Chapter 5 Other Systems

5.1 Introduction

The Virtual Active Vision tools which I have developed and used with the ALVINN lane keeping system are designed to be independent of the lane keeper, only requiring that the system use images to drive. Because of this, other researchers have been able to directly use these tools, as well as the concepts that they embody, to develop and enhance two other vision based lane keeping systems. The first system, called ROBIN, uses virtual cameras to achieve better lane keeping performance than it was previously capable of attaining. ROBIN uses virtual cameras to create images which have a more appropriate feature distribution than those taken directly from the video camera. Unlike ROBIN, which used the same implementation of virtual cameras that ALVINN uses, the RALPH driving system uses adaptations of three Virtual Active Vision techniques to achieve state-of-the-art lane keeping performance. A more detailed description of each system and the role that Virtual Active Vision tools play in increasing their performance follows.

5.2 ROBIN

ROBIN is a neural network autonomous lane keeping system first developed at the University of Maryland and then extended at Lockheed Martin Corporation (LMC) in Denver, Colorado. ROBIN is based on the radial basis function network architecture and learns how to drive by capturing prototypical driving scenes for a discrete set of steering directions during a training phase.

In autonomous driving mode ROBIN generates a degree-of-match between the input image and each prototypical driving scene captured during training. It then uses the steering directions associated with the prototypical scenes, weighted by the degree of match between the prototypical scene and the current scene, to control the vehicle. Figure 62 shows a diagram of the ROBIN network architecture.

In order to differentiate between different steering situations, there must be a detectable difference between the corresponding driving scenes for those situations. The more detectable the scene difference, the more reliable ROBIN's determination of the appropriate steering direction will be when driving in autonomous mode. As would be expected, features in the road scene provide the key cues for scene differentiation. ROBIN primarily exploits the contrast between the on-road and off-road portions of the image, since these are typically the most prominent features in road scenes. Other features that are evident in typical driving scenes are road lines and oil spots, but due to reduction in resolution from the camera image (640x480) to the low resolution input image (32x30), they are not used very well by ROBIN [44].

The maximum variation of road scenes occurs when the number of pixels representing off-road features approximately equals the number of pixels representing the on-road features, leading to an off-road to on-road pixel ratio of one. On LMC's testbed vehicles, the camera is mounted on a pan/tilt platform so that it can be mechanically moved to optimize the off-road to on-road pixel ratio for a particular road-scene. The optimization of the off-road to on-road pixel ratio is limited, however, to the mechanical range of the camera platform, and often the desired pixel ratio of one cannot be achieved. In order to better maintain a ratio of one, virtual cameras have been added to the ROBIN system. Because virtual cameras can be positioned anywhere in space, it is possible to find a location where the off-road to on-road pixel ratio is approximately one.

Before virtual cameras were added to ROBIN, its ability to control LMC's testbed vehicle at moderate speeds was unsatisfactory. When driving down a straight stretch of road, the vehicle, a converted U.S. Army HMMWV, would exhibit oscillatory behavior, gradually swerving from road edge to road edge. This behavior was due to the pixel ratio being only about 1 off-road pixel for every 8 on-road pixels. In order to equalize the number of off-road and on-road pixels, a vir-

