Panacea: An Active Sensor Controller for the ALVINN Autonomous Driving System

Rahul Sukthankar, Dean Pomerleau and Charles Thorpe

CMU-RI-TR-93-09

The Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

April 1993

This research was partly sponsored by DARPA, under contracts "Perception for Outdoor Navigation" (contract number DACA76-89-C-0014, monitored by the US Army Topographic Engineering Center) and "Unmanned Ground Vehicle System" (contract number DAAE07-90-C-R059, monitored by TACOM).

Contents

List of Figures

Abstract

Panacea is a modular system which incorporates a steerable sensor into an existing neural network driving system, ALVINN. A fixed camera cannot see the road when it makes sharp bends. For a vision system that builds a map of the road, it is straightforward to point the camera down the road; but ALVINN directly outputs a steering command without generating an intermediate road representation. Insight from the training scheme used in ALVINN, however, provides an interpretation of the steering command in terms of the road geometry and appropriate camera pointing strategies. Tests on the Carnegie Mellon Navlab II with a steerable camera have shown that the system significantly improves ALVINN's performance, particularly in situations requiring sharp turns and quick responses.

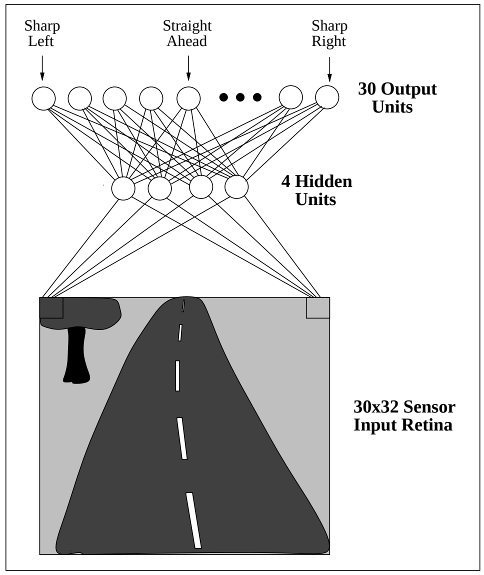

Figure 1: ALVINN driving network architecture.

1. Introduction

ALVINN (Autonomous Land Vehicle in a Neural Network) is a neural network based system which has been successful in driving robot vehicles in a variety of situations [1, 2]. However, since ALVINN maintains no state information about the world, but processes each sensor frame individually, it can become confused on sharp curves when the field of view no longer displays the important features in the scene. A steerable sensor allows the perception system to select the desired field of view to maximize the information content of a sensor frame [3]. For a vision system that builds a map of the road, it is straightforward to point the camera in the desired direction, but ALVINN directly outputs a steering command, without generating an intermediate road representation. Panacea interprets this steering command as a point on the road and pans the camera in the desired direction. However since ALVINN is trained with a fixed sensor orientation, the position of the sensor during training is implicitly encoded in the weights and moving the camera results in the outputs of the network being invalid for the given configuration. Panacea solves this problem by post-processing the steering response of the neural network as a function of the current sensor configuration. A significant advantage of this approach is that existing networks can run under this new system without any modification or retraining. Panacea was implemented on the Carnegie Mellon Navlab II and has demonstrated improved performance of ALVINN networks, particularly on roads with sharp curves.

2. ALVINN Architecture and Training

The ALVINN system's basic architecture is a three layered artificial neural network shown in Figure 1. A reduced resolution camera image is fed into a 30x32 array of input units, which are fully connected to a hidden layer of 4 units. The hidden units are fully connected to a vector of 30 output units, and the steering response is given as a Gaussian activation level centered on the correct steering curvature.

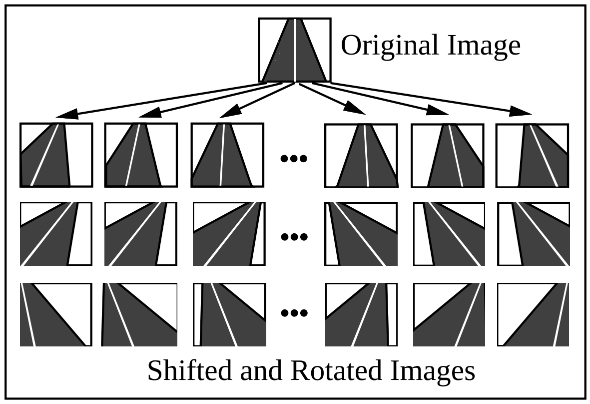

Figure 2: The single original video image is shifted and rotated to create multiple training exemplars in which the vehicle appears to be at different locations relative to the road.

