory approach was selected for use as this is a fundamental statistic approach to the problem of pattern classification associated with this application. It makes the assumption that the decision problem is posed in probabilistic terms, and that all of the relevant probability values are known. The basic idea underlying Bayesian decision theory is very simple. However this is the optimal decision theory under Gaussian distribution assumption [7]. Therefore, most of pixel classification is done by Bayes classifier.

In this project, the decision boundary that was used is given by

$$\frac{1}{(2\pi)^{d/2}|\Sigma_1|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^T\Sigma_1^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)\right] = \frac{1}{(2\pi)^{d/2}|\Sigma_2|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^T\Sigma_2^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)\right] \quad (1)$$

where μ_1 and Σ_1 are the mean vector and covariance matrix of the drivable-area R,G,B pixels in training data and μ_2 and Σ_2 are those of background pixels.

The decision boundary is simplified as follows:

$$(\mathbf{x}-\boldsymbol{\mu}_1)^T\Sigma_1^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)+\ln|\Sigma_1| = (\mathbf{x}-\boldsymbol{\mu}_2)^T\Sigma_2^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)+\ln|\Sigma_2| \quad (2)$$

In most pixel classification problems, the logarithm components in Eq. (2) is not a dominant factor for classification and therefore these two values are not computed to save time since this application requires a real time implementation. So, RGB pixels of color X which belong to a class are computed based on the distance $(\mathbf{x}-\boldsymbol{\mu}_1)^T\Sigma_1^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)$ [4].

2.3 Segmentation Method

1) Pixel-based segmentation:

After computing the threshold value from the discrimination method, each pixel in an image is then classified as drivable road area or non-drivable background. Although this is a simple approach, classifying every pixel is time consuming and can lead to results that have significant noise (see Figure 7). Often, post processing is used to reduce the noise effect. In this application, however, a block-based segmentation procedure is used which is discussed next.

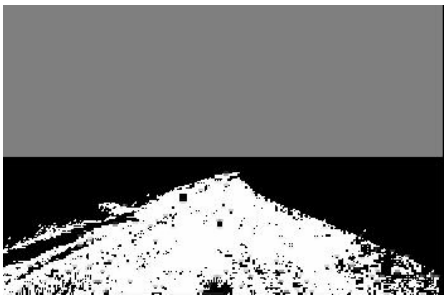


Fig. 7 Pixel-based classification result

2) Block-based segmentation:

A block-based segmentation method is used to reduce the segmentation processing time. 9 × 9 pixel regions are clustered together and replaced by their RGB mean value as calculated by

$$\mu^L_{(x,y)} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}P^L{}_{(i,j)} \quad (3)$$

where μ is new pixel mean value for 9 x 9 block, P is raw pixel data, (i, j) is the raw pixel orientation, (x, y) is the new pixel orientation, $L \in \{1,2,3\}$ for RGB , and N is block size.

The clusters, or blocks, are then segmented, and the result as shown in Figure 8 has less noise. Also the segmentation process is accomplished 3.5 times faster than pixel-based classification with a 320 × 240 image. A disadvantage, however, is that edges are blurred and not as distinct. However, in NaviGator vision system application, an offset of 9 pixels corresponds to 9.9 cm in the bottom of the image.

Fig. 8 Block-based classification

2.4 Averaging

When the vehicle is operating autonomously, many unforeseen events may occur. Significant factors related to the vision processing task is the change in light (intensity) conditions that may be caused by a cloud, a shadow such as caused by auto-iris lens response, a tree, or the change in the type of road or terrain that is encountered, for example a change from a paved to a dirt road. A primary objective of this project is to achieve color constancy for accurate classification.

In the ALVINN (Autonomous Land Vehicle In a Neural Network) project at Carnegie Mellon University, normalizing RGB is used to suppress the effects of shadows [5]. It heightens pixel contrast within images and decreases variations in overall intensity between images. However,

normalizing RGB data has not in itself led to consistency in classification.

The approach used here is to maintain a buffer of statistical information such as the mean and covariance of each RGB channel, in order to reduce the sharp change. For example Figure 9 shows two kinds of plots. The first dotted lines are the mean of the RGB channels for a series of frames. The second solid line shows the mean value of a frame averaged with the mean values of the preceding 10 frames. This averaging, or buffering, reduces rapid changes in the calculated mean value, and thus the corresponding classification threshold value. Figure 10 shows the regular block-based segmentation result compared with the buffered segmentation result where more of the image is now classified correctly. It works better at the edge of the road. The buffer size is determined empirically and it depends on the vehicle speed.

