Using Virtual Active Vision Tools to Improve Autonomous Driving Tasks

Todd M. Jochem

CMU-RI-TR-94-39

The Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

October 1994

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) as well as a DARPA Research Assistantship in Parallel Processing administered by the Institute for Advanced Computer Studies, University of Maryland.

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. government.

Abstract

[Note: This technical report is my unchanged thesis proposal document. It is published as a technical report so that it can be more easily referenced.]

ALVINN is a simulated neural network for road following. In its most basic form, it is trained to take a subsampled, preprocessed video image as input, and produce a steering wheel position as output. ALVINN has demonstrated robust performance in a wide variety of situations, but is limited due to its lack of geometric models. Grafting geometric reasoning onto a non-geometric base would be difficult and would create a system with diluted capabilities. A much better approach is to leave the basic neural network intact, preserving its real-time performance and generalization capabilities, and to apply geometric transformations to the input image and the output steering vector. These transformations form a new set of tools and techniques called Virtual Active Vision. The thesis for this work is:

Virtual Active Vision tools will improve the capabilities of neural network based autonomous driving systems.

Using Virtual Active Vision Tools to Improve Autonomous Driving Tasks

Todd M. Jochem

1. Introduction

ALVINN is a simulated neural network for road following. In its most basic form, it is trained to take a subsampled, preprocessed video image as input, and produce a steering wheel position as output. ALVINN has demonstrated robust performance in a wide variety of situations, but is limited due to its lack of geometric models. Grafting geometric reasoning onto a non-geometric base would be difficult and would create a system with diluted capabilities. A much better approach is to leave the basic neural network intact, preserving its real-time performance and generalization capabilities, and to apply geometric transformations to the input image and the output steering vector. These transformations form a new set of tools and techniques called **Virtual Active Vision**. My thesis is:

> **Virtual Active Vision tools will improve the capabilities of neural network based autonomous driving systems.**

I will demonstrate several new capabilities, using virtual active vision tools and ALVINN, and the CMU Navlab vehicles. These capabilities will include intersection detection and traversal, lane changing, and a variety of confidence measures.

1.1 Autonomous Driving Overview

Autonomous driving has matured to the point where there are many competent systems which can perform small parts of the problem very robustly and reliably. Systems have been developed using a wide variety of techniques which can drive autonomous vehicles on roads [2] [5] [10] [13], avoid obstacles [4] [6], plan safe paths for the vehicle to follow [1], and allow the vehicle to exhibit intelligent, goal directed behavior [9]. There have been attempts to connect these component modules into a robust, comprehensive autonomous driving system. Although many of these attempts have produced good results, there is much work to do in creating a competent, general purpose, autonomous driving system.

The foundation of any complete driving system is a robust road follower. The earliest road followers were strongly model based - they incorporated a priori knowledge of some important road feature (color, lane markings, etc.) and attempted to use this feature to locate the road. These systems performed well in situations where their built-in knowledge accurately characterized the road, but in cases where the feature they key on was not present or did not correctly define how to drive, failure was inevitable. In many cases the domain of operation was restricted so that the potential failure modes were never encountered. Although model based systems were not as robust as was

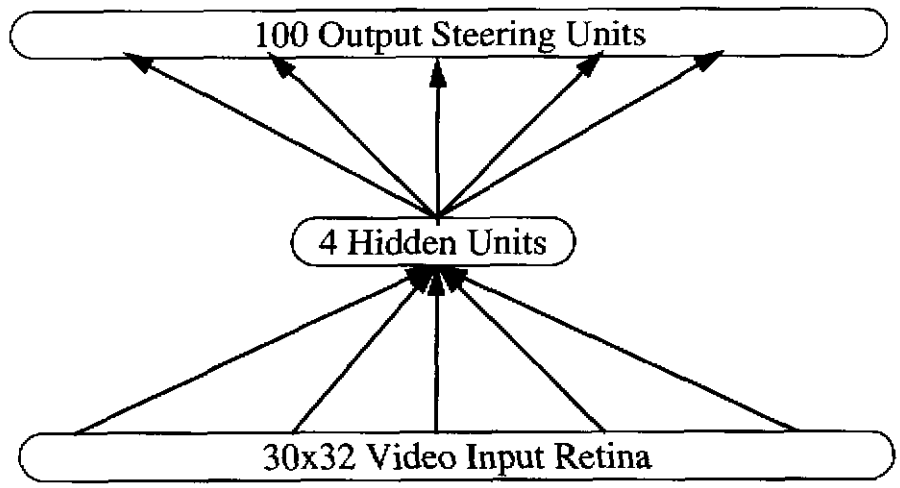

Figure 1. ALVINN network architecture.

desired, some of their characteristics, such as knowledge of where they thought the road was actually located, were ideas which were useful in merging these systems with higher level modules.

A major drawback of model based road following systems is that they are inherently tailored to finding one (or a few) features(s) which indicate(s) a road is present in an image. This problem was a factor which led to the development of neural network based road followers. The basis for using a neural system is that it can presumably learn the correct road model for any type of road. This gives the system the ability to drive on many road types instead of only ones which fit the model incorporated by the programmer into the system. Arguably the most successful neural network based road following system is ALVINN [10].

1.2 ALVINN Overview

ALVINN (Autonomous Land Vehicle In A Neural Network) has shown that neural techniques hold much promise for the field of autonomous road following. Using simple color image preprocessing to create a grayscale input image and a 3 layer neural network architecture consisting of 960 input units, 4 hidden units, and 100 output units, ALVINN can quickly learn, using backpropagation, the correct mapping from input image to output steering direction. See Figure 1. This steering direction can then be used to control our testbed vehicles, the Navlab and a converted U.S. Army HMMWV called the Navlab II.

ALVINN has many characteristics which make it desirable as a robust, general purpose road follower. They include:

- ALVINN learns the features that are required for driving on the particular road type for which it is trained.
- ALVINN is computationally simple.
- ALVINN learns features that are intuitively plausible when viewed by a human.
- ALVINN has been shown to work in a variety of situations.

ALVINN can learn to drive on lined city streets, jeep trails, and interstate highways and has successfully driven the Navlab II for over 90 miles at speeds exceeding 50 miles per hour. Because it has proven to be reliable over a wide range of road types and because it uses very basic input to produce its output, it is a natural choice on which to build more elaborate and comprehensive autonomous driving systems.

2. Towards a Comprehensive System

Previously, ALVINN aided high level modules in determining appropriate vehicle action in a very minimal way. Its primary function, besides staying on the road, was to change networks when the high level module requested. Some other high level module was in charge of determining when intersections and other road transitions occurred, figuring out which ALVINN network was required for the new road type, and relaying this information to ALVINN. ALVINN provided little [10] or no [12] feedback to the high level module about where it thought the transition was or how it was performing. This should not be the case. *Knowledge from high level modules should be used to guide, rather than force, lower level modules in their search for information relevant to satisfactory completion of high level goals. Additionally, lower level modules should play an important role in updating temporal and spatial information associated with appropriate vehicle action.* The research proposed in this document is strongly based on these two tenets.

To facilitate this work, a new sensor called a virtual camera is used. Developing more robust autonomous road following systems and improving driving performance is possible using virtual cameras. Virtual cameras are the fundamental tool upon which all other virtual active vision tools and techniques presented in this proposal are based. These tools create a natural link to high level modules and the important information that they contain as well as provide a mechanism for determining the appropriateness of vehicle actions. Finally, virtual cameras do not compromise the robust driving performance of the original ALVINN system.

All of the tools developed in this proposal lie in a field we call **Virtual Active Vision**. **Virtual** because all the methods use artificially created sensors which can be robustly manipulated to suit our needs and **Active** because the techniques move sensors to locations where the images they create will enhance system performance.

In order to more tightly integrate virtual cameras into a high performance driving system, one significant change to the basic ALVINN system is needed. This change is moving away from a system that produces a steering arc to one which produces the location of the center of the road or driving lane at a pre-specified distance in front of the vehicle. In effect, the new system will produce a point to drive over rather than an arc to drive. By using this point-based approach, ALVINN and other high level modules will have a common reference frame in which to communicate.

