Proc. 2001 Australian Conference on Robotics and Automation
Sydney, 14-15 November 2001

Computer Vision for Vehicle Monitoring and Control

Luke Fletcher, Nicholas Apostoloff, Jason Chen, Alexander Zelinsky
Robotic Systems Laboratory
Department of Systems Engineering, RSISE
The Australian National University
Canberra, ACT 0200 Australia
[luke|nema|chen|alex]@syseng.anu.edu.au

Abstract

A platform for intelligent transport systems and autonomous vehicle research is in development at the Australian National University. The vehicle's first application will be a test bed for on-road driver and vehicle monitoring as well as computer vision based autonomous steering and velocity control.

1 Introduction

Many Australians spend large amount of time behind the wheel. Unlike other accident risk factors such as drink driving and speeding, there is no clear metric to detect fatigued drivers. The effects of fatigue are a judgement call on the part of the driver, and ironically it impairs that judgement. The seriousness of the problem was highlighted recently in a House of Representatives Standing Committee on Communications, Transport and the Arts inquiry into *Managing Fatigue in Transport*. The committee stated that a conservative estimate is that 20 to 30 percent of all car crashes can be linked to driver fatigue [HoRSCoCTA, 2000].

With increased computing power, Vision researchers are getting the courage to move outside of highly controlled laboratory and factory environments. The road environment offers an excellent compromise between the rigid constraints available in the laboratory and the complexity of the outside world. Roads are designed by humans to be: high contrast, predictable in layout and predominately free of out of context objects (any object unexpected can savely regarded as something to be avoided). Road vehicle control also has some significant niceties. Unlike many other real world robot deployments, there is a finite amount of a priori knowledge required by the robot, a lot of which has been explored and documented to educate human drivers. Also there is a significant body of literature avaliable regarding the dynamics of road vehicles.

Automation in road vehicles is being looked to as a possible tool to combat fatigue. Analogous to the deployment of industrial robots, automation offers the possibilities of continous attentiveness and exceptional endurance. The point has not been lost on vehicle manufacturers, most of which are now incorporating systems into their prototype cars and actively researching move advanced driver aids [Kato and Ninomiya, 2000] [Maurer, 2000].

Figure 1: The Autonomous Vehicle.

At the ANU's Autonomous Vehicle Project (AVP) preliminary research into autonomous systems for cars has begun. Previously in Australia the Safe-T-Cam project has been very successful at using computer vision to monitor the speed and trip times of trucks. And [Jarvis, 2000] are developing a sensor equipped vehicle for driver assistance. Internationally a large number of research groups are investigating a variety of topics. Arguably the most successful group has been the Universität der Bundeswehr München (UBM) where vehicles have been developed that can drive autonomously down freeways at speeds of over 140kph. The vehicles, predominately using computer vision, can complete complex tasks such as overtaking other vehicles[Dickmanns, 1999]. Carnegie Mellon University's Navlab project and its successors have also made significant contributions with the RALPH lane tracking system and trinocular obstacle detection [Batavia *et al.*, 1998][Williamson and Thorpe, 1999].

Initially the focus of the AVP is to explore computer vision as a means of identifying the state of the vehicle under the control of a human driver. Future experiments will introduce computer control of the vehicle. The research has two fronts: to explore the use of a computer vision (and other sensors) to control a vehicle and to explore how a human driver goes about controlling a vehicle. Vehicle monitoring can be thought of as the study of how people actually perceive the road environment while driving. Likewise vehicle control can be thought of as the study of how people act while behind the wheel. Although there are numerous differences between any possible computer based system and a human driving system, the project attempts to structure the road scene perception task in a way that can mimic the human system.

In the next section we give a hardware overview of the vehicle. Then we discuss the issues involved in computer vision for lane following. Finally we demonstrate an approach to moving obstacle detection.

