Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Robotics and Autonomous Systems 52 (2005) 71–84

Robotics and Autonomous Systems

www.elsevier.com/locate/robot

Correlating driver gaze with the road scene for driver assistance systems

Luke Fletcher[a,*], Gareth Loy[b], Nick Barnes[c,a], Alexander Zelinsky[d]

[a] *Department Information Engineering, Australian National University, Canberra, Australia*
[b] *Computer Vision and Active Perception Laboratory, Royal Institute of Technology, Stockholm, Sweden*
[c] *Autonomous Systems and Sensing Technology, National ICT Australia, Canberra, Australia*
[d] *ICT Centre, CSIRO, Sydney, Australia*

Received 5 March 2005; accepted 21 March 2005
Available online 24 May 2005

Abstract

A driver assistance system (DAS) should support the driver by monitoring road and vehicle events and presenting relevant and timely information to the driver. It is impossible to know what a driver is thinking, but we can monitor the driver's gaze direction and compare it with the position of information in the driver's viewfield to make inferences. In this way, not only do we monitor the driver's actions, we monitor the driver's observations as well. In this paper we present the automated detection and recognition of road signs, combined with the monitoring of the driver's response. We present a complete system that reads speed signs in real-time, compares the driver's gaze, and provides immediate feedback if it appears the sign has been missed by the driver.

Keywords: Driver assistance; Sign detection

1. Introduction

Cars offer unique challenges in human-machine interation. Vehicles are becoming, in effect, robotic systems that collaborate with the driver. As the automated systems become more capable, how best to manage the on-board human resources is an intriging question. Combining the strengths of machines and humans, and migitating their shortcomings is the goal of intelligent-vehicle research.

Cars also provide unique challenges for robotic vision. They operate in an environment that can be highly dynamic and subject to extremes of illumination. It is, however, a well-structured environment, designed for easy perception. Though things can move faster than in most robot-vision environments, many features are known in advance and their appearance is well-constrained.

In this paper, we demonstrate how robot-vision can be used to create context sensitive driver aids. Using

[*] Corresponding author.
E-mail address: luke.fletcher@anu.edu.au (L. Fletcher).

0921-8890/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.robot.2005.03.010

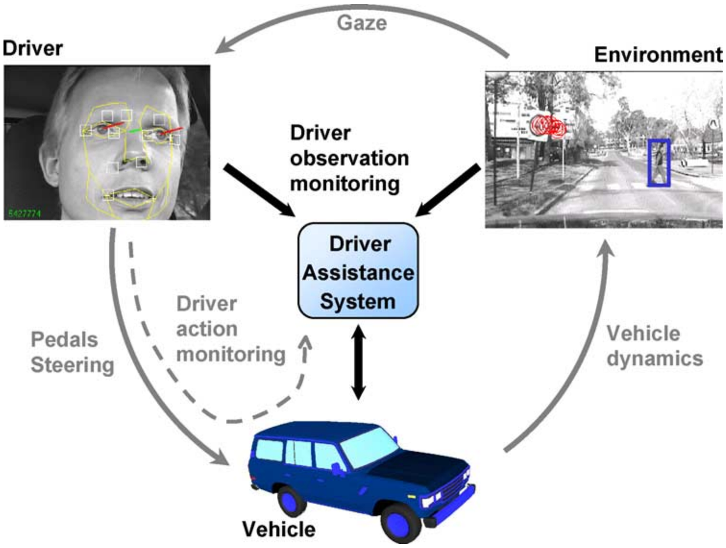


Fig. 1. Introducing driver observation monitoring to supplement driver action monitoring.

developed vision systems that are effective in the dynamic road environment, features relevant to the driver can be robustly identified. Then, by combining direct driver monitoring with the extracted traffic features we can infer far more about the driver's behaviour, not only through monitoring the driver's actions, but also by monitoring the driver's observations (as illustrated in Fig. 1).

Driver monitoring and real-time sign recognition are combined to correlate eye gaze with the sign direction. From this and the vehicle state we develop a system that determines whether the driver should be made aware of the detected sign. A fast image enhancement technique for sign recognition is also demonstrated.

2. Automation in vehicles

Early research in autonomous vehicles focused on fully autonomous driving, generally controlling steering. The revolutionary work by Dickmanns and Graefe [8,7] was able to steer a vehicle at over 100 km/h on well-formed roads. In the early 1990s more robust image processing enabled the SCARF system from the CMU Navlab [5] to handle more degraded roads. Subsequent demonstrations by these groups in the mid 1990s showed impressive robustness. CMU's famous 'No Hands Across America' demonstration, for example, was able to steer autonomously for 98% of a 302 mile trip [13].

However, there are two key problems that prevent the idea of a fully autonomous car being a reality in the near future. First, the remaining few percent required for a vehicle to gain full autonomy is highly challenging. Having an automated system handle all conceivable scenarios is extremely difficult. As accident statistics show, even humans cannot perform this task perfectly. The second problem is insurance. Currently, there must be a human driver.

Driver support, on the other hand, offers immediate possibilities. Here we support the driver by ensuring awareness of relevant aspects of the driving environment, aid simple aspects of driving, while leaving critical decisions to the driver, and importantly, maintaining the driver's sense of control over the vehicle. Giving the driver an increasingly higher level of support is also a path toward full autonomy.