ALVINN's neural net is trained "on the fly", and the human driver's steering responses are used as the teaching signal. ALVINN is able to learn from this limited data by artificially expanding its training set. Each original image is shifted and rotated in software to create 14 additional images in which the vehicle appears to be situated differently in relation to the road (See Figure 2). The training signal for each of these new images is calculated by assuming a pure-pursuit [4] model of driving and transforming the original steering response accordingly. One of the advantages of using a weak model like pure-pursuit is that it is independent of the driving situation. Figure 3 illustrates this model. With the vehicle at position A, the pure pursuit model assumes the goal is to bring the vehicle to the road center at the target point T, a predetermined distance ahead of the vehicle. After transforming the image with a horizontal shift s and rotation θ to make it appear that the vehicle is at point B, the appropriate steering direction according to the pure pursuit model should also bring the vehicle to the target point T. Mathematically, the formula to compute the radius of the steering arc that will take the vehicle from point B to point T is

$$r = \frac{l^2 + d^2}{2d} \tag{1}$$

where r is the steering radius, l is the lookahead distance and d is the distance from point T the vehicle would end up at if driven straight ahead from point B for distance l. The displacement d can be determined using the following formula:

$$d = \cos\theta \cdot (d_p + s + l\tan\theta) \tag{2}$$

where d_p is the distance from point T the vehicle would end up if it drove straight ahead from point A for the lookahead distance l, s is the horizontal distance from point A to B, and θ is the vehicle rotation from point A to B. The quantity d_p can be calculated using the following equation:

$$d_p = r_p - \sqrt{r_p^2 - l^2} \tag{3}$$

where r_p is the radius of the arc the person was steering along when the image was taken.

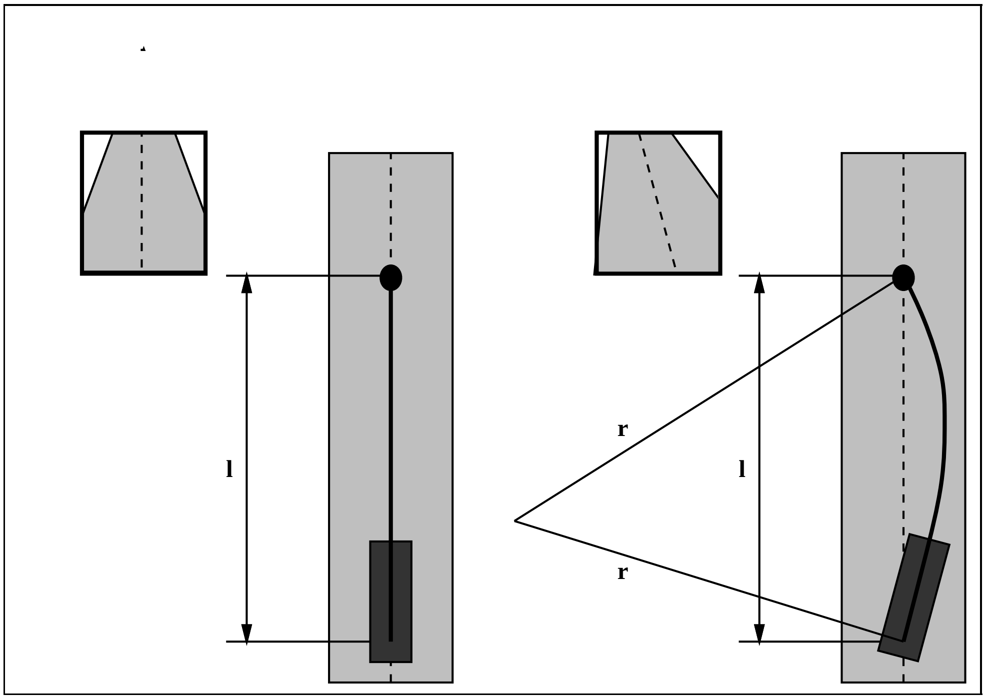

Figure 3: Illustration of the "pure pursuit" model of steering.