2 Hardware Overview

The test-bed vehicle in our project is a 1999 Toyota Landcruier 4WD (Fig. 1). A 4WD vehicle was chosen for a number of reasons: it provides a strong and robust platform capable of surviving the rigors of experimentation; it has a large amount of interior space for installing sensors/computers; and it allows the option of performing research into off-road autonomous driving. As already stated, the project is ongoing, and so installation of hardware into the vehicle is not yet fully complete. In the following paragraphs we present a review of the hardware we have already installed and which we plan to install into the vehicle. For the purpose of this discussion, we note that hardware can be divided into three types: sensing, actuation, and communication/processing.

The main mode of **sensing** used in the vehicle will be vision. Two separate vision systems are planned. First, an active vision head (called CeDAR developed previously at the ANU - see [Truong *et al.*, 2000]) will be mounted with two stereo camera pairs. One pair will have a short focal length, and concentrate on the near field of view, while the other pair will have a longer focal length, and concentrate on looking further along the road. The second vision system involves using a stereo pair looking from the dash back toward the driver's face. By monitoring the driver useful information as to their intention can be gathered as well as verification that they have seen a detected dangerous situation. This system is based on the faceLAB system from [Seeing Machines, 2001]. Apart from vision sensing, a Global Positioning System (GPS), Inertial Navigation Sensor (INS), and laser range finder have been installed into the vehicle. The 6 DOF INS is mounted close to the vehicle's centre of gravity at a point between the two rear-seat foot-wells. It provides a continuous stream of linear and angular acceleration data that can be used to keep track of vehicle dynamics. The GPS provides data that can be used for high-level, navigation problems, but is also very useful for correcting drift in the INS output. The laser range finder has been mounted looking forward on the vehicle's bull-bar. Its purpose will be to identify obstacles, both stationary (eg. guard-rails, parked cars, etc.) and moving (eg. other vehicles), and will provide an additional source of information for our obstacle avoidance algorithms.

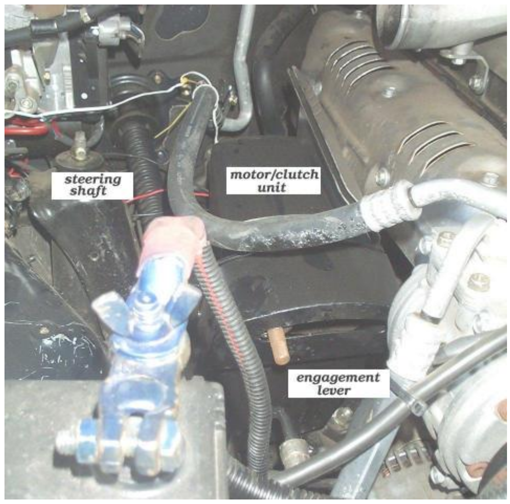

Figure 2: Steering mechanism of vehicle including drive motor/clutch unit (left), idler gear (centre) and gear on steering shaft (right).

Three **actuation** sub-systems are required in the vehicle: steering, braking, and throttle. We achieve throttle control by interfacing with the vehicle's cruise control module. The steering sub-system is based around a Raytheon rotary drive motor/clutch unit, which was designed for use in yacht auto-pilot applications. It was installed in the engine bay alongside the steering shaft of the vehicle. Power from an electric motor is transferred to the steering shaft using three spur gears: the first is attached to the steering shaft, the second to the motor shaft, and the third, being an idler gear, sits between the first two. A key feature in the design is that the idler gear can be engaged and disengaged from the drive-train using a lever protruding from the assembly. Then for "manual" driving of the vehicle, the idler gear can be disengaged, providing the safeguard that the autonomous steering assembly cannot impede normal steering in any way. A photo of the steering sub-system is shown in Figure 2. Note the lever used to engage and disengage the idler gear. Also note the rotary drive motor/clutch unit, and the vehicle's steering shaft.