3. The Virtual Camera

A virtual camera is simply a camera which can be placed at any location and orientation in the world reference frame. It creates images using actual pixels imaged by a real camera that have

been projected onto some world model. By knowing the location of both the actual and virtual camera, and by assuming a flat world model, accurate image reconstructions can be created from the virtual camera location. A flat world model has been chosen as a first approximation of the actual world because in most road following scenarios, it accurately represents the world near the vehicle. Virtual camera views from many orientations have been created using images from three different actual cameras. The images produced by these views have proven to be both accurate and usable by the ALVINN system to navigate successfully. Figure 2 shows some typical virtual camera scenes.

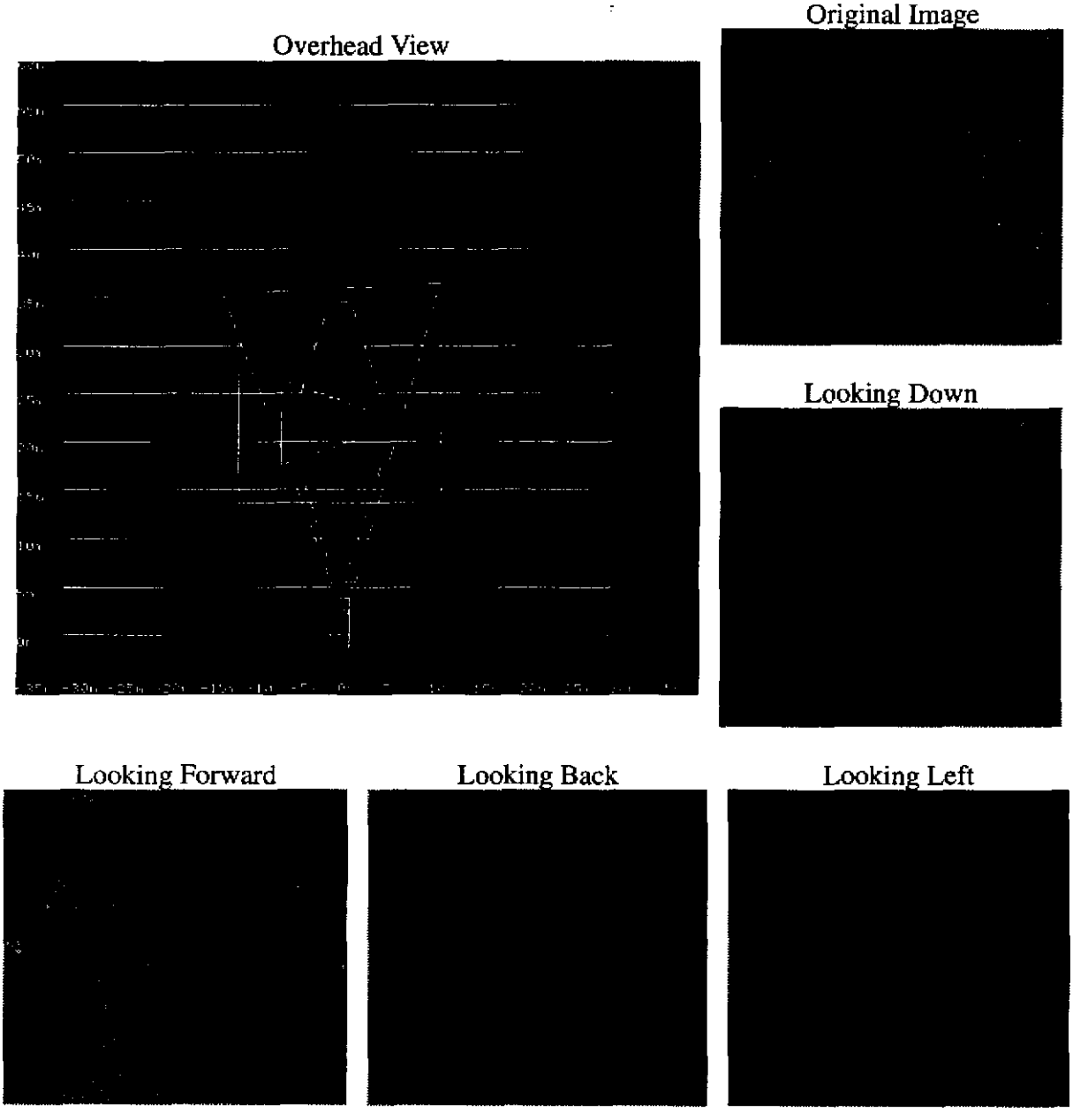

Figure 2. Typical virtual camera scenes.

An interesting issue that is a general theme of this proposal is showing that virtual cameras are well suited for merging neural systems with symbolic ones. A partial answer, and one which will be extensively explored, is that virtual cameras impose a model on the neural system. In our case, the model is not a feature in an image, but rather a canonical image viewpoint which ALVINN can interpret. This idea can be better understood by looking at a virtual camera from the point of view of both the high level module and of ALVINN.

From the standpoint of the high level system, a virtual camera isn't a camera, or even a sensing device, but instead an abstract object that ALVINN uses to get its job done. The single most important thing about a virtual camera to the high level module is that it can be placed at an arbitrary location in the real world. Location is a concrete concept which many high level systems use as an integral part of their operation. Location can be relative ("past the grocery store") as well as absolute ("sail to 10 latitude and 5 longitude"), and need not necessarily be specified by a number ("at the grocery store"). These kinds of ideas are all ones which high level systems handle well.

To ALVINN, the virtual camera is a sensing device. It is ALVINN's only link to the world in which it operates. ALVINN doesn't care where the virtual camera is located, only that it is producing images which can be used to locate the road. This interpretation may seem to trivialize ALVINN's functionality, but in reality, finding the road is what ALVINN is designed to do best. The virtual camera, guided by high level modules, insures that ALVINN gets images which will let it do its job to the best of its ability.

So in essence, the virtual camera imposes a model on the neural system without the neural system knowing, or even caring, about it. The model helps both the high level system as well as the neural system do their respective jobs better and allows them to seamlessly cooperate to exhibit goal directed, intelligent behavior.

4. System Overview and Comparison

A graphical comparison between the current ALVINN system and the system proposed in this document is shown in Figure 3. The main differences that should be noted are:

- ALVINN is insulated from the outside world in the new system by virtual active vision tools.
- There is more information flow from ALVINN to high level modules in the new system.

These differences and the reasons for them are described in the next paragraphs.

4.1 The Current ALVINN System

The current ALVINN system can be described as a tightly closed system. ALVINN directly senses the world and sends steering commands to the vehicle controller. This closed loop is necessary to achieve real-time performance but makes interacting with ALVINN especially difficult. As a consequence, providing ALVINN with information that allows it to exhibit intelligent, goal directed behavior has been done in a somewhat ad-hoc manner. Interpreting information from ALVINN has been even harder. New ideas about integration with other systems and improving performance tend to be tacked onto the sense-react cycle instead of integrated with it. Addition-

