Driver Assistance Systems based on Vision In and Out of Vehicles

Luke Fletcher, Lars Petersson and Alexander Zelinsky
Robotic Systems Laboratory
Department of Systems Engineering, RSISE
The Australian National University
Canberra, ACT, Australia
[luke|larsp|alex]@syseng.anu.edu.au

Abstract

As computer vision based systems like lane tracking, face tracking and obstacle detection mature an enhanced range of driver assistance systems are becoming feasible. This paper introduces a list of core competencies required for a driver assistance system, the issue of building in robustness is highlighted in contrast to leaving such considerations to a later product development phase. We then demonstrate how these issues may be addressed in driver assistance systems based primarily on computer vision. The underlying computer vision systems are discussed followed by an example of a driver support application for lane keeping based on force-feedback through the steering wheel.

1 Introduction

Almost every driver has experienced a warning from a passenger about an obscured car while merging or a jaywalking pedestrian in a blind spot. These forms of assistance save countless lives every day.
Computer based assistive systems have already found their way into vehicles in the form of anti skid braking (ABS) and cruise control systems, but after more than a decade, autonomous systems such as those developed by [7] [15] [4] have not been realised commercially. A key distinction between the commercial systems and the promising R&D systems is acceptance on the part of the stake holders: vehicle manufacturers, traffic authorities and the driver. A paradigm shift to autonomous vehicles is unlikely so autonomous technologies must make in roads elsewhere. An alternative is driver support sub-systems. These small sub-systems can be tailored to solve well defined tasks that attempt to support, not replace the driver. These sub-systems are the focus of research at the ANU's road vehicle project.

Examples of such sub-systems could be:

- driver fatigue or inattention detection
- pedestrian spotting and blind spot checking
- driver feedback for lane keeping
- merging assistance
  (i.e. is there sufficient clearance between cars?)
- road context monitoring
  (i.e. speed zone, traffic signal tracking)

Systems which perform these types of supporting tasks can generally be called *Driver Assistance Systems (DAS)*. The possible benefit of these systems can be illustrated by looking at the driver fatigue case. It is estimated that around 30 percent of all accidents involve driver fatigue [10], imagine the difference a vigilant assistant could make.

Robustness is of paramount importance when creating systems to be used in cars that are driving on public roads. The sensing and detection problems must be solved reliably. Fortunately, roads are designed to be: high contrast, predictable in layout and governed by simple rules. This makes the sensing problem somewhat easier, although by no means trivial. Complementary sensors and algorithms can be used to reduce the likelihood of a catastrophic failure but robust systems require performance metrics and graceful failure modes built-in from the start.

Systems also need to be operable in all driving environments. This means urban roads as well as highways need to be serviced. Cluttered urban roads have proved difficult in the past due to an explosion in road scene complexity [5]. Human drivers cope by relying much more extensively on predicting the behaviour of other road users and pedestrians in these situations than on highways. These powers of higher reasoning, which can often involve making eye contact with other road users, are not easily modelled and will not come easily to artificially intelligent systems.

We consider two key approaches:

1. A robust multi-hypothesis, multi-visual cue algorithm to adapt to changing conditions.
2. Driver monitoring not only for fatigue detection but for road scene validation as well.

Next we will identify the core competencies required by a driver assistance system. A description of the research vehicle and underlying vision systems will follow. Then an example driver assistance sub-system namely lane keeping using force feedback will be given.

2 Driver Assistance Systems

By conducting a brief analysis of the different assistance scenarios listed in the introduction some obvious knowledge bases and actions can be identified. Be it human co-pilot or automatic system the driver assistant requires knowledge of: speed, acceleration, direction, position on road, lie of road ahead, location of vehicles & potential obstacles, an a priori model of the dynamics of the vehicle and lastly the driver's own behaviour. The assistant also needs to be able to deliberate on the before mentioned knowledge and the consequences over time and act through communication with the driver or secondary control of the vehicle.