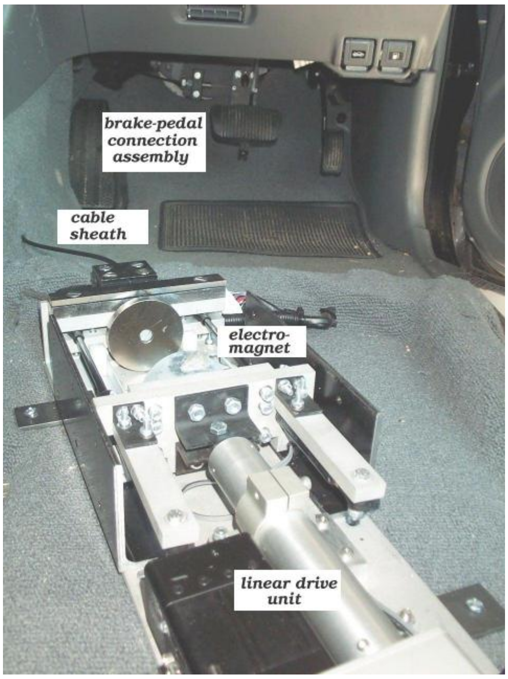

Figure 3: Braking device showing linear actuator & cable mechanism.

The braking sub-system is based around a linear drive unit (produced by *Animatics*), and an electromagnet. The linear drive is connected to one end of a braided steel cable via the electromagnet. The cable passes through a guiding sheath to reach, at its other end, the brake pedal. Braking is then achieved by having the linear drive unit pull on the cable. The electromagnet must be powered in order for braking to occur (ie. if it is unpowered, then the linear drive cannot pull on the cable to activate the brake). In an emergency, power can be cut to the electromagnet so that all braking control is returned back to the driver. In our implementation, an emergency scenario is communicated to the autonomous driving system by having the human activate an emergency stop button. The braking subsystem is shown in Figure 3. In the foreground the figure shows the linear drive and electromagnet, while in the background the brake pedal and its connection with the cable is shown.

Processing and communication hardware is required to fuse together the various sensing and actuation subsystems into a cohesive, single unit. Our approach in this area has been to favor the use of standard PC and networking hardware. Such hardware is readily available, easily upgradable, and cheap. Currently we have two PCs installed. One is a 1.4GHz Pentium IV, and is used for processing the video streams from the stereo pairs on the active head. The other is a dual 750MHz pentium III, and will be used for processing the video streams from the cameras focussed on the driver's face. An additional PC will be installed to process non-vision sensing data, and to control the throttle, steering, and braking subsystems. Communication is achieved between PCs via ethernet, with a connection from the vehicle back to a base station possible via a radio ethernet link. Due to the large number of sensing and actuation devices that communicate over serial lines, a serial port server has been installed. This device allows communication between a PC and serial devices as though these devices were connected directly to local serial ports on a PC. Finally, a SNAP I/O module (made by *opto22*) has been installed to provide a low level communication interface between PCs and various other devices (eg. cruise control system, steering motor control, steering angle potentiometer). This module connects into the ethernet, and provides a number of functionalities, including A/D and D/A conversion, PID control, timers, etc.

3 Vision Based Lane Tracking

Two main areas of thought exist on the construction of Intelligent Transport Systems (ITS): a new road infrastructure can be created to be an integral part of the ITS (i.e. through the use of lane marker beacons) or intelligent vehicles can be designed to work with the existing road infrastructure. This project focuses on making autonomous vehicles that can deal with existing road infrastructures. Two main issues support this decision: the obvious financial restrictions associated with the modification of existing road infrastructures; and the additional flexibility of designing a vehicle that can deal with a real world environment.

The task of *Lane Tracking* is to detect and track the boundaries and lane markings of a road, and is fundamental for autonomous vehicles. It is also useful for driver monitoring and assistance systems, such as Lane Departure Warning (LDW) and Adaptive Cruise Control (ACC) systems. The majority of lane tracking projects have focused on using one or at most two visual cues for the detection of lane boundaries. These systems often fall prey to robustness issues due to the inadequacies of the cues chosen. However, by combining a number of cues it is feasible to increase the robustness of the solution. We hope to exploit this idea in the lane tracking research outlined in this paper.