2.1. *What is a driver assistance system?*

A driver assistance system (DAS) is an automated system used to relieve the driver of tedious activities; warn about upcoming or missed events; and possibly take control of the car if an accident is imminent. Depending on the task to be performed, a DAS must have appropriate competencies in a number of areas.

A useful analogy for a driver assistance system is a vigilant co-pilot. Almost every driver has experienced a warning from a passenger about a hidden car or a jaywalking pedestrian. This kind of assistance saves lives every day.

If we momentarily consider a human co-pilot, it is easy to identify the important requirements. The co-pilot must be aware of what is going on outside of the car, e.g., Are there any pedestrians in sight? How is the road turning? Next, to make good judgements, the co-pilot must know where the vehicle is going, how fast, and whether it is braking, accelerating, etc. Moreover, we would like our co-pilot to warn us if we have not noticed an upcoming situation. That means the co-pilot should not only be aware of what is going on outside the car, but also what is happening inside i.e., the driver's responses. A successful driver and co-pilot team requires good communication, but the co-pilot should not be intrusive by presenting the driver with too much or repetitive information.

In summary, a co-pilot/driver assistance system must work intuitively, unobtrusively and be overridable. *Intuitively* in that the behaviour of the system must make immediate sense to the driver in the context of the standard driving task. On the whole, *unobtrusively*, as driver assistance is only an aid if it is not distracting or disruptive unnecessarily. Be *overridable*, in that ultimate responsibility rests with the driver.

The potential benefit of these systems can be highlighted by examining the contribution of inattention in accidents. It is estimated that at least thirty percent of fatal road accidents involve driver inattention [11]. Imagine the difference a vigilant co-pilot could make.

2.2. Gaze monitoring in vehicles

Direct driver monitoring has been the subject of clinical trials for decades, but such monitoring for use in driver assistance systems is relatively new. Head position and eye closure have been identified as strong indicators of fatigue [11]. When augmented with information about the vehicle and traffic, additional inferences can be made. Gordon [10] analysed the motion of the road scene from the driver's view point to draw conclusions as to what perceptual cues were used for driving. In on-road systems, Land and Lee [16] investigated the correlation between eye gaze direction and road curvature, finding the driver tended to fixate on the tangent of the road ahead. Apostoloff and Zelinsky [1] used lane tracking to verify this correlation on logged data, also observing that the driver frequently monitored oncoming traffic. Takemura et al. [27] demonstrated a number of correlations between head and eye movement and driving tasks on logged data.

Hence, in addition to direct observation for fatigue detection, driver monitoring is useful for validating road scene activity. By monitoring where the driver is looking, many unnecessary warnings can be avoided. This is a key mechanism for implementing an *unobtrusive* and *intuitive* system: unnecessary warnings can be suppressed and necessary warnings can be made more relevant. As long as a hazard, such as an overtaking car, or wandering pedestrian, is noted by the driver no action needs to be taken.

2.3. A context sensitive sign recognition DAS

This paper presents a context sensitive sign recognition DAS. An autonomous detection system recognises inportant signs. At the same time, a driver monitoring system verifies whether the driver has looked in the direction of the sign. If it appears the driver is aware of the sign, the information can be made available passively to the driver. If it appears the driver is unaware of the information, the information can be highlighted. In this case, if the driver appears to have seen a speed sign, the current speed-limit can be simply recorded on the dashboard adjacent to the speedometer. However, if it appears the driver has not looked at the sign, and over time, a speed adjustment is expected and has not occurred, a more prominent warning could be given. This still leaves the driver in control of the critical decision, but supports him or her in a way that aims not to be overly intrusive. Warnings are only given when the driver is not aware of the change of conditions. Finally, the warning can be cancelled by observing the driver: a glance at the speedometer confirms that the driver is aware of his or her speed and the detected limit.

2.4. Sign recognition

Sign recognition is an important task for a driver support system. Signs give information relevant to local conditions. They appear clearly in the environment, but a driver may not notice a sign due to distractions or another driving task. In this case it may be helpful to

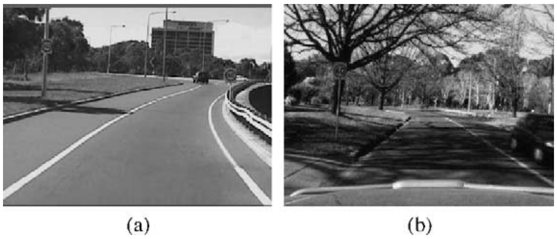


Fig. 2. (a) Algorithms assuming strong road scene structure can miss valuable information, such as a sign on an unexpected side of a lane. (b) Assuming common regions such as sky and road can be identified by large areas of self-similar colour is not always valid.

alert the driver to the information that he or she has missed.

Road sign recognition research has been around since the mid 1980s. A popular approach is to use separate stages for sign detection and classification [20,19,14,23], making use of a detection stage to focus classification at a few potential sign locations. The most common means of detecting potential locations is colour segmentation [23,22,20,14,9,25,12]. These methods typically achieve some robustness under varying lighting conditions by considering the apparent chrominance of signs. However, they are not robust to changing chrominance of the incident light, which can occur in real driving conditions (compare fluorescent and tungsten streetlights and sunlight, for instance). Another approach to detection is *a priori* assumptions about image formation and scene structure [12,22], such as assuming the road is approximately straight, that the road and sky will appear as large uniform regions, or that signs will appear only at particular locations within the image. However, such assumptions can break down, as shown in Fig. 2: signs can occur on either side of a road, at ground level (in the case of temporary roadwork signs) or directly overhead on major highways. Roads can be curved or bumpy, and often cluttered, making it difficult to predict where signs will appear in an image from an on-board vision system.