The above analysis can be categorised as follows:

- Traffic situation monitoring
- Driver's state monitoring
- Vehicle state monitoring
- Communication with the driver
- Vehicle control
- Reasoning system

Diverging from the human co-pilot analogy a automatic driver assistance system must be also be intuitive, unobtrusive and overridable. *Intuitive* in that the behaviour of the system makes immediate sense in the context of the standard driving task. On the whole *unobtrusive* as driver assistance should aid the driver not distract or disrupt unless deemed necessary. *Overridable* in that ultimate control rests with the driver. The driver should be able to refuse assistance.

Several groups have been looking at driver assistance style systems. Adaptive cruise control (ACC) such as the DISTRONIC system offered by Daimler Chrysler is a good example of autonomous technologies integrated into driver assistance subsystem, though ACC systems using vision alone are not as mature. Adaption to varying visual conditions is a theme identified and addressed by several groups. Carnegie Mellon University's lane tracking systems in particular, the enhanced ALVINN and RALPH lane tracking systems explicitly dealt with changed road conditions [15]. Though Kalman filtering has proved a corner stone of vision based autonomous technologies we are unaware of any previous use of particle filtering (aka condensation) applied to the road scene vision problem. Stereo disparity and optical flow have both been applied to the road obstacle detection problem [4][19]. Franke et. al. [9] have shown a DAS style system that combines stereo disparity and optical flow for near field obstacle detection. While Labayrade et. al. have demonstrated a simple yet effective mechanism for recovering higher level information from stereo disparity maps based on an accumulator array [12].

Figure 1: The vision platforms in the vehicle. The CeDAR active vision head and FaceLAB passive stereo cameras are labelled.

3 TREV: the Transport Research Experimental Vehicle

The transport research experimental vehicle (TREV) used in this project is a Toyota Land Cruiser with a variety of sensors and actuators installed to support a variety of ITS related research. Vision is the primary sense used on board the vehicle, which has two major systems installed (see Fig. 1). A CeDAR stereo active camera platform [20] is mounted in place of the rear view mirror and is used for monitoring the road scene in front of the vehicle. This system carries 4 cameras - one pair used for stereo vision in the near-field, and one pair for mid-field stereo experiments. For driver monitoring a FaceLAB head & eye tracking system has been mounted on the dashboard, this system is discussed further in the next section.
The experiments in this paper also use a Hall effect speed sensor, steering shaft encoder and a steering angle potentiometer. A steering actuator based on a

geared down DC motor is capable of controlling the power steering assisted vehicle, in this case it is used to apply a feedback to the driver, not steer the vehicle. Strain gauges are also fitted to the steering shaft to measure torque applied by the driver.
Various other sensors and actuated systems have been fitted to the vehicle more detailed description of the experimental vehicle is given in [8].

The processing is done on a standard PC architecture. There is a primary PC that controls the actuators and reads all non-vision sensors, a second PC supports the FaceLAB system and a third PC that processes the road scene images. As the vehicle is to be used for a variety of driver assistance systems a great deal of flexibility is desired in terms of which devices are to be used and how. To support this flexibility and to try and make the complexity of the system tractable we have opted for a common object request broker architecture (CORBA) based interprocess communication (IPC) system [14]. This gives an implicit client server style modularity across the system. For example the steering actuator can be controlled by the vision PC by accessing a remote 'steering' object provided by the primary PC. The fact that the 'steering' object is really implemented as a process on a remote host does not affect the application. CORBA has the benefit over traditional IPC techniques of being inherently object oriented, platform independent and implementation independent hence several operating systems and programming languages can be used together in concert.