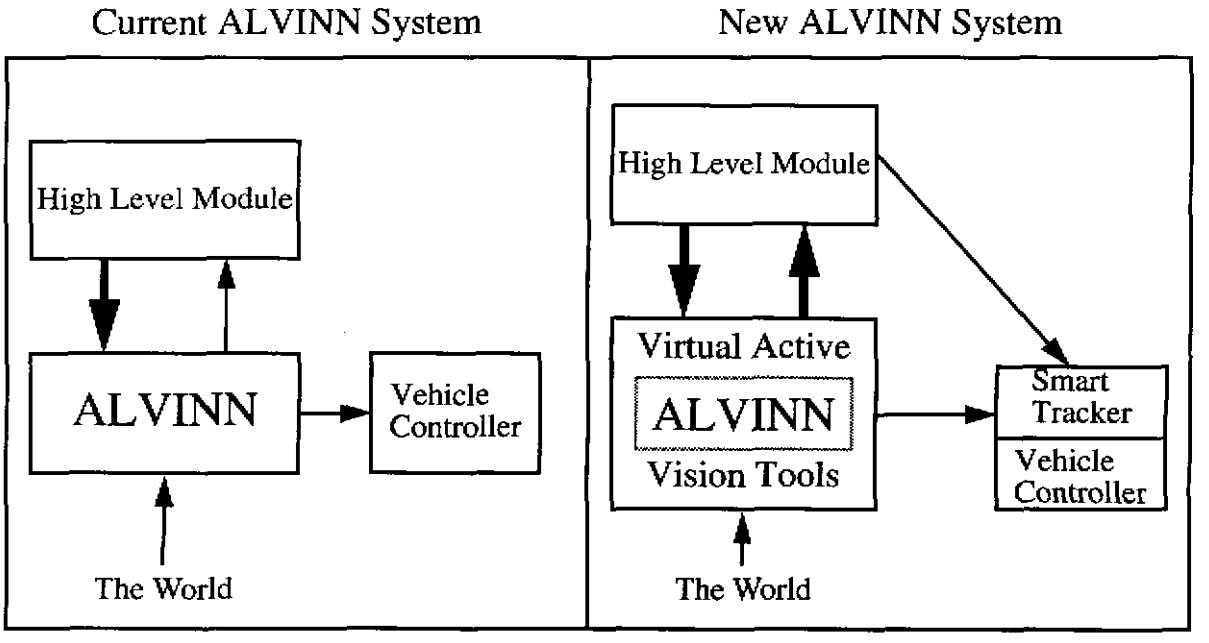

Figure 3. System Diagram. Arrows indicate direction of data flow. Arrow stem width indicates amount of data flow.

ally, this tightly closed sense-react processing loop makes it hard to determine how ALVINN is performing in manner that will not jeopardize real-time performance.

4.2 The New ALVINN System

In the ALVINN system proposed in this document, the tightly closed sense-react loop is maintained, but isolated from the world and high level systems by virtual active vision tools. These tools allow ALVINN and other systems to work in a common language - location. Data interpretation problems associated with heterogeneous (subsymbolic and symbolic) systems are overcome using this language. Virtual active vision tools still allow ALVINN to perform in real-time, but serve as a useful, bidirectional gateway to high level systems and the knowledge they contain. They allow ALVINN to 'focus its attention' on the tasks which are critical for achieving goal directed behavior while still maintaining the robust performance of the original ALVINN system. Using virtual active vision tools, ALVINN can overcome the problems which have hampered it in previous attempts at exhibiting goal directed behavior [12] [3] [10].

5. Other Similar Systems

An early attempt to use an artificial viewpoint to more easily interface low level robot control with high level functionality was reported by Wallace [14]. This work used a local flat world assumption to create a bird's-eye view (looking straight down) of the area in front of a mobile robot. From this viewpoint, the actual road structure became explicit, and when road edges were extracted, they could be easily matched to an actual map of the area.

More recently, Meng and Kak [8] have presented work on a system for an indoor mobile robot that couples low level, neural network based navigation systems with higher level, semantically based planning. Their system uses neural networks to identify when landmarks such as hallway junctions and deadends are reached, as well as to navigate correctly down the center of the corridor. They use parts of a Hough transform created from an edge image of the current scene as a training signal for each of the networks. The part of the Hough transform which each network sees is determined a priori by the researchers and corresponds to where relevant features in the Hough space a likely to appear. Each of their landmark detecting networks is trained to create a "near," "far," or "at" output. These outputs are then directly fed to the planner so that the robot can localize itself and replan if necessary. An interesting point about this work is that absolute position is never used; only relative, semantic positioning ("near the junction"). The claim is that humans use only relative positioning to navigate, but by giving up absolute locational information, many important task that we are attempting to address in this proposal, such as latency compensation and reliability estimation, become extremely difficult, if not impossible. While such a scheme may work for indoor mobile robots, it is insufficient for the road following task.

Work done by Lotufo [7] uses a bird's-eye view to simplify processing which must be done to detect road edges. In this work, an input image of a road is transformed onto the ground plane where specialized vertical edge detectors are applied. Because the road edges in the transformed image appear to be parallel, vertical lines, the edge detector can quickly and accurately find them.

6. Research Agenda

The research proposed in this document falls in the following areas:

- Improving Driving Performance.
- Improving Reliability Estimation.
- Merging Neural and Symbolic data.
- Active Vision and Sensor Fusion.

The general goal of the research proposed in this document is to develop methods that will allow autonomous vehicles to exhibit reliable and robust performance in complex, real world driving scenarios. The necessity of more robust performance in each of the above areas will become clear and potential research topics for achieving this end will be outlined. Each item can benefit from what is learned by exploring the others, but none depend solely on the development of another for individual success. As stated earlier, all of these topics fall into what we call Virtual Active Vision and use an already developed sensor called a virtual camera.

7. Improving Driving Performance

7.1 Optimal Camera Placement

One way that virtual cameras can improve ALVINN's driving performance is that they will allow ALVINN's neural network to be trained on a camera view from which it can learn the image to

road location mapping most quickly and accurately. Consider the case when the actual camera is oriented in such a way that the horizon line is in the image. Parts of the image which would ideally contain useful information about the road, now contain data which is not necessary for road following, such as trees and power lines. This extra, and often contradictory, information makes it more difficult for ALVINN to learn the correct mapping. By using a virtual camera, it is possible to create a view which contains only information that is useful for driving.

It is possible to find the optimal virtual camera view by training several networks using images created by virtual views in which the parameters which identify the pose of the virtual camera are changed. Virtual camera pose parameters which can be tested are the lateral offset from the center line of the vehicle (x), the virtual camera's height (z), the offset from the front of the vehicle (y), the virtual camera's tilt (pitch), and the virtual camera's yaw. Once each network has been trained, the mean error can be computed by comparing human driving performance with the output of the network. The virtual camera view which has the lowest mean error is judged as optimal for driving on the current road type. It cannot be judged as optimal for all road types because the location of important features for driving may change as road types change. For example, on an unlined bike path, it is reasonable to assume that the optimal camera offset is zero (i.e. centered) because both edges of the road are important for driving. But for driving on a lined city street, it may be more important to keep the yellow center line in the field of view at all times while sacrificing some of the outside road edge.

Additionally, selecting an optimal camera view is closely related to eliminating the need for adding structured noise to images during ALVINN training. [11] (Structured noise refers to a spatially coherent features in the image which should not be, but often are, mistaken for the real features which are required for driving.) Adding structured noise is required for driving in situations where transitory features could be mistaken for the actual features required for correct driving. For example, suppose a network is trained to drive on a typical four lane highway, having broad shoulders and a grassy median. During testing, it is observed that the network becomes confused when driving over a bridge that has concrete jersey barriers at the road side. Even though the lines marking the lanes (the real features required for correct driving) do not change, system performance degrades. To prevent this degradation, structured noise is added during training to the areas of the image where this type of transitory feature is likely to be encountered. By adding noise in these areas, the network will learn to only key off the important features required for driving. An alternative solution to this problem using virtual cameras is to pick a virtual camera view which does not include the area where transitory features occur.

7.2 Latency Compensation

Another potential advantage of using virtual cameras in ALVINN is for latency compensation. Latency compensation refers to modelling and correcting for the delay between when an image is digitized and when the steering command ALVINN's neural network produces in response to that image is executed. On the Navlab II this latency time is approximately 400 msec. This delay in response can result in control instability and oscillations, particularly when driving at high speeds. By correcting for this delay, better driving performance can be achieved.

Currently, to compensate for latency in ALVINN, modification of the training and driving processes is needed. Using a virtual camera positioned at the latency distance in front of the vehicle

and by training the neural network to produce the center of the road at its output rather than a steering arc, ALVINN's output can be directly used by a point tracking module to drive the vehicle. Additionally, virtual cameras provide a very simple way to adjust the latency distance dynamically, based on current vehicle speed. See Figure 4.