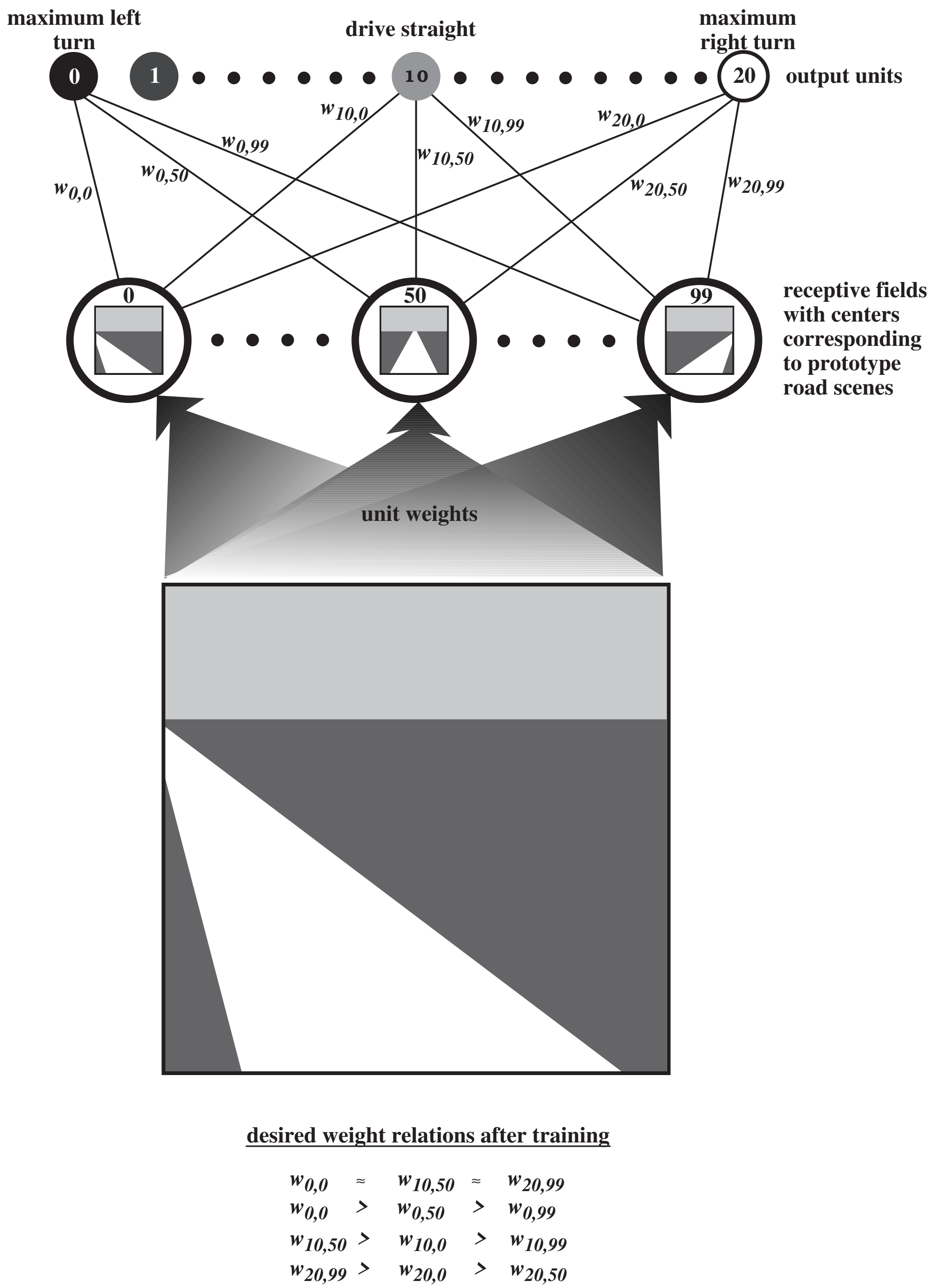

Figure 62 ROBIN network architecture.

tual camera was used to resample the input image from a location that was more appropriate. (The actual camera was mounted on the right side of the vehicle, about a foot above the hood. From this viewpoint, the vertical field of view of the lens caused the camera to image above the horizon.) When tested in this new configuration, ROBIN's oscillatory driving behavior was eliminated and the system was be able to drive LMC's testbed vehicle on four lane divided highways at speeds up to 45 m.p.h.[45]

5.3 RALPH

RALPH (Rapidly Adapting Lateral Position Handler) is a vision based, adaptive lane keeping system that has driven our testbed vehicle over 7000 autonomous miles in the last 9 months. This includes a 2850 mile transcontinental trip from Washington D.C. to San Diego, CA, during which RALPH autonomously steered the Navlab 5 98.2% of the way. RALPH decomposes the problem of steering the vehicle into three steps, sampling the image, determining the road curvature, and determining the lateral offset of the vehicle relative to the lane center. The results of the last two steps are combined into a steering command which can either be used to control the steering wheel of our testbed vehicle or provide an estimate of the proper direction to drive, if the system is operating in a warning capacity [42].