Preliminary results at the ANU using a simple feature extraction algorithm [Dickmanns and Zapp, 1987] have reinforced this idea. An overview of the algorithm is shown in Fig. 4. This algorithm is used to extract edges at specific angles, θ, to detect road boundaries and lane markings from an image.

A low-pass filtering of the search region is performed by condensing a pixel field oriented at the angle θ shown, into a single vector. This condensed vector is then searched for edge pixels via a ternary correlation. The algorithm is particularly useful for searching for edges at predicted angles, but can also be used for initialisation of lane tracking algorithms by searching over several angles. Sensitivity of the edge detector to

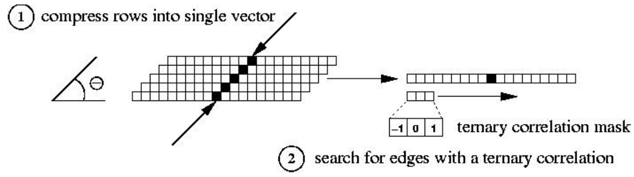

Figure 4: Edge detection algorithm by Dickmanns

curved edges and varying edge widths is controlled by the parameters of the ternary mask, and the number of rows chosen to be condensed together.

Figure 5 shows the edges found using this algorithm on a road image containing various factors that make road detection difficult (reflections in the windscreen, structures that run parallel to the road and shadows across the road). The black lines show the road and lane boundary edges that were detected. The white lines indicate invalid edges that can be easily removed via simple continuity constraints and road model fitting, while the dashed lines indicate edges that could be incorrectly detected as road or lane boundary edges.

This presents the question, "how does one decide which edges are part of the road boundary and which ones are not?". Further information or *a priori* knowledge must be used to filter out invalid information. For example, the system presented in [Suzuki *et al.*, 1992] utilises a global constraint on the scene (that the projective transformation induced by the camera, projects the parallel lines of the road so that they meet at the vanishing point). This does not help with any edges that are found to be parallel with the road (cracks, oil stains, fences etc.). Additionally, many systems rely on the use of lane markings to search for dark–light–dark regions within the image. However, it is undesirable to rely solely on this information considering the lack of detectable lane markings on Australian roads. Therefore it is beneficial to use a set of substantially different cues for the detection of road regions that do not have the same inadequacies as each other. If enough redundant cues are used so that the inadequacies of each can be factored out, then a robust lane detection algorithm can be constructed.

Some possible cues that will be investigated in this project include:

- Feature edge detection.
- Colour segmentation.
- 3D stereopsis with the assumption of relatively constant road width.
- The 4D approach [Dickmanns and Zapp, 1987].
- The assumption of the road as a plane.

The architecture of the system uses a probabilistic approach to combine data from the different cues and a condensation algorithm [Blake and Isard, 1998] to control hypothesis generation. Additionally, it is desirable to perform such computations in an efficient

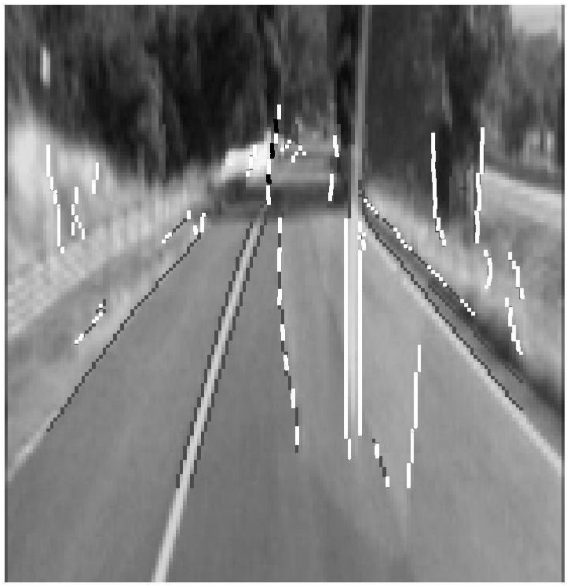

Figure 5: Lane edge feature detection using a single edge based cue

manner by limiting the use of extra cues to periods of high uncertainty. This is particularly useful considering that initialisation is the most difficult task of the lane tracking process. Once the road position is known, robust prediction of future road positions can be made using the continuity of the vehicles motion. This allows higher probabilistic outcomes of the cues and a resulting reduction in computational complexity.