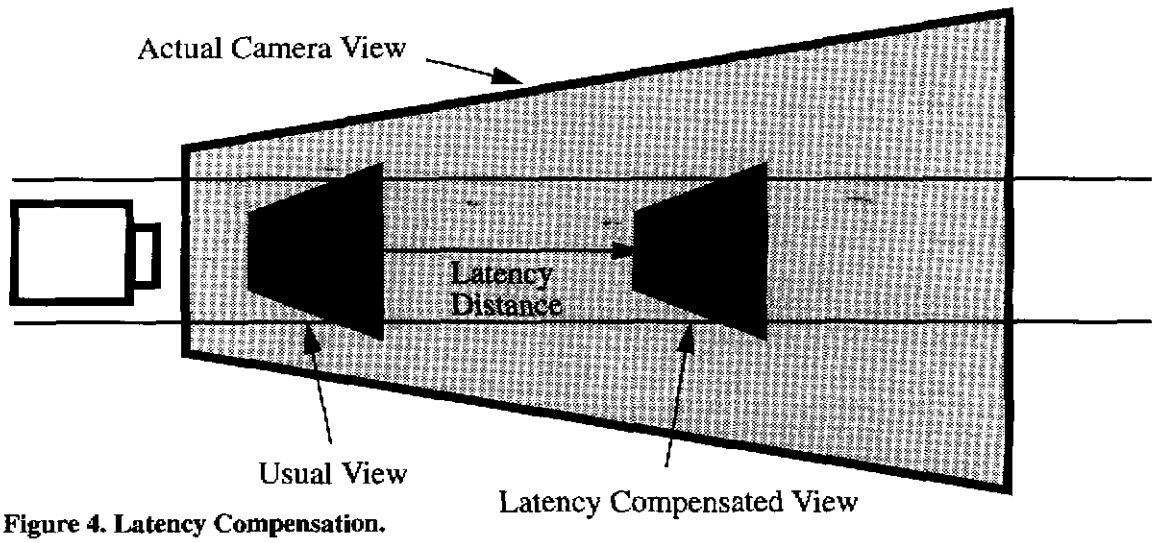

Figure 4. Latency Compensation.

8. Improving Reliability Estimation

A major factor that has thus far prevented wholehearted acceptance of ALVINN, as well as other neural network based applications, is the inability to precisely understand what the network is learning and how it will react to new situations. In model based road following systems, it is easy to understand that when a particular feature is not present in the road image, the system will not function properly. The system designer accepts this because he created the system to look for that feature - thus he knows why and when it will fail. In ALVINN, it is much more difficult to determine when the system will fail and almost impossible to tell precisely why it fails. This is because it learns different features for different road types and because it usually encodes what it learns in a way that is hard to accurately interpret. These factors make analyzing neural network systems especially hard.

One method that attempts to provide a quantitative measure of the accuracy and reliability of ALVINN is called Input Reconstruction Reliability Estimation (IRRE). In the IRRE method, the network is trained to not only produce the desired steering direction but also to reproduce the input image at its output. By comparing this reconstructed image with the actual image presented to the network it is possible to compute a confidence measure. The reasoning behind this method is straightforward: If the network can accurately reproduce the input image, it has most likely been trained on similar images and it is reasonable to assume that the output steering direction is also accurate. Although this technique works well and will form a basis for the techniques described later, it can only predict what the vehicle will do at this instant and assumes that a single, accurate reconstruction can always be related to a correct action. There is no way to determine the future consequences of the current action. Because the confidence value is computed

from just one image, there is a chance that the network will accurately reconstruct a random image, leading to belief in the output road position - and a potentially catastrophic situation.

A Virtual Active Vision method called Consequence Based Reliability Estimation (CBRE) provides a way to make quantitative measurements of network confidence and reconcile the vehicle's current actions with their future consequences. Virtual Active Vision techniques also provide a way to merge the meaning of output activations from many virtual views and to create a spatially relevant measure of network reliability which relies on the consistency of the IRRE measure of the network when presented with images from known, multiple virtual cameras. (This type of reliability estimation is known as Output Consistency Reliability Estimation or OCRE.) These methods, both grounded by action in the physical world, are discussed in greater detail in the next two sections.

8.1 Consequence Based Reliability Estimation

Consequence Based Reliability Estimation is an iterative method in which the action resulting from ALVINN's output is simulated and used to develop a confidence measure. The method is a four step process which involves:

- Imaging the scene at the current vehicle position.
- Producing an output and IRRE measure based on this scene.
- Compounding this IRRE measure with any previous ones.
- Updating the current vehicle position based on the network output.

This technique uses ALVINN's output to predict where the vehicle will be at some point in the future and moves the virtual camera view to that location, computing a confidence measure along the way.

The first step in CBRE is to place a virtual camera view immediately in front of the vehicle. The image produced by this view will be passed through ALVINN's neural network and will create some output activation. Because this output actually specifies a point on the ground plane some distance in front of the vehicle, it is possible to simulate moving the vehicle to this location and to re-image the same scene with another virtual camera view. This 'image and move' process is repeated for several iterations. See Figure 5.

The intuitive reasoning behind CBRE is that if the network produces an accurate output, not only will IRRE confidence be high, but also the new position computed using this output will be on the road. When a scene is imaged from this new location, and if the previous output did actually place the vehicle on the road, IRRE confidence will again be high. If the output was inaccurate, the imaged scene will not be on the road and IRRE confidence will be low. Combining a series of these measurements will yield a confidence measure that is grounded in vehicle action.

Used alone, IRRE suffers from the drawbacks mentioned earlier, but by incorporating several measures, each derived from an action predicted by the trained network, better reliability estimations can potentially be developed. Figure 6 and Figure 7 shows how the virtual images might look for two different road types.

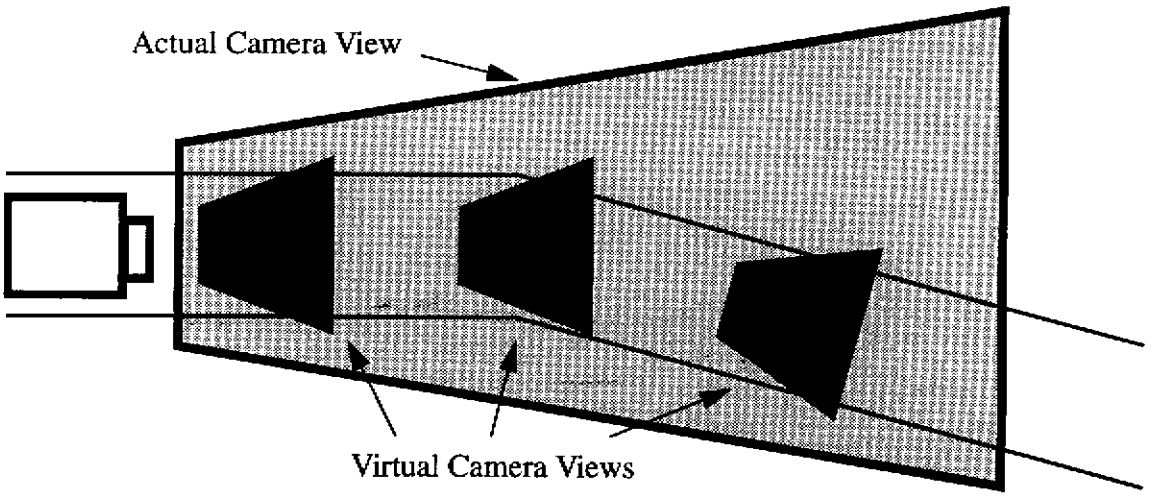

Figure 5. Consequence Based Reliability Estimation.