However, the appearance of road signs is highly restricted. They must be of a particular shape, colour and size, and face on-coming trafffic. Signs are placed to be easily visible, so a driver can see them without looking away from the road. Owing to the orthogonal alignment of signs with the road, the apparent shape of relevant signs is constant. It does not change under different driving conditions and is a strong cue for detecting potential signs.

3. Detecting road sign candidates

Signs are detected by locating sign-like shapes in the input image stream. Australian speed signs are required to have a dark (typically red) circle enclosing the speed-limit. These circles provide a strong visual feature for locating speed signs. We apply the fast radial symmetry operator [18] to detect these circular features, and thus potential speed sign candidates, as demonstrated in [3]. This method can also be extended to detect triangular, diamond (square) and octagonal signs [17]. Here we focus on the speed sign case, and for completeness of presentation include a brief description of the fast radial symmetry detector, summarising from Loy and Zelinsky [18].

3.1. Radial symmetry detector algorithm

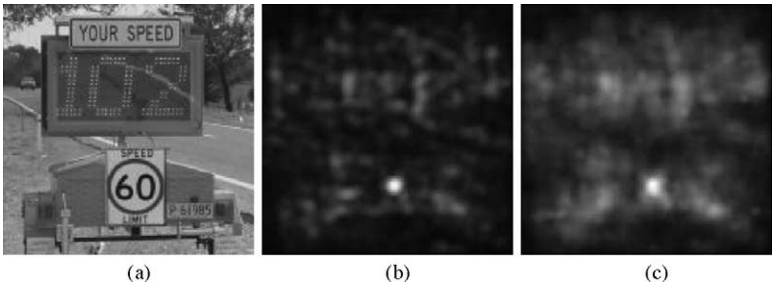
For a given pixel, p, the gradient, g, is calculated using an edge operator that yields orientation, such as Sobel. If p lay on the arc of a circle, then its centre would be in the direction of the gradient, at distance of the circle radius. Robustness to lighting changes is achieved by applying the discrete form of the detector, and insignificant gradient elements (those less than a threshold) are ignored. The location of a pixel that will gain a vote as a potential shape centroid is defined as

$$p_{+\mathrm{ve}} = p + \mathrm{round}\left(\frac{g(p)}{\|g(p)\|}n\right), \quad (1)$$

where $n \in N$ is the radius, and N is the set of possible radii. (A negative image is defined similarly, facilitating constraining the operator to find only light circles on dark backgrounds and vise-versa.) A histogram image O_n is defined by counting the number of votes awarded to each pixel, and truncated to form $\tilde{O}_n$ as follows

$$\tilde{O}_n(p) = \begin{cases} O_n(p), & \text{if } O_n(p) < k_n, \\ k_n, & \text{otherwise.} \end{cases} \quad (2)$$

where k_n is a scaling factor that subsequently normalises $\tilde{O}_n$ across different radii.



(a) (b) (c)

Fig. 3. The radial symmetry detector running on a still image, (a) shows the input image obtained from the internet, (b) the response at radius 20, and (c) the sum of responses for radii 15, 20 and 25.

The response for radius n is then determined as

$$S_n = G\left(sgn(\tilde{O}_n(p))\left(\frac{|\tilde{O}_n(p)|}{k_n}\right)^{\alpha}\right), \tag{3}$$

where G is the Gaussian, and α is the radial strictness parameter (typically 2).

Each radii of N votes into a separate image to facilitate recovery of radius. The full transform is the mean of the contributions over all radii considered

$$S = \frac{1}{|N|}\Sigma_{n\in N} S_n. \tag{4}$$

See Loy and Zelinsky [18] for full details. Fig. 3 shows the result of running this detector on a still image containing a speed sign. The speed sign centroid appears as the dominant maximum in both the response at radius 20 (the closest to the true target radius) and the full response (radii 15, 20 and 25).

This shape-based approach to sign detection has strong robustness to changing illumination as it detects shape based on edges. It returns the centroid of candidate signs as well as their scale. Subsequent computation for classification is well targeted, and comparatively little further computation is needed to assess a candidate.

3.2. Implementation and real-time issues

To adapt the algorithm for the road sign detection case, we apply it only to radii that are practical for detecting signs in traffic images. Small shapes can be ignored, as if there are insufficient pixels present to discern what the sign says there is no point in further processing: we should wait until the sign is close enough to be recognised. In normal driving conditions a sign will never appear closer to the camera than several metres. Given a camera of approximately known focal length, we can impose an upper limit on the possible radii that we are interested in. As the basic shape of a sign will be clear, even if it is faded, we set a threshold that requires a large number of the possible edge pixels to be detected. Further consistency checking can be performed over time i.e., a shape must appear for at least two concurrent frames, its radius must not have changed greatly during that time, and it must not move far in the image.