4 Vision In and Out of Vehicles

4.1 Vision Inside Vehicles: FaceLAB

FaceLAB is a driver monitoring system developed by Seeingmachines [17] in conjunction with ANU and Volvo Technological Development. It uses a passive stereo pair of cameras mounted on the dashboard to capture video images of the driver's head. These images are processed in real-time to determine the 3D pose of the persons face ($\pm$1mm, $\pm$1 deg) as well as the eye gaze direction ($\pm$3 deg), blink rates and eye closure. Clinical trials show that head position and eye closure are key indicators for the detection of fatigue in drivers [10]. When augmented with information about the vehicle and traffic situation additional inferences can be made. Apostoloff et. al. [3] was able to show a clear correlation between the eye gaze direction and the curvature of the road, particularly an apparent monitoring of the oncoming traffic.

In addition to the direct observation of the driver for fatigue and inattention detection, driver monitoring is used to validate the road scene monitoring applications.

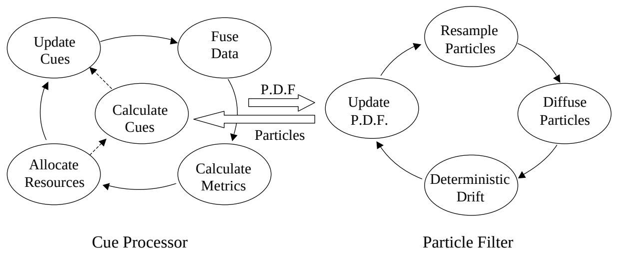

Figure 2: Distillation: a visual cue processing framework

By monitoring where the driver is looking, many false alarms can be avoided. If the driver is looking at a potential problem/uncertain area in the road scene, a warning is irrelevant. This is based on the higher goal of driver assistive systems to assist the driver by informing them of occurrences they may not be aware of, not second-guessing the driver's choices when they are paying attention. So in a complex traffic scene as long as an identified hazard, such as an overtaking car, or an unpredictable hazard, such as wandering pedestrian, are noted by the driver no action needs to be taken.

4.2 Vision Outside Vehicles

Despite many impressive results in the past, it is clear that no single visual processing method can perform reliably in all traffic situations. Many groups have reported issues regarding road appearance changes due to shadows, solar glare and road works [6][15]. With the exception of where the physical limits of the sensor plays a role, in order to achieve robust lane tracking and obstacle detection multiple methods of image processing need to be undertaken and selected based on the prevailing conditions faced by the system. We have developed a visual cue processing framework, "Distillation", to accommodate these kinds of robustness issues. The framework is based on a visual cue scheduling and integration system in conjunction with a particle filter (or condensation algorithm) (see Fig 2). The framework is able to *distill* the contributions of various visual cues and test hypotheses into an overall best estimate of the state space distribution.

The Distillation framework provides:

- the fusion of visual cues using Bayesian theory
- the dynamic allocation of resources over a suite of cues based on merit
- a top-down hypothesis testing approach
- performance metrics
- the ability to combine cues at different rates
- the ability to exploit the reuse of image processing steps common between cues.

A more detailed discussion of the algorithm is in [2]. The Distillation vision framework has also been demonstrated in people tracking [13].

4.3 Lane Tracking

In the lane tracking application visual cues such as edges for finding lane marks and road colour consistency are combined with cues based on physical world constraints such as vanishing points and plausible road shapes to distill a winning hypothesis of the vehicle's position with respect to the road and the geometry of the road ahead. Currently the application estimates for the the road width, the lateral offset of the vehicle from the centerline of the road and the yaw of the vehicle with respect to the centerline of the road. Future enhancement will see estimates the horizontal and vertical road curvature in the mid to far-field ranges. More details about the lane tracking system specifically is available in the companion paper [1].

4.4 Obstacle Detection and Tracking

As mentioned in Frank et. al. [9] the range of optical flow values and disparities encountered in the road scene is large. Disparities and image motion in a single instance can easily range from 0 at the horizon to over 64 pixels in the near field. Frank et. al. limited the vehicle speed in their experiments so that the gradient based optical flow estimation constraint of image motion of less than 2 pixels per frame was honoured. They mention that future work could include a solution using gaussian image pyramids to enhance the dynamic range possible in the flow estimation. We have adopted a image pyramid technique both in the optical flow and disparity map estimation. For the case of optical flow we implement a method similar to Simocelli [18]. The optical flow is computed for the most coarse images then the result used to warp the next higher image resolution to maintain an acceptably small image motion at each level. The penalty for using a coarse to fine approach is that any errors occurring at any image resolution are propagated and amplified into the finer images.