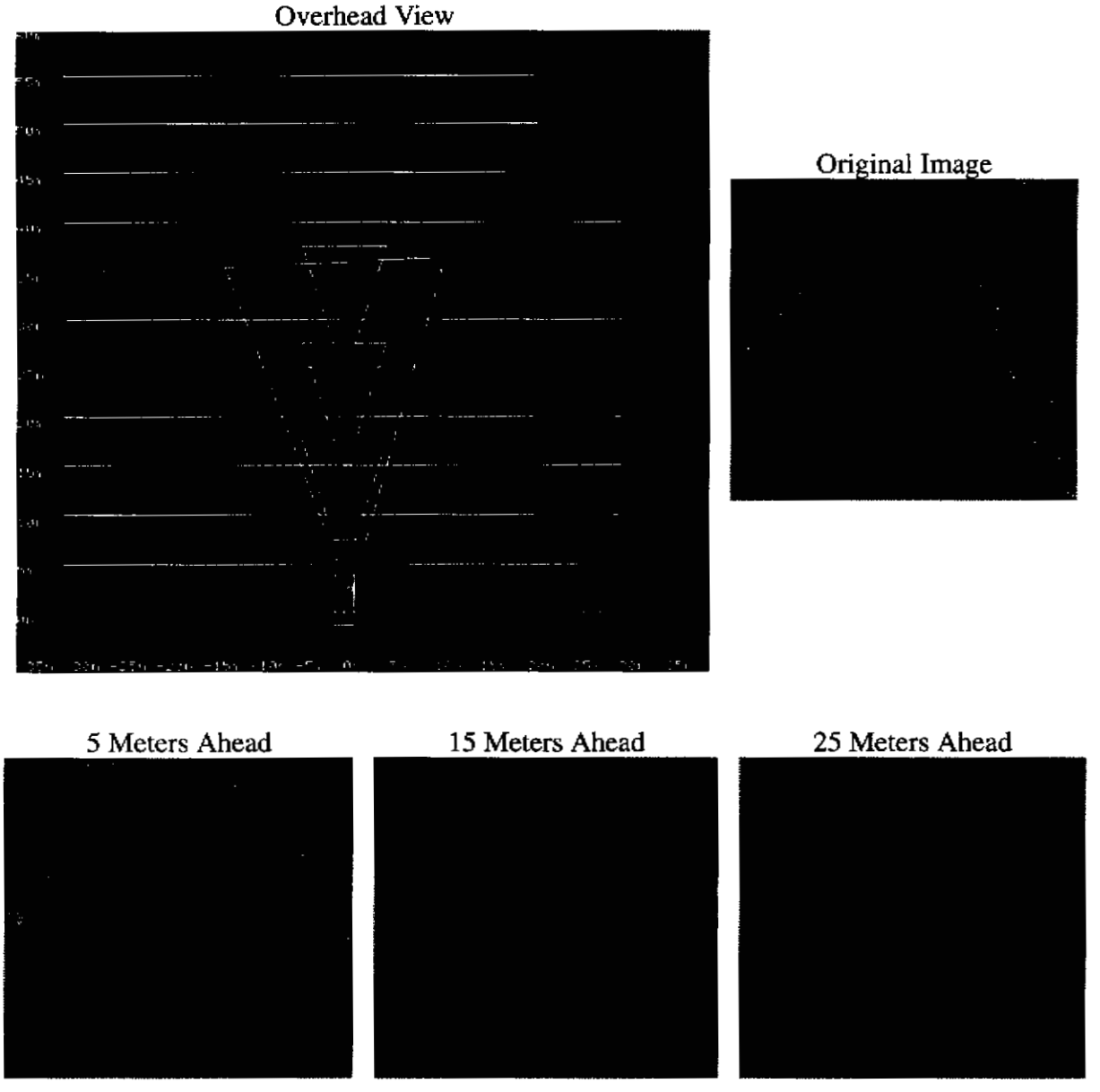

Figure 6. Simulated CBRE images on an unlined, straight, paved path.

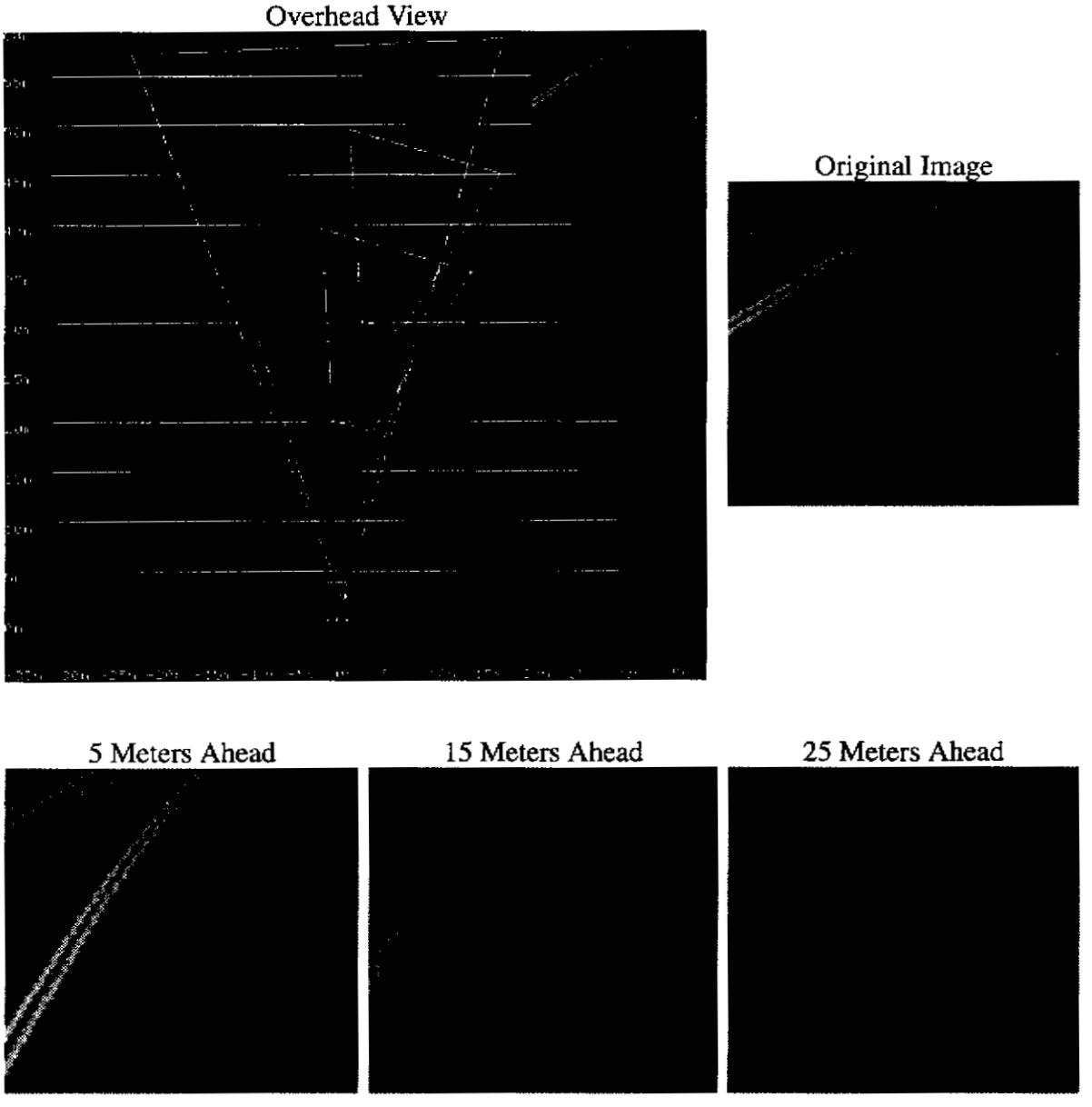

Figure 7. Simulated CBRE images on a lined, curving, city street.