3. Panacea

Panacea uses the pure-pursuit driving model to adjust an existing ALVINN network's steering output in response to variations in sensor orientation. Since the model is also used internally by ALVINN during training, the same assumptions are made in the two modules. When used with a fixed sensor, both systems produce identical responses.

ALVINN outputs a steering response which can be symbolically interpreted as a turning radius, or a desired arc. In the pure-pursuit model, every such arc maps to a single *target point* TP, at the specified look-ahead distance from the sensor. Thus for a given vehicle pose, the position of the TP should remain invariant under changes in sensor orientation. In other words, the pure-pursuit model implies that there is a "correct" TP for the current vehicle pose, which is independent of the sensor pan. ALVINN's response is in sensor coordinates since it implicitly assumes that the camera is pointing directly ahead. However, since the sensor is not in its original orientation, the turning radius given by ALVINN no longer steers the vehicle towards the target point. Therefore we have to compensate for the change in sensor orientation, and generate the arc which correctly steers the robot towards the TP corresponding to the vehicle's actual position.

Panacea thus converts ALVINN's outputs into a target point representation, and generates the arc (in the current vehicle frame) which drives the robot towards the TP. Figure 4 illustrates this transformation. The equations for this transform are derived below:

$$d = r - \mathrm{sgn} r \sqrt{r^2 - l^2} \quad (4)$$

$$l' = (l-a)\cos\theta - d\sin\theta + a \quad (5)$$

$$d' = (l-a)\sin\theta + d\cos\theta \quad (6)$$

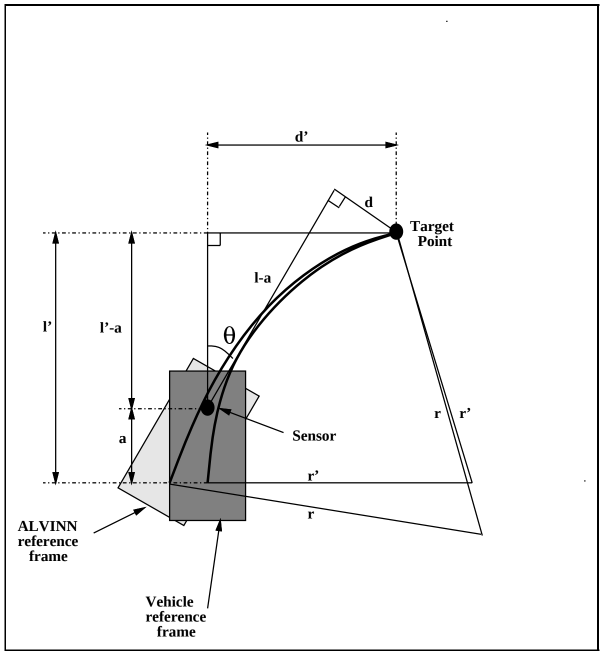

Figure 4: Sensor pan compensation using Panacea.

$$r' = \frac{d'^2 + l'^2}{2d'} \tag{7}$$

where r is the steering radius reported by ALVINN and r' is the compensated radius calculated by Panacea, while d and d' are the offsets. l' is the analog of l, ALVINN's lookahead distance, in the vehicle reference frame. The steering radius r' reported by Panacea is used to control the vehicle.

To gain a better understanding of the equations, a surface plot of the compensation against the input parameters was made. For clarity, turning radii were converted to curvatures, and the compensation expressed as the difference between the input and output curvatures. Figure 5 displays compensation as a function of input curvature and camera pan angle for two different lookahead distances. The graph on the left corresponds to a typical Navlab II configuration ($l = 10$ meters, $a = 3.3$ meters). The compensation seems to be independent of the input curvature, and varies proportionally with the camera pan angle over the values encountered in practice. However it is interesting to note that this is not true in general. The graph on the right shows the same surface with an extreme value for $l = 250$ meters. Note that the compensation is no longer independent of the input curvature. Although the implementation on the Navlab II could have been approximated using a planar model of the surface, the computational savings would be insignificant since the original equations are already quite simple. Therefore Panacea computes the precise compensation using equations 4 to 7.