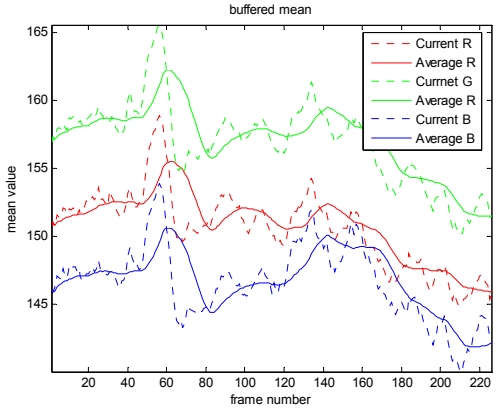


Fig. 9 Buffered mean and current frame mean

(a)

(b)

Fig. 10 (a) Block-based classification result
(b) Block-based with buffering method classification result

3. CONFIDENCE FACTOR

Even though the performance of the system to date has proven to be reliable, there will be occasions when poor or even erroneous output may occur and these cases need to be identified. Ideally it would be nice to have the Bayesian probability for a correct output. In practice, this is clearly impossible. However, since images are typically acquired at a rate of 10 frames/sec, more confidence will be placed in segmentation results that do not vary much between subsequent frames. The exception here, however, is if two subsequent images result in very few pixels classified as road area. If this is the case, then the confidence factor that is calculated and associated with each block will be lowered.

The final goal of the vision system is to send traversability information to a Smart Arbiter component. The Smart Arbiter is charged with intelligently fusing the input data from disparate sensors into the basis for real-time, on-board planning and decision-making [6]. It takes raster-oriented grid information from the sensors and fuses them into a composite grid of obstacles and traversability values. The vision system estimates traversability based on the classification result.

It was stated that there should be a correlation between frames because the scene changes slowly with a camera acquiring 10 frame/sec. For this camera acquisition speed, the vehicle can move 0.89 m per frame when traveling at a speed of 20 miles per hour. For this reason we will have to settle for a more approximate estimation of confidence. The confidence factor was found to depend on two primary variables. i.e. the number of pixels classified as road in the image and the difference in the classification of a block between consecutive frames. Further when the vehicle drives on a very rough gravel road, the images are shaken and out of focus. In this case, the sky border changes rapidly and this must also be considered. Figure 11 illustrates how a confidence factor is determined for the classification results in and between the images.

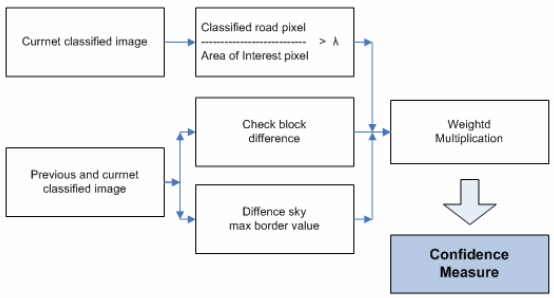


Fig. 11 Block diagram of the confidence estimation

4. SUMMARY AND CONCLUSION

Evaluation of the performance of the original vision system that was used in the inaugural 2004 DARPA Grand Challenge identified the need for an improved ability to classify terrain. The developments associated with the use of monocular vision to be used in the upcoming 2005 event have been presented in this paper, and the results to date have been promising. Continued advancements in the areas of mobility, sensing, data interpretation, and planning will ultimately make the vision of autonomously navigating vehicles a reality.

ACKNOWLEDGMENTS

The authors would like to acknowledge Smiths Aerospace who is the primary sponsor of the Team CIMAR entry into the 2005 DARPA Grand Challenge. The authors would also like to acknowledge the support of the University of Florida, Autonomous Solutions, Inc., and the Eigenpoint Company.

REFERENCES

[1] Otsu, N., "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, PP 62-66, 1979.

[2] R.Guo, S.M. Pandit, "Automatic threshold selection based on histogram modes and a discriminant criterion", *Machine vision and Application*, PP.331-338, 1998.

[3] Rand C. Chandler, "Autonomous Agent Navigation Based on Textural Analysis", 2003.

[4] Charles Thorpe, Martial H. Hebert, Takeo Kanade, "Vision and Navigation for the Carnegie-Mellon Navlab.", *IEEE Transactions on Pattern Analysis and Machine Intelligence.* Vol. 10, No. 3, May 1988.

[5] J. Hancock, C. Thorpe. "ELVIS: Eigenvectors for Land Vehicle Image System", *tech. report CMU-RI-TR-94-43,* Robotics Institute, Carnegie Mellon University, December, 1994.

[6] Carl D. Crane III, David G. Armstrong, Mel W. Torrie and Sarah A. Gray, "Autonomous Ground Vehicle Technologies Applied to the DARPA Grand Challenge, *ICCAS2004,* 2004.

[7] R.D. Morris, X. Descombes and J. Zerubia, "Fully Bayesian image segmentation - an engineering perspective" *In Proceedings of IEEE International Conference on Image Processing,* Santa Barbara, October 1997.