8.2 Output Consistency Reliability Estimation

Output Consistency Reliability Estimation is based on the agreement of the output of ALVINN's network across several input images. The specific output produced by each image is not and should not be identical, but rather should specify the same point in front of the vehicle. If several virtual cameras are created at the same distance in front of the vehicle, but at different offsets from the vehicle's center line as shown in Figure 8, all of their outputs should specify the same point. The degree to which they do so can be thought of as a measure of the reliability of the system. Also, because multiple driving point are available, they can be combined into one point using their respective IRRE measure as a kind of weighting factor.

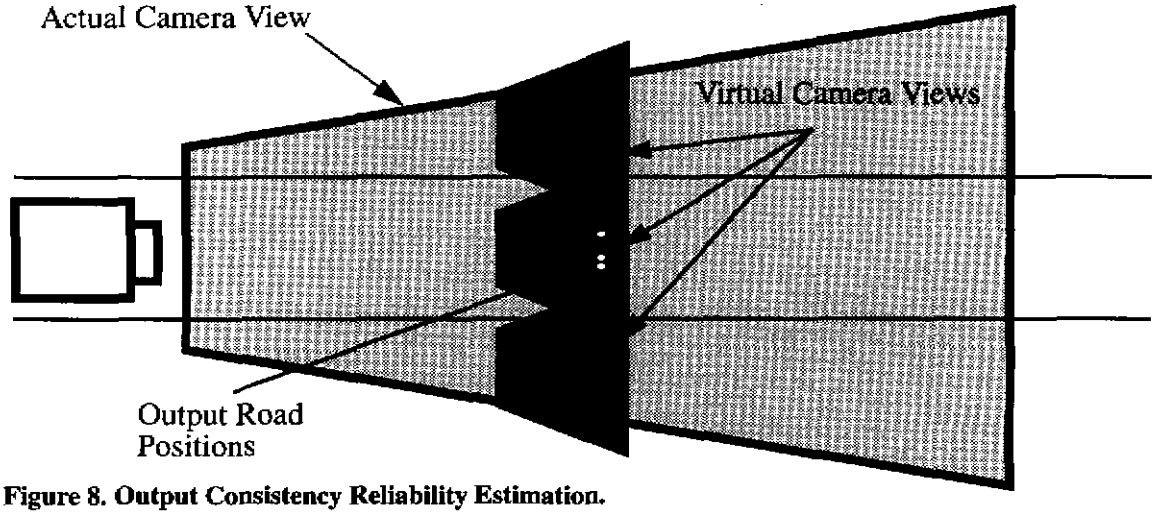

Figure 8. Output Consistency Reliability Estimation.

The basic OCRE method could work as follows. Suppose the network is given images taken at each of the virtual camera locations. For each of these images, it produces a road position along with an IRRE measure. The IRRE measure acts as weighting factors for combining this output displacement with others to create a new output. Intuitively, this means that virtual views which produce a high IRRE measure are weighted more than ones which produce a low measure.

A potential way to obtain a confidence measure of this new output is to examine the sum of the IRRE measures. A large sum value would indicate high confidence while a small value would indicate low confidence. A method such as this has the advantage that if one virtual camera views caused the network to produce an incorrectly high IRRE measure, the other views will not likely suffer the same error and low confidence will be correctly reflected in the final, combined confidence measure. Figure 9 shows how virtual images used in OCRE might look.

Another interesting offshoot of this idea is not only using different virtual camera views but also using different networks, trained from different views, to produce a road position. Using networks trained with images from different camera poses, perhaps some focusing entirely on particular features in the image like the center line or the road edge, a more accurate road position could be derived. An approach such as this, in which the location of important features is determined a priori, has a definite model based flavor, but it is an interesting way to incorporate some of the strengths of the model based approaches into the neural paradigm.

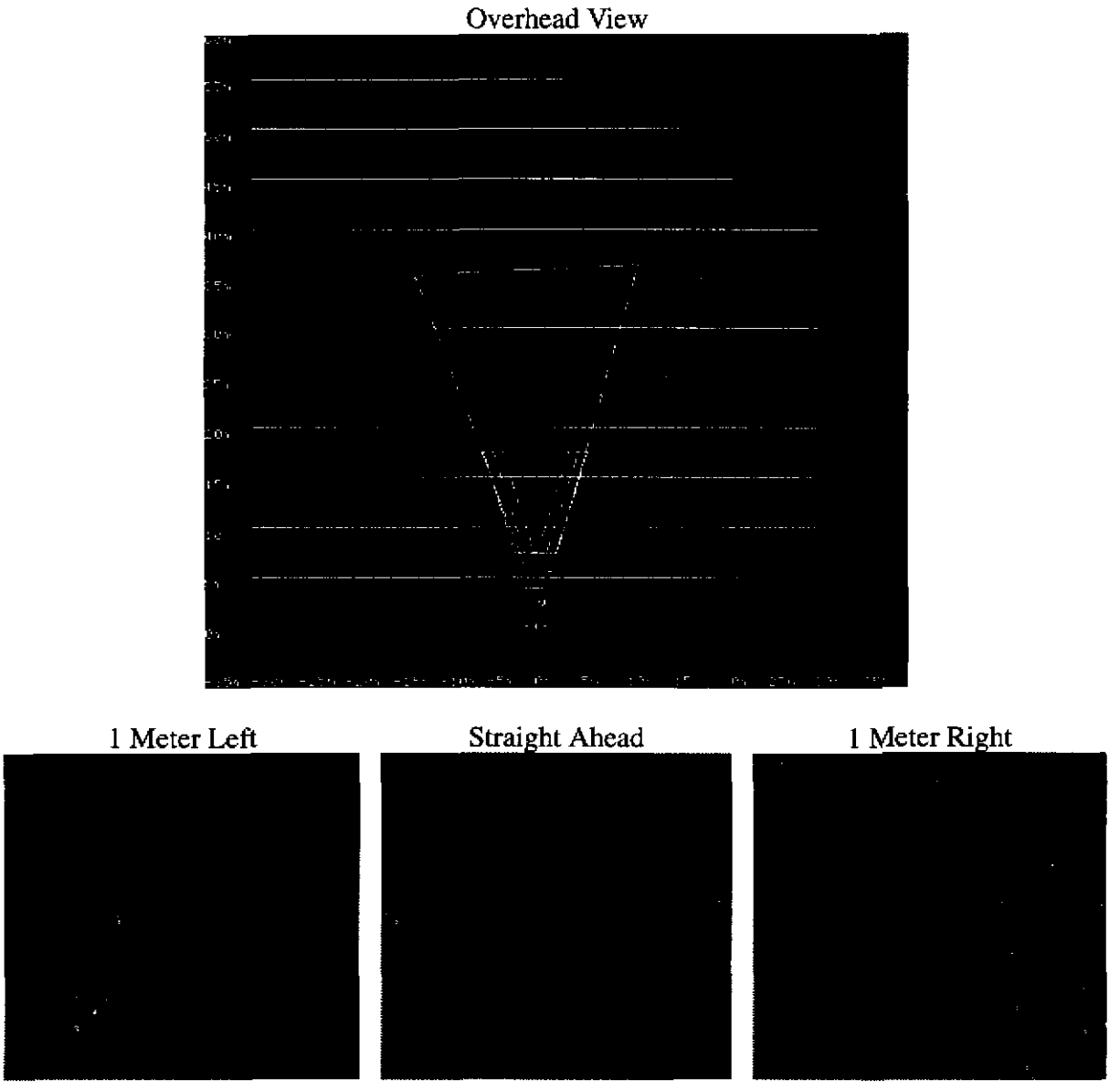

Figure 9. Simulated ORCE images on a straight road.

Finally, the OCRE method is especially well suited for the task of merging the output of networks which use different sensing modalities into a single road position. The combination of infrared and video data for day and night driving is one such example.

9. Merging Neural and Symbolic Data

Virtual cameras provide a way to link ALVINN with other, mostly symbolic, autonomous navigation systems that provide information such as local planning and global positioning. The information contained in these other systems was previously difficult to use because there was no real common ground between them and ALVINN. Instructions were issued and ALVINN was assumed to have responded properly. Feedback was almost non-existent. Below are three examples of real situations which require high level knowledge. A potential solution to each using virtual cameras is outlined.

