Proc. 1994 Government Microcircuit Applications Conference, pp. 358-362, San Diego, CA, Nov. 1994

DEFENSE AND CIVILIAN APPLICATIONS OF THE ALVINN ROBOT DRIVING SYSTEM

Dean **A.** Pomerleau

School of Computer Science
Carnegie Mellon University
Pittsburgh, **PA 15213**

Abstract

Perhaps the most successful development of the ARPA Unmanned Ground Vehicle (UGV) program is the ALVINN autonomous driving system. ALVINN is a neural network system that locates the road in images from a video camera mounted on a UGV and steers the vehicle to follow it. ALVINN has been demonstrated on several HMMWV test vehicles driving at speeds of up to 70 mph, and for distances of over 90 miles without human intervention.

ALVINN was originally designed as part of an unmanned vehicle for the modern battlefield, performing reconnaissance, surveillance and nuclear, biological and chemical (NBC) detection missions. However we are currently adapting the system for civilian use, as part of the Intelligent Vehicle Highway System (IVHS) initiative. This paper describes the ALVINN system, and its potential applications to both defense and civilian applications. It also describes the the technical issues involved in converting the system from military to civilian operation, and the steps we are taking to address them.

Introduction

An application domain which has proven quite amenable to artificial neural networks is vision-based autonomous driving. In this domain, the objective is to steer a robot vehicle, like the Navlab shown in Figure 1, based on input from an onboard video camera. The vehicle is equipped with motors on the steering wheel, brake and accelerator pedal, enabling computer control of the vehicle's trajectory. A typical autonomous driving task which has been addressed using artificial neural networks is road following[1,2]. For this task, the input consists of images from the video camera, and the output is a steering command which will keep the vehicle on the road.

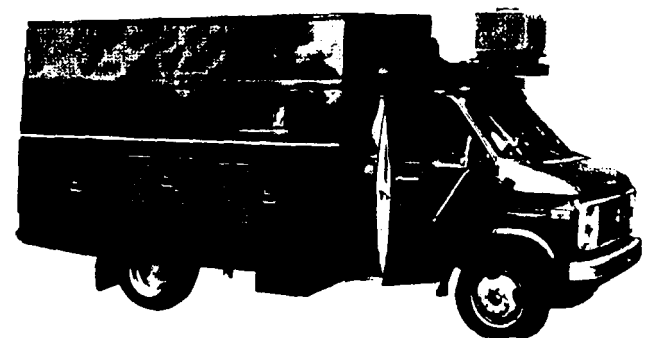

Figure 1: The CMU Navlab autonomous navigation testbed vehicle.

Neural Network Model

The connectionist model for autonomous road following used in the ALVINN system is the feedforward multi-layer perceptron shown in Figure 2. The input layer consists of a single 30x32 unit "retina" onto which a video image is projected. Each of the 960 input units is fully connected to the four unit hidden layer, which is in turn fully connected to the output layer. Each of the 30 output units represents a different possible steering direction. The centermost output unit represents the "travel straight ahead" condition, while units to the left and right of center represent successively sharper left and right turns.

To drive the Navlab, an image from the video camera is reduced to 30x32 pixels and projected onto the input layer. After propagating activation through the network, the output layer's activation profile is translated into a vehicle steering command. Instead of simply selecting the output with the highest activation level, as is typically done in classification tasks

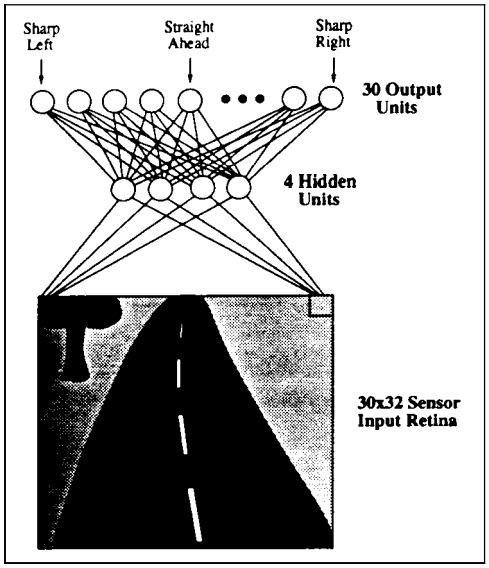

Figure 2: Architecture of the network designed for autonomous driving.

like character recognition and speech recognition, the steering direction dictated by the network is taken to be the center of mass of the "hill" of activation surrounding the output unit with the highest activation level. Using the center of mass of activation instead of the most active output unit to determine the direction to steer permits finer steering corrections, thus improving ALVINN's driving accuracy. This distributed output representation also makes generalization to similar situations easier for the network, since slightly shifts in the road position in the input image lead to only slight changes in any single output unit's target activation.

Training "On-the-Fly"

The most interesting and novel aspect of the ALVINN system is the method used to train it. In this technique, called training "on-the-fly", the network is taught to imitate the driving reactions of a person. As a person drives, the network is trained with back-propagation using the latest video image as input and the person's steering direction as the desired output.