4. Sensor Pointing

Panacea also addresses the issue of intelligent sensor control. ALVINN's output, which may be interpreted as a TP on the center of the road ahead of the vehicle, can be used to pan the camera in order to keep the road in view. The following equation relates the position of the TP to the pan

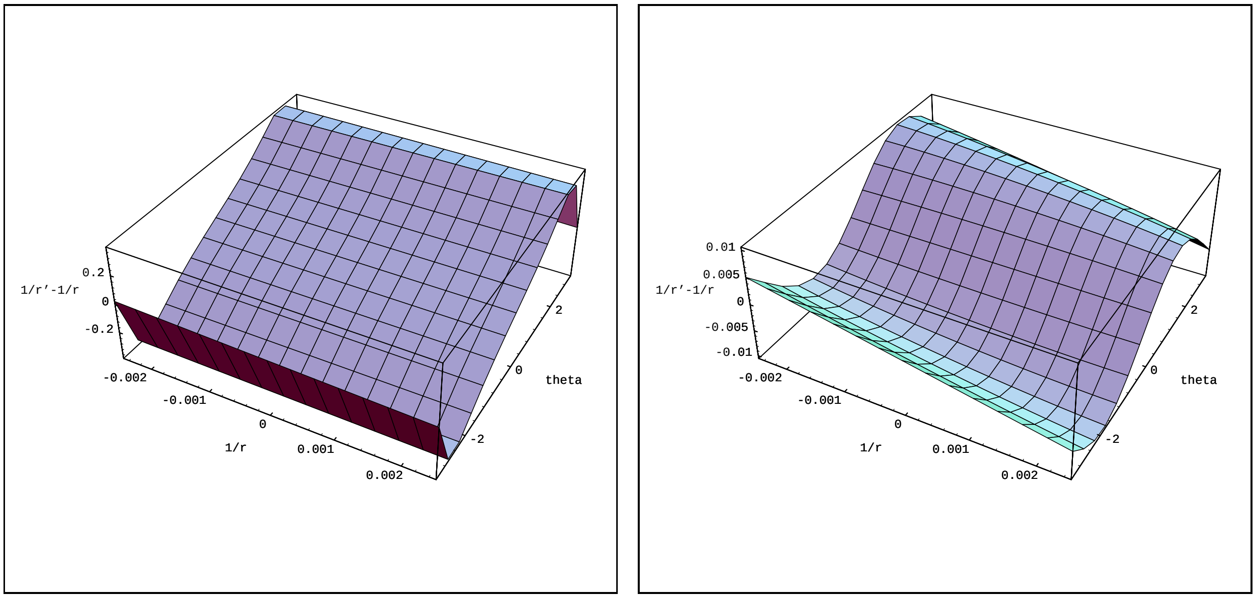

Figure 5: Curvature compensation with lookahead of 10m and 250m respectively.

angle:

$$\phi \quad = \quad \tan^{-1} \frac{d'}{l' - a} \tag{8}$$

where l' and d' are defined in Equations 5 and 6 respectively. This allows us to control the sensor directly from the output of our neural network, in a manner which is completely consistent with the pure-pursuit model. The actual implementation is somewhat complicated by control issues such as oscillations caused by the dynamics of the system. In practice this was solved by introducing a damping term which smoothed the sensor's response.

There are a number of advantages associated with controlling the sensor based on the network's output:

- By directing the sensor towards the TP, the important features of the scene as perceived by ALVINN are centered in the field of view.
- Images of this type are closer to those seen during training, and therefore accuracy of the network is increased.
- Since the sensor responds more quickly than the robot vehicle, the network is able to "look before it leaps".

Panacea is implemented so that the compensation for sensor displacement and the control of the sensor are decoupled. Thus ALVINN can drive the vehicle even when the sensor is being used to look at other features in its environment, such as signs, provided that the road remains at least partially in the field of view.