Initial results using only one edge based cue show that the condensation algorithm has a positive effect on the detection and tracking performance of the system. Due to the "reverse" nature of the condensation algorithm where hypothesized road models are verified instead of searched for, many of the *a priori* constraints mentioned above are imposed indirectly on the system, increasing the robustness of the system.

4 Vision Based Obstacle Detection & Tracking

Obstacle detection faces many of the same issues as lane tracking and similarly we believe robust fusion of visual cues and expectation based image processing is the way to maximise the information gleaned from the cameras. In related research an condensation based architecture for face tracking has been developed. The result of multiple visual cues are integrated with varying update frequencies to maximise the tracking performance under a finite computational resource constraint [Loy *et al.*, 2002]. An extension of this architecture will be used for vehicle obstacle tracking. Different methods of image processing will be used depending on the certainty that objects in the visual field are successfully described by a list of current known objects. If tracked objects don't behave as their motion model predicts or if the system is just starting, bootstrapping algorithms are used which perform computationally intensive "bottom up" image processing.

The primary bootstrapping technique we are investigating is realtime 3D depth flow.

Realtime 3D depth flow fuses a stream of stereo depth map information with conventional 2D optical flow to derive an estimate of the 3D position and relative velocity of in the direction of each pixel in the stereo image [Kagami *et al.*, 1999]. The method relies on area based correlation between the images. On the active camera platform the camera pair share the same tilt axis. When the cameras are parallel the lines of disparity due to depth (that is the epipolar lines) are also parallel. When the cameras are verged or diverged a homographic transformation to each image is required can restore the parallel imaging geometry. With parallel images the depth of a point in 3D space corresponds to a horizontal disparity between the matched points in the images. 2D optical flow is calculated using matched regions in consecutive frames from one camera, in this case there can be a vertical and horizontal component to the disparity.

In both the disparity map and the optical flow field consistency checking is used to cull noisy results. Consistency checking is a particularly convenient validation process as most of the computations required are done as a product of the original disparity map and optical flow calculation. Sub pixel interpolation is also done using a quadratic approximation.

The 3D depth flow is produced by combining the depth map and the 2D optical flow. The 2D optical flow can be computed on either the left or the right image sequence. The choice is abitrary as it will only change the orientation of the 3D depth flow coordinate frame relative to the vehicle. We use the left image sequence.

The 2D optical flow gives the X and Y velocity of objects projected into the image plane, the Z velocity is determined by finding the change in depth of the object from time t_k to t_{k+1}.
That is the change in depth ΔZ at time k is given by:

$$\Delta Z_{k+1} = Depth_{k+1}(X_{k+1}, Y_{k+1}) - Depth_k(X_k, Y_k) \tag{1}$$

where X_k & Y_k represent the projection of the 3D point in the image plane at time k. $Depth_k(X, Y)$ represents the projected depth in the direction (X, Y) at time k.

Figure 6 shows: (a) the left image from the stereo camera pair, (b) a surface generated from the disparity map and (c) a section of the 3D depth flow around the overtaking car. The overtaking car shows up as a lump on the right hand side of the surface (Fig. 6b), the tree line can be seen in the top right corner. The road is visible across the bottom and to the left. Spikes such as the one in the top left of the surface are noise caused by aliasing. In the 3D depth flow (Fig. 6c) instantaneous velocity vectors of points on the overtaking car