Using image pyramids for disparity map estimation give a couple of added benefits in addition to increasing the range of disparities estimated, in the case of the disparity map estimation no image warping between resolutions is performed. Correlation techniques are plagued with the issue of using the correct correlation window size. A large window allows a more reliable match but causes overly smooth disparity maps. A small window size allows for finer features to be represented but introduces noise due to erroneous matches. Fortunately obstacle detection in road scenes usually support the rule of thumb that close objects are large and distant objects, such as vehicles down the road, are small. Using an image pyramid and calculating the disparity for each image resolution with the same size correlation window means that the correlation window is effectively halving for each image resolution going from course to fine. This property is exactly what we would prefer to match large objects at large disparities and smaller objects at small disparities (like near the horizon). Also, by not warping the image between resolutions we can avoid the propagation of errors between image resolutions. At higher resolutions we are interested in finding distant objects with small disparities, where as larger objects such as a close vehicles are recovered at a coarse image resolution. There will be a issue that coarse resolution images can only resolve disparities to half the accuracy of the next higher resolution images but as this works in opposition to the property of disparity estimates deteriorating as distances increase the effect on the resultant disparity map is acceptable.

The obstacle detection and tracking system is composed of 3 main levels. The most primitive level uses a set of “bottom up” whole image techniques to search the image space for likely obstacle candidates. Stereo disparity and optical flow are principally used at this level. While very noisy the disparity and flow information can be combined to form a 3D depth flow field [11] (see Fig 3). The first iteration of our technique used template matching to find the disparities and optical flows. This is the method used in Fig 3, the second iteration of our technique uses image pyramids as discussed above. Possible obstacle candidates can also be derived using colour consistency. Sets of particles representing each obstacle candidate are injected into the particle filter state-space inside the Distillation framework in a gaussian distribution around the estimated location. The obstacles are tracked in a state-space consisting of a bearing from the center of mass of the research vehicle and one dimension representing the estimated obstacle size. Particles representing unsubstantiated obstacle candidates are eventually resampled to other obstacle candidates. The remaining potential obstacles are tracked within Distillation framework between frames. Each cluster of particles that survive a minimum number of iterations is then checked to against a uni-modal gaussian distribution at it's centroid. If the gaussian distribution adequately describes the cluster a uniqueness operator is applied to the region of the obstacle in each if the stereo images and a set of correlation templates are taken at the most unique points of the obstacle. If the correlation templates are tracking reliably and the distribution is still sufficiently gaussian a Kalman filter is spawned solely to track this obstacle. The obstacles are then tracked using correlation templates alone. If the Kalman filter starts to diverge the location is treated as a obstacle candidate again and particles are injected back into the particle filter and the Kalman filter is discarded.

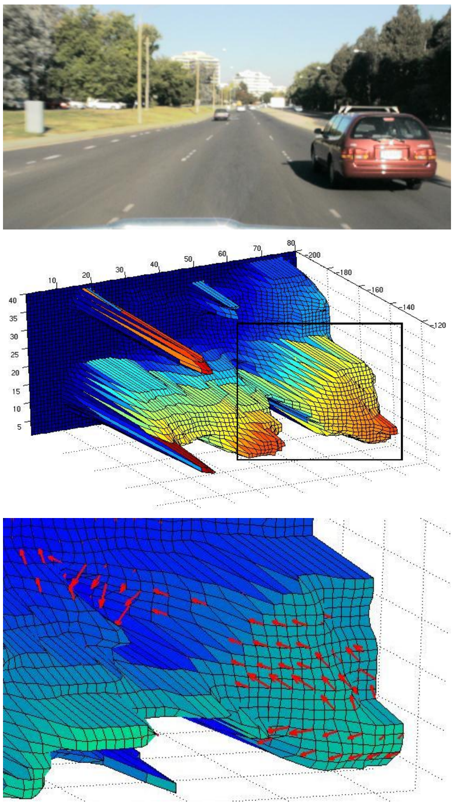

Figure 3: A.(top): left image from stereo pair, B.(middle): 3D surface from stereo disparity (rectangle indicates region of 3D depth flow), C.(bottom): 3D depth flow.