Figure 6: Panacea successfully negotiates a sharp fork in Schenley park.

5. Results and Discussion

This system was implemented on the Carnegie Mellon Navlab II, using a video camera on a pan/tilt mount (with constant tilt used throughout the experiments). Tests were conducted on a single-lane bicycle path, and on a two-lane street. The network was trained with the video camera pointing directly ahead. In the first experiment, the camera was offset at a constant angle and the vehicle switched to autonomous control. Panacea compensated correctly for the change in orientation and drove successfully. Subsequent tests were conducted with the sensor under Panacea's control and the system drove as reliably as the unmodified ALVINN system. A comparison between the two systems was then made at a sharp fork in the road (See Figure 6). With a fixed camera, ALVINN was unable to negotiate this stretch of the road. The main reason for ALVINN's difficulty in this situation is that road features on a sharply curved road fall outside a fixed camera's field of view. In addition, the robot vehicle reacts slowly to steering commands whereas a steerable sensor can pan fast enough to keep the road in sight at all times.

A sensor which pans under Panacea's control results in improved performance since the view seen by the sensor tends to correspond more closely to the images in the training set. Since the sensor points towards the TP, the important features in the scene are always within the field of view and the network is less likely to make steering errors. In particular, when the robot sees a fork in the road, the new system is less likely to dither over the decision since whichever road segment first appears most appropriate is immediately centered into the field of view, and the chance of the network choosing the other fork is thus substantially reduced. Higher level planning systems could exploit this by pointing the sensor in the appropriate direction at an intersection, causing ALVINN to choose one fork over another. This extension has not yet been implemented.

Panacea embodies the following beneficial attributes:

- Sound theoretical basis: Since Panacea uses the pure-pursuit model, which is already implicit in ALVINN, no additional assumptions are introduced. Furthermore, when the sensor configuration is static, the outputs of both systems are identical, so Panacea is transparent in that case.
- Modularity: Panacea is designed as a post-processing module for existing ALVINN systems. No additional time is required to train ALVINN driving networks. This also means that networks trained on a fixed sensor can be used without modification in the new system.

- Efficiency: The equations given above are very efficient, and the overhead of using Panacea on the ALVINN system is negligible.

6. Future Work

Panacea has shown that active perception and neural networks can be successfully integrated into a modular system for autonomous driving. Although the implemented system already demonstrates some advantages of this merger, there are many interesting topics which merit further exploration. In particular, the notion of decoupling the sensor motion from the driving network can be exploited further.

One application where it may be desirable to point the sensor at the TP without necessarily driving towards it is during obstacle avoidance. Here it is important that the video camera used for road following continue to focus its attention on the road, even during the temporary evasive maneuvering so that the driving algorithms can continue uninterrupted after the obstacle has been successfully avoided.

Conversely, an example where it may be desirable to point the sensor away from the center of the road, while continuing to drive towards it, is in road sign detection. This is also an example of how multiple systems could successfully share the same active sensor, since the ALVINN system, when augmented by Panacea, does not need the sensor to point at the center of the road as long as the relevant features remain visible in the sensor's field of view.

Although this paper focuses on Panacea as integrated into the ALVINN driving system, the same approach can be easily applied to any other road follower, as long as the system can provide information concerning the position of the road ahead of the vehicle.

Acknowledgements

This research was partly sponsored by DARPA, under contracts “Perception for Outdoor Navigation” (contract number DACA76-89-C-0014, monitored by the US Army Topographic Engineering Center) and “Unmanned Ground Vehicle System” (contract number DAAE07-90-C-R059, monitored by TACOM).

References

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

[2] Pomerleau, D.A. (1992): Neural Network Perception for Mobile Robot Guidance. Carnegie Mellon technical report CMU-CS-92-115.

[3] Turk, M., Morgenthaler, D., Gremban, K., and Marra, M. (1988): VITS — A Vision System for Autonomous Land Vehicle Navigation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 10, Number 3*

[4] Wallace, R., Stentz, A., Thorpe, C., Moravec, H., Whittaker, W., and Kanade, T. (1985): First Results in Robot Road-Following. *Proc. IJCAI-85*