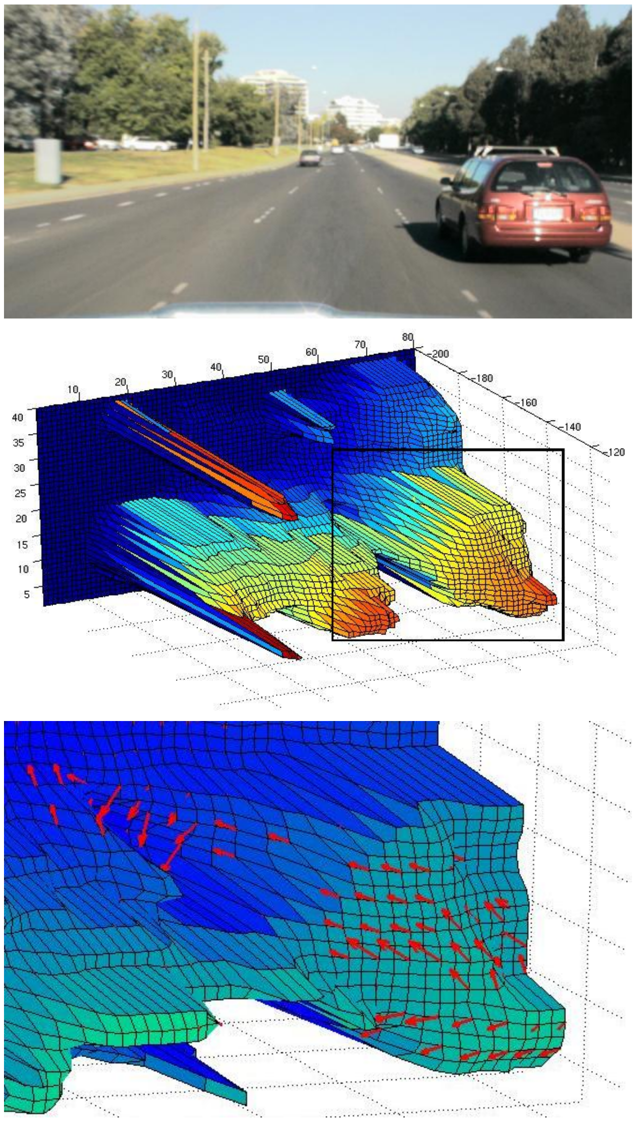

Figure 6: A.(top): left image from stereo pair, B.(middle): 3D surface from stereo disparity (rectangle indicates region of 3D depth flow), C.(bottom): 3D depth flow.

are shown, like 2D optical flow, the measurements are noisy but can be combined to get a better result.

Our disparity map implementation uses a Laplace of Guassian (LoG) filter followed by sum of absolute difference (SAD) template correlation. Normalised cross correlation (NCC) has also be implemented, however our of LoG & SAD correlation is currently faster than the NCC algorithm and as [Banks *et al.*, 1997] has shown the results are similar. To acheive real time performance, both the depth map and the optical flow are computed recursively (which reduces redundancy in the calculations), use the cache efficiently by grouping operations operating on the same image region and use Single Instruction, Multiple Dataset (SIMD) instructions (MMX & SSE) [Kagami *et al.*, 1999][Kagami *et al.*, 2000].

Although the 3D depth flow is optimised it runs is still too slow to be used at full resolution at frame rate (currently around 5Hz at 640x480 & 10Hz at 320x240). Instead the 3D depth flow will run as a background processing task performing a "catch all" function for detecting new obstacles. Obstacles are segmented from the 3D depth flow by looking for regions bounded by discontinuities of depth and relative velocity. Detected Obstacles will be tagged as known objects and tracked using consistent colour, template tracking and a simple motion model.

Conclusion

An introduction to the Autonomous Vehicle Project (AVP) at the ANU has been given with a hardware overview of the vehicle. The AVP will concentrate on computer vision for lane tracking and obstacle detection. A preliminary investigation into lane tracking reinforces the need for the fusion of multiple image primatives for reliable results. An approach to obstacle detection & tracking is outlined including the intended bootstrapping mechanism.