9.1 Exit Ramp Detection

Suppose a mapping and path planning module tells ALVINN that in order to reach some final goal point, an exit ramp off the interstate needs to be taken. This module could supply ALVINN with general information about where the exit ramp is located along with other relevant information that could aid in its detection. ALVINN might then use this information to create virtual camera views that are positioned in such that they will image areas to the right of the interstate (where exit ramps usually occur.) At the same time, another virtual camera could keep the vehicle on the road until the exit ramp is found. See Figure 10. With an approach such as this, ALVINN can

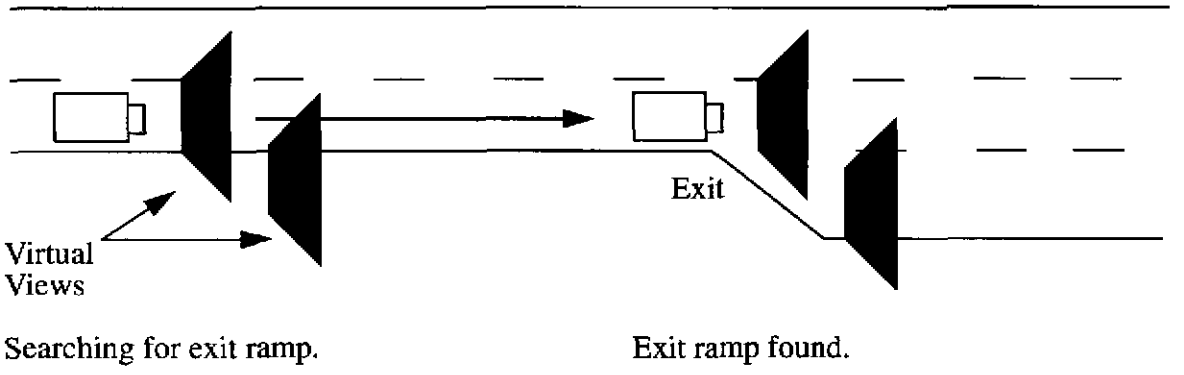

Figure 10. Detecting exit ramps.

simultaneously follow the road as well as carry out goal directed behavior requested by high level systems. Once the exit ramp has been found, possibly by applying the reliability estimation techniques mentioned earlier, its location can be refined in the high level module's map. Figure 11 shows how this scenario might look in on a real highway.

9.2 Intersection Detection and Traversal

Another example where virtual cameras can be used to merge high level, symbolic data, with ALVINN is that of detecting and moving through city intersections. Suppose a high level mapping module tells ALVINN to "turn left at the next four way intersection." Ideally, you would like ALVINN to take this instruction, find and go through the intersection, and inform the high level module that the request had been successfully completed. By placing virtual cameras in the manner shown in Figure 12, this task can be completed. Used in this manner, virtual cameras can provide a way to navigate through intersections without having to rely solely on blind, dead

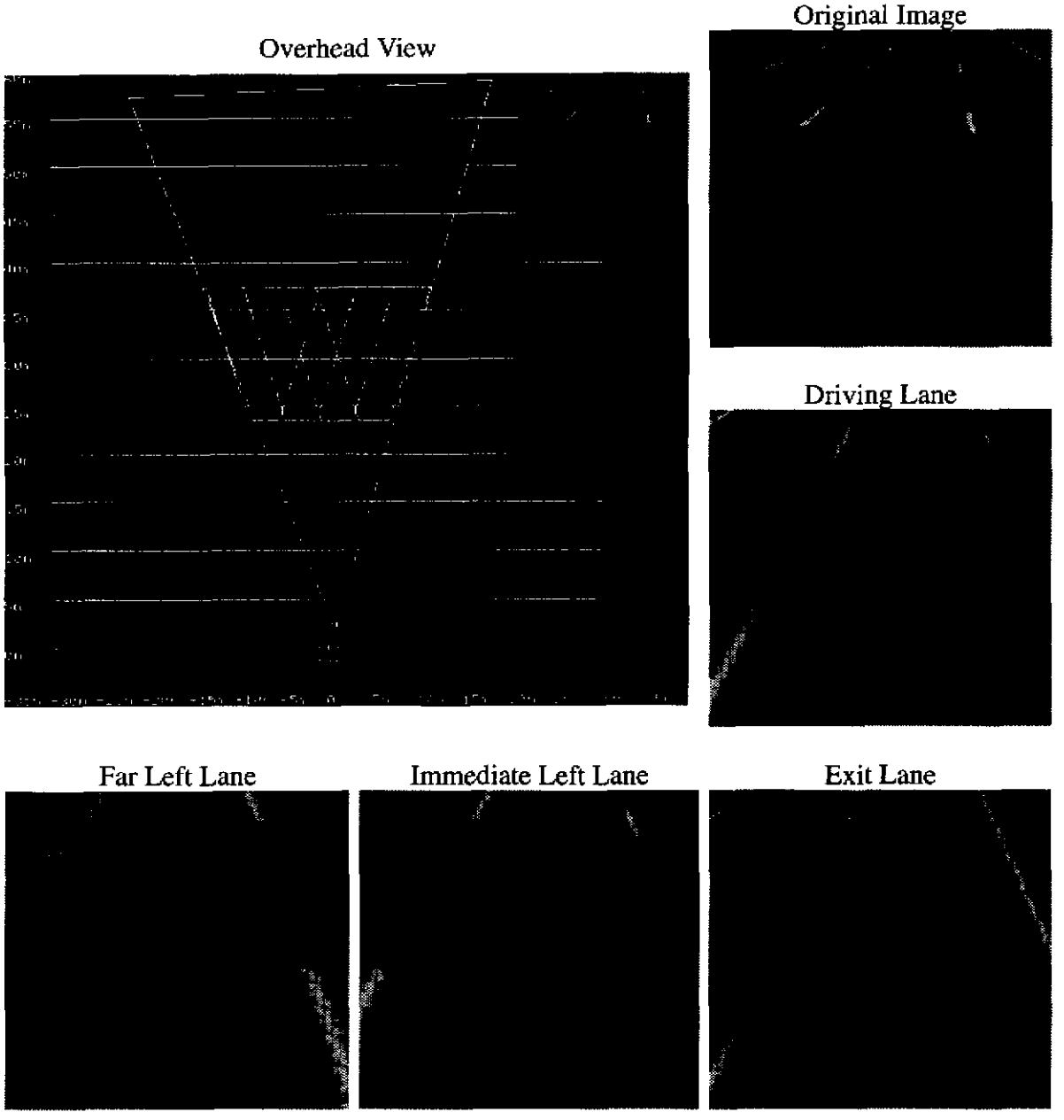

Figure 11. Offramp and lane detection using virtual

reckoning techniques. Because it is unlikely that a single camera will be able to sufficiently image the entire intersection, the need for integrating more traditional active vision techniques is highlighted. Some of these techniques, such as panning the actual camera and combining multiple actual camera views, are discussed in Section 10.