Previously [3] we demonstrated that shape detection can combine effectively with classification. Using the fast radial symmetry detector, we were able to detect and classify speed signs correctly in the majority of cases using normalised cross-correlation. The full classification was implemented in C++ to evaluate real-time performance. It was found that for a 320 image × 240 image, the full radial symmetry detection and classification was able to be run at 30 Hz, with classification taking ≤1 ms.

3.3. Classification

As the detection phase of the sign recognition process is very effective at culling potential sign candidates i.e., circular objects moving consistently over time, only a simple classification scheme is necessary. The classification needs to reject circular objects that are not signs and differentiate between a small set of potential symbols. Circular objects that are not signs are rejected as a consequence of symbol classification. It is extremely rare that a feature will have a circular border, move consistently and have a high correlating symbol within. Classifying the sign between a set of potential symbols is more difficult: symbol misclassifications far outweigh false sign detections. This topic is examined further in Section 5. The top three peaks detected in the output of the shape detector are tracked over time. Any peak that exhibits temporal and spatial consistency (i.e., small movements) over three frames begins to be enhanced and classified. The classification is done with a (NCC) template correlation of the text as used for the verification above. For these trials the speeds ‘40’, ‘60’, ‘70’ and ‘80’ were classified. The sign detection system signals a sign found when three consecutive frames are classified consistently.

4. Driver awareness

The behaviour of the driver several seconds before and after a sign is detected is used to decide whether to issue a warning. Driver monitoring is achieved via an eye gaze tracking system and vehicle speed monitoring.

4.1. Correlating eye gaze with the road scene

Scene camera and eye configuration is analogous to a two-camera system (see Fig. 4). Gaze directions trace out epipolar lines in the scene camera. If we had a depth estimate of the driver’s gaze, we could project to a point in the scene camera. Similarly, if we had the sign depth we could re-project on to the eye and estimate the required gaze. A depth estimate of the gaze is hard to obtain. A common assumption is that the gaze angle and angle in the scene camera are the same. In practice this assumption amounts to supposing that either the scene camera is sufficiently close to the driver’s head (equivalent to a trivial baseline) or that the objects of interest are near infinity [16,27,26]. In these cases error bounds on the gaze direction (not fixation duration) are infrequently used and even less frequently justified.

The sign depth could be estimated using a second scene camera running the same detection software or assumptions on sign size and/or road layout. However, it is desirable to maintain flexibility of the implemented sign detection system which only uses a single camera and has no strong assumptions on the sign scale. If the depth of the sign is unknown we can instead model the effect of the disparity in our confidence estimate.

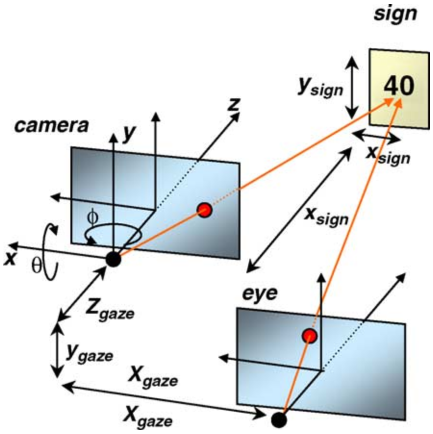


Fig. 4. The scene camera and gaze direction is analogous to a two camera system.

The effect of an unknown stereo disparity will be a displacement along the epipolar line defined by the gaze direction on to the scene camera. The disparity, as with any stereo configuration, will be most apparent for close objects and reduce by a $1/x$ relationship with distance from the baseline. The angular deviation reduces as the angle becomes more obtuse. To get an upper bound of the likely disparity deviation we can compute the worst case disparity for our camera configuration. With reference to Fig. 4, and using the scene camera centre as a world reference frame, the scene camera and gaze angles for a sign at $(X_{\text{sign}}, Y_{\text{sign}}, Z_{\text{sign}})$ can easily be derived as the following equations

$$\begin{aligned}\Delta\theta &= (\theta_{\text{cam}} - \theta_{\text{gaze}})\\ &= \arctan\frac{X_{\text{sign}}}{Z_{\text{sign}}} - \arctan\frac{X_{\text{sign}} + X_{\text{gaze}}}{Z_{\text{sign}} + Z_{\text{gaze}}}, \qquad (5)\end{aligned}$$

$$\begin{aligned}\Delta\phi &= (\phi_{\text{cam}} - \phi_{\text{gaze}})\\ &= \arctan\frac{Y_{\text{sign}}}{Z_{\text{sign}}} - \arctan\frac{Y_{\text{sign}} + Y_{\text{gaze}}}{Z_{\text{sign}} + Z_{\text{gaze}}}. \qquad (6)\end{aligned}$$

The worst case disparity then translates to when the sign is closest to the vehicle on the driver’s side of the road, equivalent to a sign on the right shoulder of a single lane road (note that our system is in a right-hand drive vehicle). The field of view of the scene camera limits the closest point at which the sign is visible. The closest visible sign is at (−3.0, −1.6, 8.0) for the 50° field of view of the camera. The worst case height of the sign relative to the scene camera, −1.6, would be when it is on the ground (this is actually worse than any actual case as the sign is not visible due to the bonnet). With pessimistic estimates of the driver (far) eye position relative to the scene camera manually measured to be: ($X_{\text{gaze}} = 0.22$, $Y_{\text{gaze}} = 0.1$, $Z_{\text{gaze}} = 0.2$) the final errors become;

$$\Delta\theta = (\theta_{\text{cam}} - \theta_{\text{gaze}}) = (20.6^\circ - 18.7^\circ) = 1.9^\circ, \quad (7)$$

$$\Delta\phi = (\phi_{\text{cam}} - \phi_{\text{gaze}}) = (11.3^\circ - 10.4^\circ) = 0.9^\circ. \quad (8)$$