The RALPH algorithm is designed to look for parallel, linear features on the road. These features can be solid or dashed lines, the road/shoulder boundary, tire tracks left by a preceding vehicle, or even the oil spot down the center of the lane. When imaged from a camera mounted behind the windshield, these features are not parallel, but rather converge because of perspective distortion. Because the algorithm expects the features to be parallel, the image must be resampled. The resampling is done so that the portion of the road within 1 meter on either side of the lane and extending from about 20 to 70 meters in front of the vehicles is extracted. This area is enclosed by the light colored "imaging trapezoid" shown in Figure 63 and corresponds to the area which would be imaged by a virtual camera located about 45 meters in front of the vehicle which was looking straight down at the roadway. The pixels within the imaging trapezoid are selectively sampled so as to create a low resolution (30x32 pixel) image in which important features such as lane markings, which converged towards to top of the original image, appear parallel. Although the implementation of this functionality is different than those used for the experiments described in the previous chapters, the effect is the same - the image processing task is made easier through the use of Virtual Active Vision techniques and more robust performance is possible.

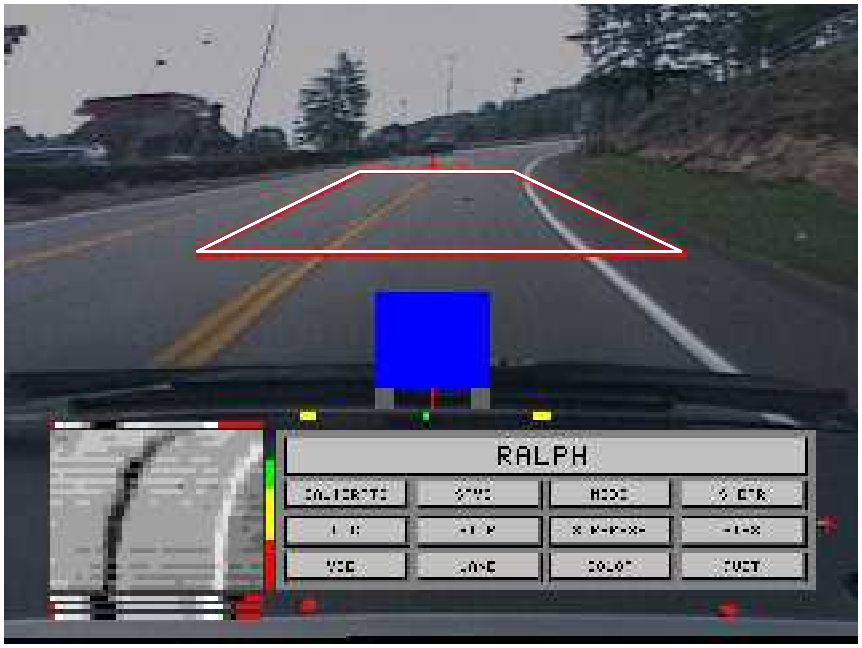

Figure 63 Ralph display.

Another instance where the RALPH system uses concepts much like those developed in this thesis is for focusing attention on the areas of the scene which contain the most important information for accomplishing the task at hand. One example of this is adjusting the location of the imaging trapezoid based on the current vehicle speed. As the vehicle speed increases, RALPH moves the imaging trapezoid further out, away from the vehicle, so that the upcoming road curvature can be determined in time for the system to react to it. This movement of the imaging trapezoid serves the same function as the virtual camera views used for lane transition and intersection detection - it allows the underlying lane keeping system to be used without modification, to accomplish a task that could typically require fundamental changes in the lane keeping algorithm. A second example of moving the imaging trapezoid to increase driving performance occurs when the system detects that the horizon, or other irrelevant feature, is in the distal portion of the imaged area. Because these features do not correspond to those in the rest of the imaged area, driving performance can suffer. To alleviate this problem, the imaging trapezoid is moved closer to the vehicle, eliminating the spurious feature from the imaged area.