9.3 Plan Improvement Through Observation

The reuse of information acquired in previous runs is a problem which has typically been hard to address when using reactive systems such as ALVINN. Although high level modules usually contain such information, merging it with the current output of low level systems has typically been

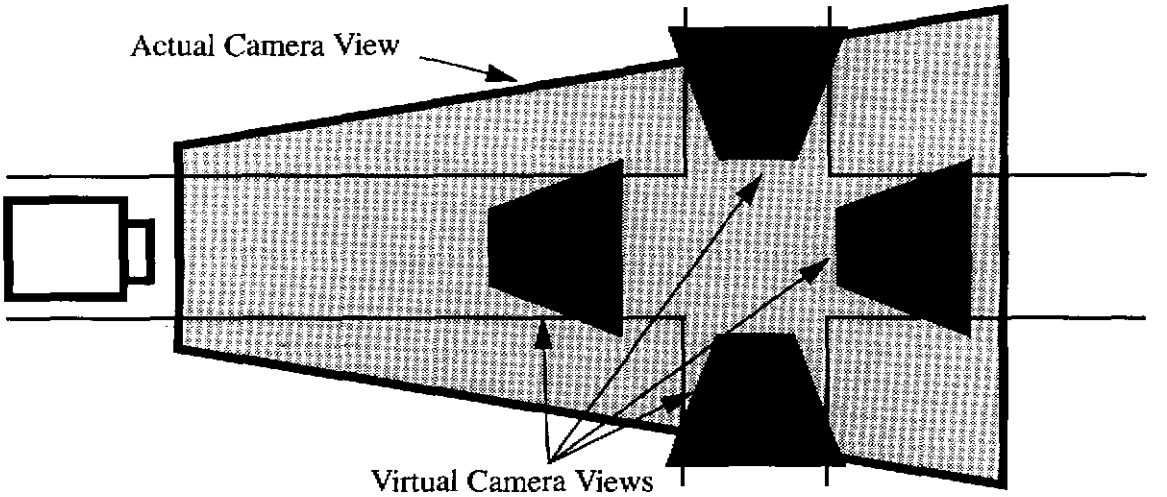

Figure 12. Intersection navigation.

difficult because of issues related to data representation. Since ALVINN will now produce points rather than arcs, new areas of information reuse can be explored. An initial area that can be explored is that of merging locational information about where the vehicle drove on previous trips on the current road with new, sensor derived data. Since it is feasible to assume that ALVINN can supply a high level module with every point over which it expects to drive, knowledge can be readily stored for reuse. In later trips over the same road, this information can be recalled and merged with the current road estimation. Over several trips, an accurate representation of the correct path can be constructed.

10. Active Vision and Sensor Fusion

Traditional Active Vision and Sensor Fusion are two areas which can also be explored using Virtual Active Vision techniques. Several potential items that relate to these areas which would benefit an autonomous road following system are discussed below. It is not yet clear which of these items will be most useful for the problems that I propose to explore, but dealing with at least the cursory issues of all the topics will likely be required.

- **Optimal sensor placement for varying navigation tasks.** Because pixel values in a virtual camera's field of view, which are not in the actual camera's, must be inferred, it is desirable to maximize the overlap of virtual and active camera fields of view. In the exit ramp example, the view required for finding the ramp could very well be out of the actual camera's field of view if the actual camera was oriented for road following. It would be better to have the actual camera pose be one which overlapped maximally with the road following and exit ramp virtual views.
- **Using more than one camera to produce virtual views.** Another solution to the maximal overlap problem mentioned above is to use more than one actual camera whose views overlap. Pixels in a virtual view which fall outside of the field of view of one actual camera may fall inside the field of view of the other. A method such as this would be desirable when movable cameras are not available and when it is necessary to image a large area quickly, such as in the case of intersection detection.

- **Using 3D elevation instead of the flat world assumption to map image data.** By using real elevation data instead of the flat world model to map image data to the ground plane, a more accurate world model can be produced. However, it is not yet clear whether this is a better or necessary model for use with an ALVINN based autonomous driving system.
- **Sensor Registration.** As new sensing modalities become available, the ability to correctly register them with each other becomes more important. Virtual cameras provide a way to image different data, like color video and forward looking infra-red (FLIR), using identical image formation models. This ability, coupled with the new sensing modalities like FLIR, provides intriguing opportunities in areas such as day-night driving.

11. Contributions

The goal of this work is to develop methods that will allow reliable and robust performance on complex, real world driving tasks. Such tasks inherently require information from high level, goal directed modules as well as stable performance from low level, reactive control systems. These tasks provide truly difficult requirements which must be sufficiently satisfied for acceptable results.

To accomplish this goal, significant progress must be made in evaluating ALVINN's performance in terms of both safety and with respect to landmark detection. (i.e. Finding offramps, intersections, and lane changes.) And to accomplish any of this, new ways to bidirectionally link ALVINN to higher level modules must be found. It is in these areas that important contributions will be made.

This work will provide significant advances in both the neural network control and the autonomous mobile robot communities.The research will explore and provide methods for estimating the reliability of artificial neural networks using extrinsic metrics (CBRE and OCRE) rather than the traditional, intrinsic metrics (IRRE). Knowledge in this area will perhaps lead to more widespread acceptance of the neural paradigm for control problems and perhaps provide a partial answer to the question, "Yes, but how do you KNOW it works?"

Additionally, it will provide a systematic method for merging high level knowledge sources with low level, reactive modules. This merger will be bidirectional, with ALVINN providing feedback to high level modules about how it is currently performing and how it expects to perform in the near future. Information contained in high level modules that is required for goal directed behavior will become easily accessible to ALVINN. Because of this, autonomous operation in new domains will become feasible and overall system performance and reliability will significantly improve.

12. Time Table

Table 1. Proposed timetable to completion.

Start	End	Research Agenda
7/1/93	9/30/93	Develop and test basic virtual camera code.
10/1/93	10/31/93	Write thesis proposal.
11/1/93	2/28/94	Best camera view, latency compensation, CBRE and OCRE.
3/1//94	3/31/94	Document results.
4/1/94	7/31/94	Merge with high level mapping module.
8/1/94	8/31/94	Document results.
9/1/94	1/31/95	Active vision and sensor fusion experiments.
2/1/94	4/30/95	Complete thesis document and defend.

13. Bibliography

[1] Brumitt, B., Kelly, A., Stentz, A. "Dynamic Trajectory Planning for a Cross-Country Navigator," *SPIE Mobile Robots VII,* Boston, MA, 1992.

[2] Dickmanns, E. and Mysliwetz, B. "Recursive 3-D Road and Relative Ego-State Recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, pp. 199-213, May 1992.

[3] Jochem, T. Pomerleau, D., Thorpe, C. "MANIAC: A Next Generation Neurally Based Autonomous Road Follower," *Intelligent Autonomous Systems-3,* February 1993, Pittsburgh, PA, USA.

[4] Kelly, A and Stentz, A. "RANGER, An Intelligent Predictive Controller for Unmanned Ground Vehicles," CMU Robotics Institute Technical Report. In progress.

[5] Kluge, K. and Thorpe, C. "Representation and Recovery of Road Geometry in YARF." *Intelligent Vehicles '92 Symposium*, June, 1992.

[6] Langer, D. and Thorpe, C. "Sonar Based Outdoor Vehicle Navigation and Collision Avoidance," *IROS '92,* 1992.

[7] Lotufo, R., Dagless, E., Milford, D., and Thomas, B., "Road Edge Extraction Using a Plan-view Image Transformation," *4th Alvey Vision Conference.*

[8] Meng, M. and Kak, A., "Mobile Robot Navigation Using Neural Networks and Nonmetrical Environmental Models," *IEEE Control Systems,* October 1993, pp. 30-39.

[9] Payton, D., Rosenblatt, J. and Keirsey, D., "Plan Guided Reaction," *IEEE Transactions on Systems Man and Cybernetics,* 20(6), 1990, pp. 1370-1382.

[10] Pomerleau, D., "Neural Network Perception for Mobile Robot Guidance," Kluwer Academic Publishers, 1993.

[11] Pomerleau, D., "Progress in Neural Network-based Vision for Autonomous Robot Driving," *Intelligent Vehicles '92 Symposium,* June, 1992.

[12] Pomerleau, D., Gowdy, J., and Thorpe, C. "Combining Artificial Neural Networks and Symbolic Processing for Autonomous Robot Guidance," *Engineering Applications of Artificial Intelligence,* Vol. 4, No. 4, pp. 279-285, 1991.

[13] Turk, M., Morgenthaler, D., Gremban, K., and Marra, M. "VITS - A Vision System for Autonomous Land Vehicle Navigation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 3, May 1988.

[14] Wallace, R., Matsuzaki, K., Goto, Y., Crisman, J., Webb, J., and Kanade, T., "Progress in Robot Road Following," *1986 IEEE Conference on Robotics and Automation,* pp. 1615-1621.