Therefore the worst expected deviation due to stereo disparity is $\pm1.9^\circ$ horizontally and $\pm0.9^\circ$ vertically which is on par with other error sources in the system. The expected deviation for the majority of cases where the sign is further away is significantly less. The deviation is twice as large in the horizontal direction, implying that a suitable approximation of the tolerance region will be an ellipse with a horizontal major axis.

To determine the overall tolerance of the system, two further factors need to be accommodated. The gaze tracking system has an accuracy of $\pm3^\circ$ and the field of view of the foveated region of the eye is estimated to be around $\pm2.6^\circ$ [29]. The accumulated tolerance is the sum of these sources which for our experimental setup comes to $\pm7.5^\circ$ horizontally and $\pm6.6^\circ$ vertically. That is, a 70 pixel × 46 pixel ellipse in the 320 pixel × 240 pixel scene camera image. This allows us to claim that the driver was very unlikely to have seen the sign if the sign and gaze directions deviate by more than this tolerance.

4.2. System setup

To align the scene camera with the gaze direction, the default rotation and scaling between the gaze coordinate system and the scene camera must be determined. While these parameters can be obtained through knowledge of the scene camera parameters and the relative position of the gaze tracking system, an online initialisation is effective and allows easy re-calibration (for zoom changes, gaze head model changes, etc.). The driver is asked to look at several (≥ 4) features visible from the scene camera. Best features are points that approximate points at infinity such as along the horizon. The gaze direction is measured and the points are manually selected in the scene camera. The rotational offset and scaling can then be computed using least squares.

4.3. Implementation

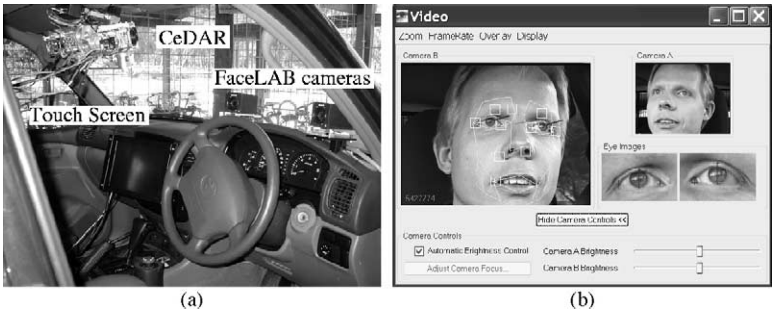
The system was implemented using a modular distributed architecture. Other systems developed within our group, including lane tracking, blind spot monitoring, pedestrian and vehicle detection, can be added and a relatively simple *DAS logic* module written to implement other similar kinds of driver assistance systems. The driver assistance system framework is designed around (see: [21,15]):

- Video based driver monitoring.
- Multi-cue video based road scene monitoring.
- Multi-hypothesis, ambiguity tolerant algorithms.
- Intuitive, unobtrusive, overridable and integrated DAS design goals.

The DAS runs on Pentium IV computers located in the rear of the vehicle. A scene camera was mounted in the centre of the vehicle with approximately the same view (though not field of view) as the driver. The vehicle is fitted with a FaceLAB eye gaze tracking system. Fig. 5(a) shows the scene camera and FaceLAB cameras in the test vehicle. FaceLAB is a driver monitoring system developed by SeeingMachines [24] 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 driver's head (to ±1mm, $\pm1^\circ$) as well as the eye gaze direction (to $\pm3^\circ$). Blink rates and eye closure can also be measured. Fig. 5(b) shows a screen-shot of this system measuring the driver's head pose and eye gaze.

The speed and acceleration of the vehicle was estimated using a hall effect sensor on the tail-shaft of the vehicle. A touch-screen monitor was used to present relevant information and allow the driver to interact with the system.

To correlate the eye gaze with the sign position, the histories of the two information sources were examined. The sign detection sub-system provides a history of the sign location since detected. This includes all frames from when the sign was first detected before the sign was able to be verified or classified. Similarly, the FaceLAB data provides the historical head pose and gaze direction. When a sign has been classified, the sign angles and gaze directions are checked back



(a) (b)

Fig. 5. (a) The cameras in the test vehicle. The CeDAR active vision platform containing the scene camera and FaceLAB passive stereo cameras are labelled. (b) Driver head pose and eye gaze tracking using SeeingMachines FaceLAB system.

in time to when the sign was first detected. If the angles from any previous frame fall within the tolerance, the sign is reported as seen by the driver. If the angles never coincide, the sign is reported as missed. The system provides a four second tolerance for the driver to achieve the speed-limit. The timer is instigated when the measured speed exceeds the limit and the measured acceleration is not significantly decreasing.