To facilitate generalization to new situations, variety is added to the training set by shifting and rotating the original camera image in software to make it appear that the vehicle is situated differently relative to the road ahead. The correct steering direction for each of these transformed images is created by altering the person's steering direction for the original image to account for the altered vehicle placement. So for instance, if the person were steering straight

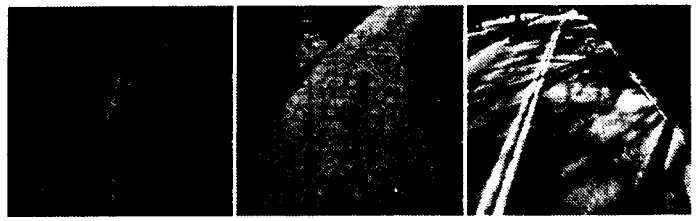

Figure 3: Video images taken on three of the roads ALVINN has been trained to handle.

ahead, and the image were transformed to make it appear the vehicle is off to the right side of the road, the correct steering direction for this new image would be to steer towards the left in order to bring the vehicle back to the road center. Adding these transformed patterns to the training set teaches the network to recover from driving mistakes, without requiring the human trainer to explicitly stray from the road center and then return.

ALVINN Driving Performance

Running on two Sun Sparcstations onboard the Navlab, training on-the-fly requires about two minutes during which a person drives over about a 1/4 to 1/2 mile stretch of training road. During this training phase, the network typically is presented with approximately 50 real images, each of which is transformed 15 times to create a training set of 750 images.

Once it has learned, the network can accurately traverse the length of the road used for training, and also generalize to drive along parts of the road not encountered during training under a variety of weather and lighting conditions. In addition, since determining the steering direction from the input image merely involves a forward sweep through the network, the system is able to process 10 images per second, and drive at up to 55 mph. This is over five times as fast as any non-connectionist system as driven using comparable hardware (Crisman, 1990; Kluge, 1990).

The flexibility provided by the neural network has allowed ALVINN to learn to drive in a wide variety of situations. Individual networks have been trained to drive on single-lane dirt and paved roads, two-lane suburban and city streets, and multi-lane divided highways. Images taken from three of these domains are shown in Figure 3. On the highway, ALVINN has driving for up to 90 miles without human intervention.

For each road type, the feature detectors the network develops are slightly different. For instance, for dirt roads the network develops "rut'' detectors.

For unlined paved roads the network developed edge detectors and detectors for trapezoidal shaped road regions. For lined highways, the network develops detectors for the lane markings. This ability to adapt its processing to fit the situation makes ALVINN more flexible than other autonomous driving systems

3.4

ALVINN-on-a-chip

In the current implementation of ALVINN, video images are digitized into a frame store. Processing to perform the neural computations is done in software running on an attached workstation. The need to acquire and move large amounts of image data from the frame store to the workstation severely limits the frame throughput of the current ALVINN implementation — only 10 frames/second can be processed. While we have shown this to be sufficient for autonomous driving at up to 55 mph, higher processing rates are required if further gains in the speed and performance of the driving task are to be obtained.

Latency is another serious problem for ALVINN when video cameras are used. The images processed by the workstation are those taken by the camera several frames back. In other words, vehicle heading is determined from road conditions that are not current. While pipelining can improve system throughput, the latency in an imaging system built around a frame store cannot be eliminated. Latency is a serious problem for vision-based control systems, severely limiting their stability. Applications like ALVINN that are sensitive to the real-time nature of the images, will therefore be limited in performance.

To overcome these problems, we are are currently integrating the sensing and processing performed by ALVINN on a single VLSI chip. The partitioning of the network in the ALVINN-on-a-chip implementation is shown in Figure 4. The chip has an array of transducers for imaging the road ahead, and analog circuitry for performing the computationally expensive the matrix-vector product necessary to propagate activation from the input to the hidden layer. After distilling the information from the image into a small number of hidden unit activations, these values are transferred off-chip and the limited number of operations required to propagate from the hidden to the output layer are performed using a conventional processor.

A more detailed design of each cell in the ALVINN-on-a-chip implementation is depicted in Figure 5. In

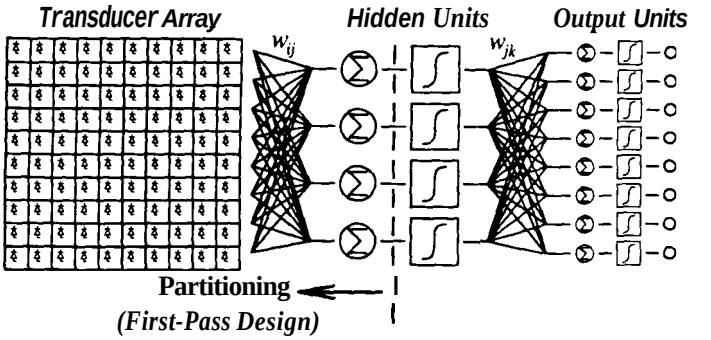

Figure 4: Partitioning the ALVINN task for VLSI.