Acknowledgements

This work is partially supported by the Centre for Accident Research and Road Safty - Queensland.

References

[Banks *et al.*, 1997] Jasmine Banks, Mohammed Bennamoun, and Peter Corke. Fast and robust stereo matching algorithms for mining automation. In *Proc. International Workshop on Image Analysis and Information Fusion*, pages 139–149, November 1997.

[Batavia *et al.*, 1998] Parag H. Batavia, Dean A. Pomerleau, and Charles E. Thorpe. Predicting lane position for roadway departure prevention. In *Proc. IEEE Intelligent Vehicles Symposium*, 1998.

[Blake and Isard, 1998] Andrew Blake and Michael Isard. *Active Contours*. Springer, Great Britain, 1998.

[Dickmanns and Zapp, 1987] E. D. Dickmanns and A. Zapp. Automonous high speed road vehicle guidance by computer vision. In R. Isermann, editor, *Automatic Control—World Congress, 1987: Selected Papers from the 10th Triennial World Congress of the International Federation of Automatic Control*, pages 221–226, Munich, Germany, 1987. Pergamon.

[Dickmanns, 1999] Ernst D. Dickmanns. An expectation-based, multi-focal, saccadic (ems) vision system for vehicle guidance. In *Proc. International Symposium on Robotics and Research*, Salt Lake City, Utah, October 1999.

[HoRSCoCTA, 2000] HoRSCoCTA. House of Representatives Standing Committee on Communications, Transport and the Arts: Beyond the Midnight Oil: Managing Fatigue in Transport. Technical report, House of Representatives Standing Committee on Communications, Transport and the Arts, October 2000.

[Jarvis, 2000] Ray Jarvis. Intelligent sensor based road vehicle driver assistance. In *Proc. Australian Conference on Robotics and Automation (ACRA)*, pages 173–178, Melbourne, Australia, August 2000.

[Kagami *et al.*, 1999] S. Kagami, K. Okada, M. Inaba, and H. Inoue. Real-time 3d flow generation system. In *Proc. IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems*, Taipei, Taiwan, 1999. IEEE Computer Press.

[Kagami *et al.*, 2000] S. Kagami, K. Okada, M. Inaba, and H. Inoue. Design and implementation of onbody real-time depthmap generation system. In *Proc. IEEE Int. Conf. on Robotics and Automation*, California, USA, 2000. IEEE Computer Press.

[Kato and Ninomiya, 2000] Takeo Kato and Yoshiki Ninomiya. An approach to vehicle recognition using supervised learning. *IEIC Transations on Information and Systems, Special Issue on Machine Vision Applications*, E83-D(7):1475–1479, July 2000.

[Loy *et al.*, 2002] Gareth Loy, Luke Fletcher, Nicholas Apostoloff, and Alexander Zelinsky. An adaptive fusion architecture for target tracking. Submission to IEEE International Conference on Face and Gesture Recognition, May 2002.

[Maurer, 2000] M. Maurer. Ems-vision: Knowledge representation for flexible automation of land vehicles. In *Proc. Intelligent Vehicles*, Dearborn, MI, October 2000.

[Seeing Machines, 2001] Seeing Machines. FaceLAB: A face and eye tracking system. http:www.seeingmachines.com, 2001.

[Suzuki *et al.*, 1992] A. Suzuki, N. Yasui, N. Nakano, and M. Kaneko. Lane recognition system for guiding of autonomous vehicle. In *Proceedings of the Intelligent Vehicles Symposium, 1992*, pages 196–201, 1992.

[Truong *et al.*, 2000] H. Truong, S. Abdallah, S. Rougeaux, and A. Zelinsky. A novel mechanism for stereo active vision. In *Proc. Australian Conference on Robotics and Automation*, Melbourne, Australia, 2000. ARAA.

[Williamson and Thorpe, 1999] Todd Williamson and Charles Thorpe. A trinocular stereo system for highway obstacle detection. In *Proc. International Conference on Robotics and Automation (ICRA99)*, 1999.