4.4. Verification

We conducted a verification experiment to test that the DAS was indeed able to detect whether the driver missed a sign. The driver was asked to fix his gaze on an object in the scene. A sign was then placed at a certain distance from the fixation point. The driver was then asked to identify the sign. The sign was one of eight possibilities. The proportion of correct classifications was logged along with the driver gaze angle and apparent sign position in the scene camera. 30, 20 and 10 m depths were tested against four different displacements between the sign and fixation point. The sign size was 0.45 m in diameter. For each combination of depth and displacement 10 trials were done.

Fig. 6 shows the driver's sign classification error rate versus the angle between gaze and sign position. Expected recognition rates fall as the sign becomes more peripheral in the driver's field of view. The results of this trial verify our expectation that while it is hard to prove the driver saw a sign, it is possible to estimate, with reasonable confidence, when the driver was unlikely to have seen a sign. A curious effect was noticed (represented by a cross in the middle of the graph) when the driver was very close to the sign. The large apparent size of the sign in the driver's field of view seemed to aid the recognition rate. However, this only occurred when signs were close to the vehicle, which is not when drivers typically read signs. The driver reported not consciously being able to see the sign in this case.

This verification demonstrates the expected strength of the system: the ability to detect when the driver has missed a sign. It is impossible to determine whether the driver saw the sign as, even with perfect measurement of a perfect gaze direction match, the driver's attention and depth of focus cannot be determined. If the driver is looking in the direction of the sign, it is an ambiguous case whether the driver read the sign, thus no warning

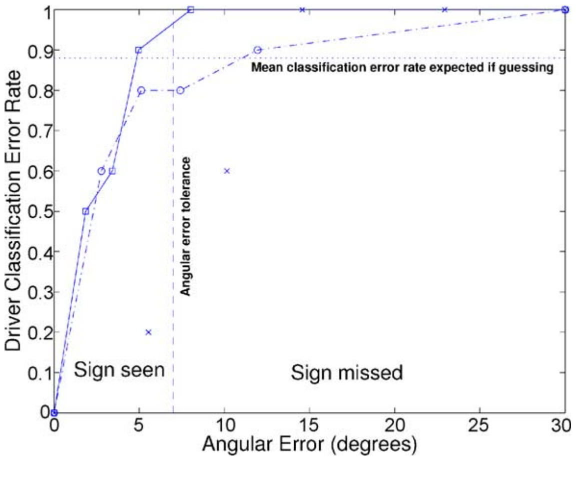
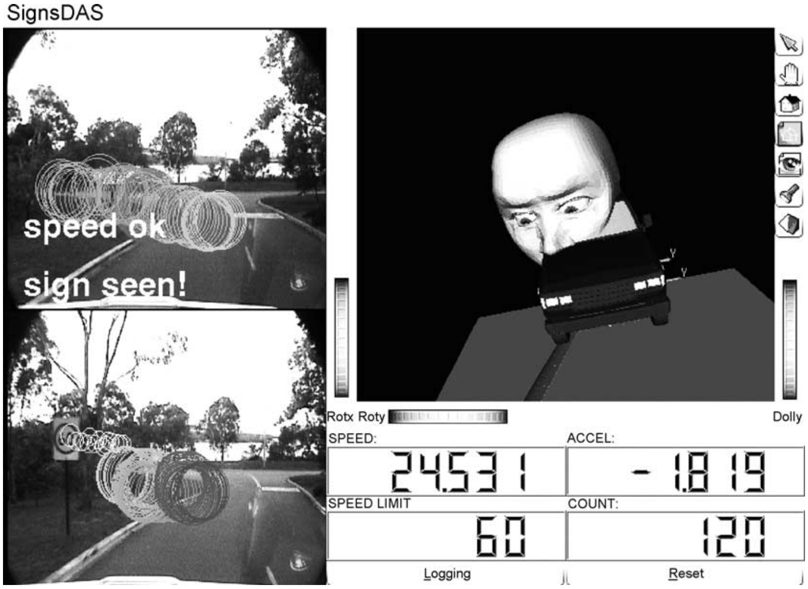


Fig. 6. Driver recognition rate of signs in peripheral vision for various sign depths. Dotted horizontal line shows expected value due to chance. Vertical dashed line represents ±7.5° derived tolerance. it squares: 30 m points. it circles: 20 m points. it crosses: 10 m points.



SPEED: 24.531

ACCEL: - 1.819

SPEED LIMIT 60

COUNT: 120

Fig. 7. Screen-shot showing ‘60’ sign detected and seen by driver. Top left: live video showing eye gaze (*large circles*) and status (*overlaid text*). Bottom left: last detected sign (*small circles*) and eye gaze (*large circles*). Top right: 3D model of current vehicle position, eye gaze (*oversize head*) and sign location. Bottom right: current detected speed-limit, vehicle speed, acceleration and count down for speeding grace period in frames.

is issued. If the driver feels uncertain of the current speed-limit he can always glance at the system and see the last detected sign. In this way the system can create a minimum amount of interference in the normal driving process whilst providing timely information when deemed suitible.