5 Lane Keeping using Force-Feedback

As mentioned earlier Driver Assistance Systems should be intuitive and unobtrusive to the driver. The principle aim of this driver assistance sub-system is to prevent unintentional lane departures not seize control of the vehicle. By applying force-feedback to the steering wheel we are leaving the underlying familiar mechanism of steering intact and instead are merely suggesting a bias in favor of the lane direction.

In this case we must use: *Traffic situation state monitoring* to estimate where the vehicle is in the lane using lane tracking, *Vehicle state monitoring* to determine the current steering angle using a potentiometer and *Driver state monitoring* to estimate the intention of the driver, strain gauges on the steering shaft allow us to register torque applied by the driver. The DAS can then adjust the actuator response to match the compliance appropriate for the lane offset.

Though not yet considered the FaceLAB system can also help identify the driver's intention. Veering due to fatigue versus planned overtaking or stopping could be detected by observing where the driver is actually

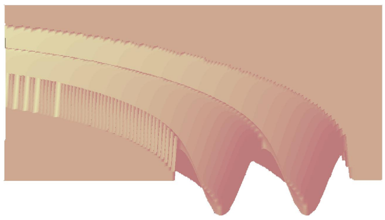

Figure 4: The potential field as the car drives on a curved section of road with two lanes.

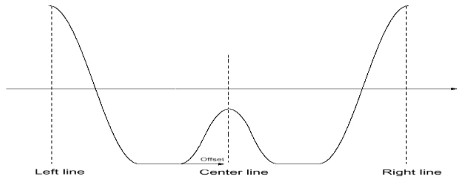

Figure 5: A potential field, V, for a two lane road. The centers of each lane are at the two local minima.

looking. During occasions of lane tracking loss if the driver is attentive no action will be taken, otherwise a warning would be issued.

The force-feedback is implemented along the lines of [16], where a virtual force framework for lateral vehicle control is developed which allows superposition of virtual forces to be applied to the car that can represent control inputs from a range of different kinds of DASs. An appropriate control force, F, is calculated using a potential field approach that is then used to derive the actual control command. Suppose we would like our car to drive in a virtual valley as is shown in Figure 4. The virtual force will become a function of the lateral offset estimated by the lane tracker. In this case virtual valley will have a sinusoidal cross section, in practice we use potential field which is more intuitive to the driver. Figure 5 shows the potential field cross section used, it has flat sections to allow the driver free movement and a smaller potential between lanes than at the road boundary to permit lane changes.

It can be shown that the required corrective virtual force is proportional to the derivative of the potential field cross section with respect to the lateral offset. The control law is also augmented with a damping term to prevent oscillations between the valley boundaries. It is worth noting that the approach allows extensions in the form of other complementary DAS such as e.g. stability control. Care must, however, be taken so that the sum of superimposed potential fields does not have local minima at the wrong places.

Some experiments have been conducted but rework to the steering actuator has prevented meaningful experimental data to date from being collected. The rework which has just been finished included adding an encoder to the steering motor and has enabled a much better response.

6 Conclusions & Future Work

In this paper, driver assistance systems were discussed. A number of necessary core competencies were identified by a comparison with a human co-pilot. The issue of robustness has been highlighted and our solution strategies of a adaptive vision framework and driver monitoring have been outlined. Three vision systems which make up the primary senses for driver assistance systems have been described. And a lane keeping system using force feedback was presented as an example of a driver assistance task oriented sub-system.
Future work includes integrating FaceLAB into the lane keeping assistant allowing the driver to depart the lane without restriction when looking in an appropriate direction, development of other vision based driver assistance sub-systems.