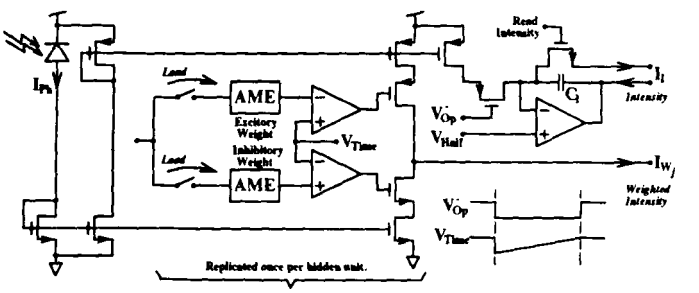

Figure 5: An ALVINN-on-a-chip cell implementation.

this design, each unit in the input layer has an associated photodetector on the chip. This detector generates a photocurrent I_{Ph} proportional to the incident light intensity level. For each hidden unit, this photocurrent is accumulated for a time proportional to the input to hidden unit weight, which is stored in the Analog Memory Elements (AME). The resulting weighted photocurrent is injected onto a line common to all cells, to sum up the input to each of the hidden units.

This ALVINN-on-a-chip system will be capable of 100 frames/second or more, a performance level that will make control of the vehicle far more robust. This will permit safe travel of the autonomous vehicle at speeds higher than that obtainable using the current implementation of ALVINN. The processing speed provided by the hardware implementation of ALVINN will allow us to simulate many network's simultaneously, each trained for a different road type. This will allow ALVINN to be more accurate, and better able to cope with changes in the roads appearance.

Applications

Even without an implementation in hardware, ALVINN has demonstrated its potential for both military and civilian applications. ALVINN has been a key technology in the successful demonstrations

conducted for the ARPA Unmanned Ground Vehicle (UGV) Demo II program[5]. The goal of this program is to demonstrate the utility of advanced UGV systems for DOD tasks such as reconnaissance, surveillance, target acquisition, and nuclear, biological and chemical (NBC) detection. As part of the first two Demo II demonstrations, conducted in July of 1993 and July of 1994, ALVINN successfully drove two specially equipped HMMWVs over dirt roads, and paved roads with and without lane markings, at speeds of up to 20 miles/hour.

Another military application of the ALVINN technology currently under investigation is to include it as part of the Off-Road Smart Mine Countermeasures (ORSMC) program. The goal of the ORSMC program is counter mines that use acoustic sensors to detonate only when tanks are nearby. The countermeasure concept in which ALVINN may play a role is an unmanned decoy HMMWV. ALVINN would drive the vehicle, while loudspeakers mounted on the vehicle generate tank sounds, designed to detonate the smart mines prematurely.

Civilian applications of the ALVINN system currently center around the domain of Intelligent Vehicle Highway Systems (IVHS). We are investigating the use of ALVINN as a lateral control device for the Automated Highway System (AHS). In this fully automated driving scenario, ALVINN would keep commuter vehicles in their lane on the highway will other systems control the vehicle's speed and prevent it from colliding with other cars.

Nearer term IVHS applications of the ALVINN system are in the area of collision warning and avoidance. As part of a US DOT sponsored program, we are currently developing a system based on ALVINN to warn drivers when they start to drift of the roadway. These types of accidents, called single vehicle roadway departures, account for over 1/3rd of all driving fatalities, and cause an estimated $100 billion in damages and delays each year.

For each of these applications, proof-of-concept systems based on ALVINN have been or are being developed. We have demonstrated that the ALVINN technology is capable of performing these tasks, but deployable solutions to these problems will require the ALVINN-on-a-chip implementation currently under development.

Acknowledgements

The principle support for the Navlab has come from DARPA, under contracts DACA76-85-C-0019, DACA76-85-C-0003 and DACA76-85-C-0002. This research is also funded in part by the US Department of Transportation, under contract no. DTNH22-93-C-07023.

References

[1] Pomerleau, D.A. (1991) Efficient Training of Artificial Neural Networks for Autonomous Navigation, *Neural Computation*, 3:88-97.

[2] Pomerleau, D.A. (1993) *Neural Network Perception for Mobile Robot Guidance*, Boston: Kluwer Academic Publishing.

[3] Crisman, J.D., and Thorpe, C.E. (1990) Color Vision for Road Following, in *Vision and Navigation: The CMU Navlab* (C.E. Thorpe, Ed.), Boston: Kluwer Academic Publishers, pp. 9-24.

[4] Kluge, K., and Thorpe, C.E. (1990) Explicit Models for Robot Road Following, in *Vision and Navigation: The CMU Navlab* (C.E. Thorpe, Ed.), Boston: Kluwer Academic Publishers, pp. 25-38.

[5] Chun, W.H., and Jochem, T.M. (1994) Unmanned Ground Vehicle Demo 11: Demonstration A, in *Unmanned Systems 12(1)*, pp. 14-20.