4.5. On road trials

The system was able to detect speed signs around the university and evaluate the implications for the driver. Fig. 7 shows a screenshot of the system demonstrating a typical case. Fig. 8 illustrates the primary scenarios encountered. In Fig. 8(a) the driver was watching a pedestrian and failed to notice a ‘40’ sign. The DAS has detected that the driver did not see the sign and has issued a red *sign: missed!* warning. Fig. 8(b) shows an instance where an ‘80’ sign was detected; the driver saw the sign and the vehicle was not speeding so no red warning was issued. Similarly, in Fig. 8(c) a ‘40’ sign was detected. The driver saw the sign, the system assumed the driver was intentionally speeding so a warning was displayed but no alert issued. In Fig. 8(d) the driver has missed the last sign and is speeding for more than a predefined grace period without decelerating. The *SLOW DOWN!* warning is shown and an alert issued.

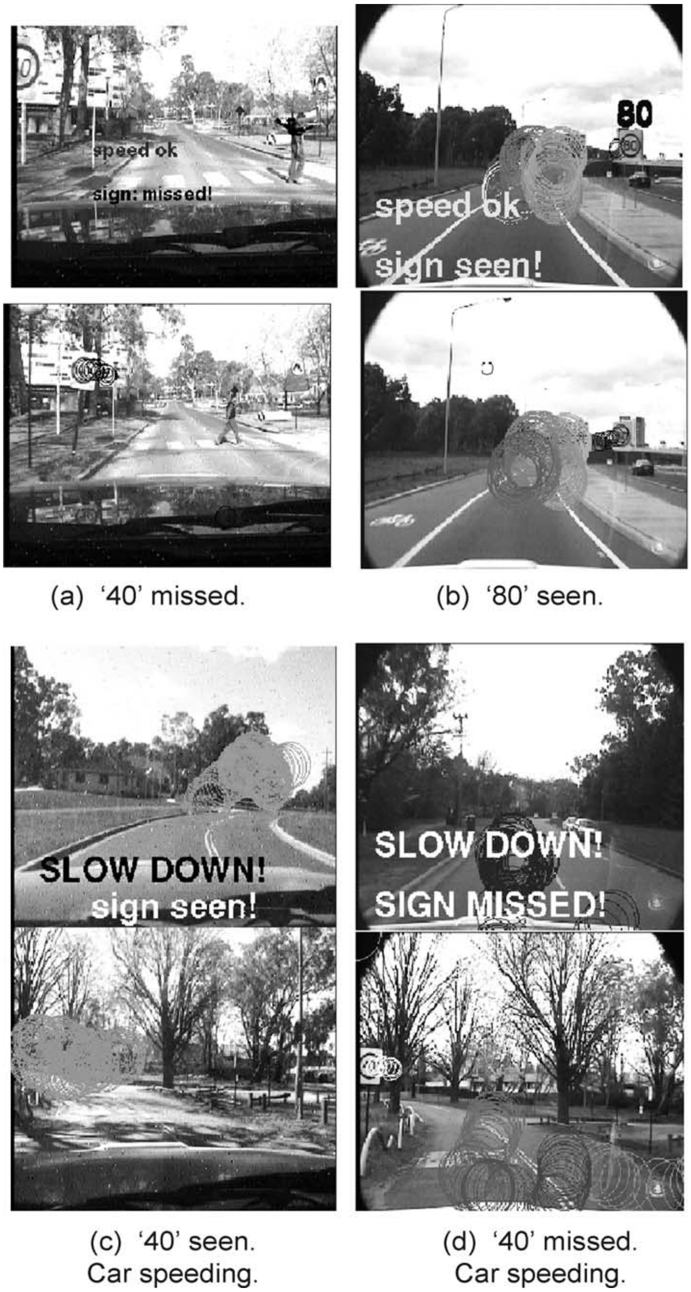
5. Sign image enhancement

By enhancing the sign image we can classify the sign reliably several frames sooner. Poor resolution hampering classification is most evident in still frames of video. Fig. 9 shows a frame from a two sequences. At a glance the speed sign seems well formed and readable, but on closer examination we find substantial degradation of the text. A much studied method of enhancing image quality in video sequences is super-resolution.

Super-resolution is the process of combining multiple low resolution images to form a higher resolution result [28]. The super-resolution problem is usually modelled as the reversal of a degradation process. A high resolution image **I** undergoes a homographic transformation followed by a motion and optical blurring, then, finally, image space sub-sampling to generate the low resolution observation images **O** [4]

$$\mathbf{O}_k = S \downarrow (b_k(H_k\mathbf{I})) + n_k, \tag{9}$$

where H_k is the homography, $b_k(\,)$ is the blurring function and $S\downarrow(\,)$ is a down sampling operation for the kth observation $\mathbf{O}_k$, and n_k is a noise term representing all other errors not modelled. The solution amounts to the 'undoing' of the degradation Eq. (9). The process consists of some form of image *registration*, which is recovering the alignment between the images, then image *reconstruction* where the images are combined to



(a) '40' missed.

(b) '80' seen.

(c) '40' seen.
Car speeding.

(d) '40' missed.
Car speeding.

Fig. 8. Primary scenarios for signs driver assistance system. Left: live video feed showing current view, eye gaze (*dots/large circles*) and current status (*overlaid text*) during screenshot. Right: last detected sign (*small circles*) and eye gaze (*dots/large circles*).