References

[1] Nicholas Apostoloff and Alex Zelinsky. Robust vision based lane tracking using multiple cues and particle filtering. In *Proc. IEEE Symposium on Intelligent Vehicles*, 2003.

[2] Nicholas Apostoloff and Alex Zelinsky. Vision in and out of vehicles: integrated driver and road scene monitoring. *International Journal of Robotics Research*, pages 5–28, 2003.

[3] Nicholas Apostoloff and Alexander Zelinsky. Vision in and out of vehicles: integrated driver and road scene monitoring. In Bruno Siciliano and Paolo Dario, editors, *Experimental Robotics VIII*, Advanced Robotics Series. Springer-Verlag, 2002.

[4] Massimo Bertozzi and Alberto Broggi. Gold: a parallel real-time stereo vision system for generic obstacle and lane detection. In *IEEE Transactions on Image Processing*, volume 7, pages 62–81, January 1998.

[5] Massimo Bertozzi, Alberto Broggi, Massimo Cellario, Alessandra Fascioli, Paolo Lombardi, and Marco Porta. Artificial vision in road vehicles. *Proceedings of the IEEE*, 90(7):1258–1270, July 2002.

[6] Massimo Bertozzi, Alberto Broggi, and Alessandra Fascioli. Vision-based intelligent vehicles: State of the art and perspectives. *Robotics and Autonomous Systems*, 32(1–16), 2000.

[7] 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.

[8] Luke Fletcher, Nicholas Apostoloff, Jason Chen, and Alexander Zelinsky. Computer vision for vehicle monitoring and control. In *Proc. Australian Conference on Robotics and Automation*, 2001.

[9] U. Franke and S. Heinrich. Fast obstacle detection for urban traffic situations. 3(3):173–181, September 2002.

[10] N. L. Haworth, T. J. Triggs, and E. M. Grey. Driver fatigue: Concepts, measurement and crash countermeasures. Technical report, Federal Office of Road Safety Contract Report 72 by Human Factors Group, Department of Psychology, Monash University, 1988.

[11] 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.

[12] Raphael Labayrade, Didier Aubert, and Jean-Philippe Tarel. Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation. In *Proc. IEEE Intelligent Vehicle Symposium*, France, June 2002.

[13] Gareth Loy, Luke Fletcher, Nicholas Apostoloff, and Alexander Zelinsky. An adaptive fusion architecture for target tracking. In *Proc. The 5th International Conference on Automatic Face and Gesture Recognition*, Washington DC, May 2002.

[14] omg.org. Corba/iiop 2.4 specification. URL:http://www.omg.org/, January 2002.

[15] Dean Pomerleau. Ralph: Rapidly adapting lateral position handler. In *Proc. IEEE Symposium on Intelligent Vehicles*, pages 506 – 511, September 1995.

[16] Eric J. Rossetter and J. Christian Gerdes. A study of lateral vehicle control under a 'virtual' force framework. In *Proceedings of the 2002 AVEC Conference*, Hiroshima, Japan, 2002.

[17] Seeing Machines. Facelab face and eye tracking system. *http:www.seeingmachines.com*, 2001.

[18] E. P. Simoncelli. *Bayesian Multi-scale Differential Optical Flow*, volume 2, chapter 14, pages 297–422. Academic Press, 1999.

[19] S. M. Smith and J. M. Brady. Asset-2: Real-time motion segmentation and shape tracking. 17(8):814–820, august 1995.

[20] Orson Sutherland, Harley Truong, Sebastien Rougeaux, and Alexander Zelinsky. Advancing active vision systems by improved design and control. In *Proceedings of International Symposium on Experimental Robotics (ISER2000)*, December 2000.