Visual Routines for Vehicle Control

Garbis Salgian and Dana H. Ballard

Computer Science Department
University of Rochester
Rochester, NY 14627 USA

1. Introduction

Automated driving holds the promise of improving traffic safety, alleviating highway congestion and saving fuel. The continuous increase in processor speed over the last decade has led to an increased effort in research on automated driving in several countries [1]. However, autonomous tactical level driving (i.e. having the ability to do traffic maneuvers in complex, urban type environments) is still an open research problem.

As little as a decade ago, it was widely accepted that the visual world could be completely segmented into identified parts prior to analysis. This view was supported in part by the belief that additional computing cycles would eventually be available to solve this problem. However the complexity of vision’s initial segmentation can easily be unbounded for all practical purposes, so that the goal of determining a complete segmentation of an individual scene in real time is impractical. Thus to meet the demands of ongoing vision, the focus has shifted to a more piecewise and on-line analysis of the scene, wherein just the products needed for behavior are computed as needed. Such products can be computed by visual routines [14], special purpose image processing programs that are designed to compute specific parameters that are used in guiding the vehicle.

This paper describes the development and testing of visual routines for vehicle control. It addresses the generation of visual routines from images using appearance based models of color and shape. The visual routines presented here are a major component of the perception subsystem of an intelligent vehicle. The idea of visual routines is compelling owing to the fact that being special-purpose vast amounts of computation can be saved. For this reason they have been used in several simulations (eg. [9]), but so far they have been used in image analysis only in a few restricted circumstances.

2. Photo-realistic simulation

Autonomous driving is a good example of an application where it is necessary to combine perception (vision) and control. However, testing such a system in the real world is difficult and potentially dangerous, especially in complex dynamic environments such as urban traffic.
Given recent advances in computer graphics, both in terms of the quality of the generated images and the rendering speed, we believe that a viable alternative to initial testing in the real world is provided by integrating photo-realistic simulation and real-time image processing. This allows testing the computer vision algorithms under a wide range of controllable conditions, some of which would be too dangerous to do in an actual car. The resultant testbed leads to rapid prototyping.

Terzopoulos pioneered the use of simulated images in his animat vision architecture. However, in their approach all the processing is carried out in software, one of the motivations for the architecture being that it avoids the difficulties associated with "hardware vision" [13]. In our case, the graphical output from the simulator is sent to a separate subsystem (host computer with pipeline video processor) where the images are analyzed in real-time and commands are sent back to the simulator (figure 2.1). The images are generated by an SGI Onyx Infinite Reality engine which uses a model of a small town and the car. Visual routines are scheduled to meet the temporary task demands of individual driving sub-problems such as stopping at lights and traffic signs. The output of the visual routines is used to control the car which in turn affects the subsequent images. In addition to the simulations, the routines are also tested on similar images generated by driving in the real world to assure the generalizability of the simulation.

The simulator can also be used with human subjects who can drive a kart through the virtual environment while wearing head mounted displays (HMD). A unique feature of our driving simulator is the ability to track eye movements within a freely moving VR helmet which allows us to explore the scheduling tradeoffs that humans use. This provides a benchmark for the automated driver and also is a source for ideas as to priorities assigned by the human driver. In particular, the fixation point of the eyes at any moment is an indicator of the focus of attention for the human operator. Experiments show that this fixation point can be moved at the rate of one fixation every 0.3