Finally, the RALPH system uses a similar concept to the lane transition methods described in Chapter 3 to transition from one lane to another. It differs in that RALPH is able to offset the vehicle from the center of the driving lane without laterally moving the imaging trapezoid. This is possible because unlike ALVINN, RALPH separately determines the vehicle's lateral displacement from the center line of the road and the upcoming road curvature. Using this knowledge RALPH is able to move the vehicle toward the centerline of the road by incrementing the target lateral offsets. Once it is at the centerline, it shifts its hypothesis of where the road features are located in the opposite direction of the lane transition. This action produces the same effect that lateral movement of virtual camera view has when used with ALVINN. In RALPH's case, when the system matches the newly shifted features to the road scene, the system will be able to finish the transition into the destination lane.

5.4 Discussion

Like ALVINN, ROBIN and RALPH are vision based lane keeping systems. Their operational domain is one in which geometric relationships are very important and useful. Each system implements and uses Virtual Active Vision tools in different ways. In the ROBIN system, the implementation of virtual cameras described in this dissertation is used only to increase the performance of an existing lane keeping system to an acceptable level. In contrast, a virtual camera-like imaging trapezoid is at the core of the RALPH system. All subsequent processing depends on the transformation that this view applies to the image. Without it, all later steps are invalid. In addition to being very dependent on a virtual camera-like sensor, RALPH uses many active view control techniques to improve lane keeping robustness. It does this by adjusting the viewing location so that it images only road features. In effect, RALPH is doing on-the-fly optimal camera placement. The ALVINN system uses virtual cameras in ways both similar to and different from ROBIN and RALPH. Like ROBIN, virtual cameras have been added to an existing lane keeping system. But they are not used to simply increase lane keeping performance. In ALVINN, active view control techniques are used to build tactical driving capabilities. But in RALPH, similar techniques are used to increase the robustness of the base lane keeping system.

By examining how ROBIN, RALPH, and ALVINN implement and use virtual cameras, domains where virtual cameras are likely to be useful can be identified and the typical progression of deployment can be better understood. Domain areas which include tasks where simple geometric transformations of the viewpoint can substantially simplify the processing are ones which are especially amenable to the application of Virtual Active Vision tools. Initially these tools will be added onto existing systems to help improve performance. For example, take a simple vision based parts identifier. The system developer could implement a system to detect parts only when they were located in a certain position and orientation on a conveyer belt. By supplementing this basic system with many virtual cameras, it would be able to identify parts which were in any position and orientation on the belt. This type of use is similar to ROBIN. The next stage of deployment of virtual cameras is represented by ALVINN and RALPH. In both of these system, virtual cameras do more than simply improve rudimentary performance. They are the foundation on which new capabilities are built. Their scope of the capabilities differ, with RALPH focussing on novel view placement procedures to eliminate spurious driving features, while ALVINN uses virtual cameras to add tactical driving capabilities. An example of this level of deployment using the parts identifier case is tracking parts and triggering some action based on where the parts were located on the belt. For example, a part could be tracked until a character recognition program was able to read its serial number using a virtual camera stabilized image.

Another example of a real vision system which is in the first stage of deployment is the face detection system being developed at Carnegie Mellon [47]. This system is designed to find faces in an input scene regardless of their location. In this system a scale dependent neural network is trained to discriminate between images which contain faces and those that do not. During testing, portions of the image are resized so that if they contain a face, it will be at the correct scale for the neural network to accurately predict it is presence. This functionality maps directly to the parts identifier example - a simple system is enhanced using a virtual camera-like tool.

It would be easy to say that all vision tasks could benefit from Virtual Active Vision tools, but at some point the exact nature of these tools must be characterized. The dividing line between Virtual Active Vision tools and simple image subsampling lies in the ability of true virtual cameras to simplify the processing of underlying systems by representing the world in a common, consistent manner. In addition to simplifying processing, true Virtual Active Vision tools enable basic systems to exhibit more advanced functionality through the use of active view control. As a result of controlling the areas in the scene which are imaged, advanced capabilities can be achieved. The

Virtual Active Vision tools present in the ALVINN and RALPH system clearly fit this definition but it is questionable whether the ROBIN and face detection systems really use Virtual Active Vision tools, or just advanced subsampling techniques. Until active view control is added to these systems, the true power of Virtual Active Vision tools has not been exploited.