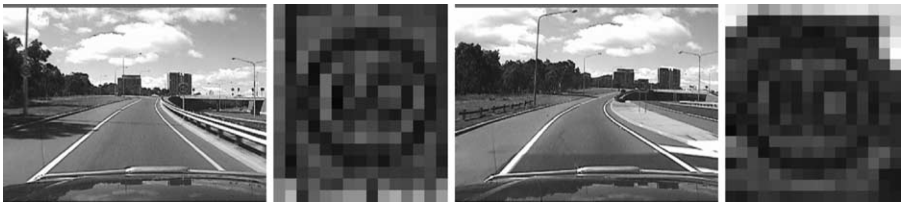


Fig. 9. Still frames from a video camera, with close ups of speed sign. Classifying the signs as 60 or 80 is not obvious.

resolve an estimate of the original image. Registration is usually done by matching feature points with one of the observation images used as a reference and computing the geometric or homographic transforms between each observation image. Reconstruction requires combining the registered images while accounting for the effects of $S \downarrow ()$, $b_k()$ and n_k.

A novel approach was advocated by Dellaert et al. [6] who tracked an object in an image sequence and used an extended Kalman filter to estimate the pose and augmented the state with the super-resolved image. With some optimising assumptions, the group was able to perform online pose and image estimation. The effect of prior knowledge is incorporated neatly into this framework in the derivation of the Jacobian matrices.

5.1. Our approach

For our application, image registration is primarily done using the sign detection algorithm. From each output frame the sign shape is cropped from the video frame and resized. The image is resized using bi-cubic interpolation to the size of the high resolution result image. Baker and Kanade [2] found, as a good rule of thumb, eight times magnification is the upper limit for significant image enhancement. In our case, the lower diameter used by the shape detector is eight pixels, so the high resolution image is chosen to be 64 pixels × 64 pixels. The image is then correlated using normalised cross correlation with the current enhanced image to locate the latest image accurately. The latest image is shifted accordingly and combined with the enhanced image. Images where the correlation coefficient is below a certain threshold (usually 0.5) are discarded as these tend to be gross errors in radius estimate by the shape detector or momentary dominant shapes near the tracked sign, such as apparent shapes caused by tree foliage or background clutter.

The reconstruction is considered as a series of incremental updates of the resolved image from the observations, as shown in Eq. (10). The equation can be rearranged into a first order infinite impulse response (IIR) filter shown in Eq. (11), allowing a fast implementation

$$\hat{\mathbf{I}}_k = \hat{\mathbf{I}}_{k-1} + \lambda c(S \uparrow (\mathbf{O}_k) - \hat{\mathbf{I}}_{k-1}) \tag{10}$$

$$\hat{\mathbf{I}}_k = (1 - \lambda c)\hat{\mathbf{I}}_{k-1} + \lambda c S \uparrow (\mathbf{O}_k) \tag{11}$$

where $S \uparrow ()$ is the up-sizing function for the kth observation $\mathbf{O}_k$ of the estimated enhanced image $\hat{\mathbf{I}}_k$ and λ is a weighting constant and c is the above mentioned normalised cross correlation result. The constant λ is set so that, when combined with the correlation coefficient, the update weighting (λc) is around 0.15–0.25. The correlation result is a scalar between 0.0 and 1.0, correlations of contributing frames are around 0.6–0.9 so λ is set to 0.25. The aim of this weighting scheme is to allow better estimates, particularly later in the sequence as the sign gets larger, to have a greater impact on the result. To recover text on a sign, we know the expected image has a smooth background and lettering with sharp edges. Thus, a suitable prior/penalty function is one that minimises local smoothness of intensity but discounts penalties for large steps in intensity. To incorporate the text prior into the real-time implementation we pre-emphasise the up-sampled images before they are integrated by Eq. (11). We perform a contrast enhancing homogeneous point operator and erosion on the grey images which sharpens the discontinuity between the foreground and background and also reduces the spread of the text (skeltonises). These steps sharpen the textual boundaries and to some extent compensate for the over-smoothing of the filter.

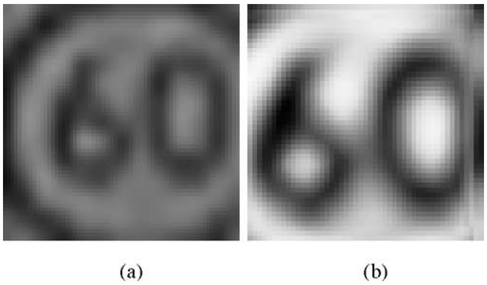


Fig. 10. (a) Resized final image in original sequence. (b) Enhanced image.

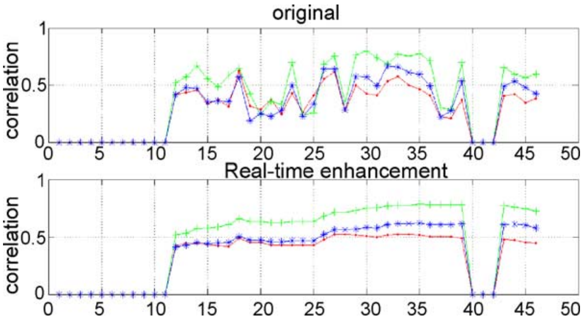


Fig. 11. Normalised cross correlation coefficient for ‘60’ sequence with ‘40’ (.), ‘60’ (+) and ‘80’ (∗) templates. Top: original resized image sequences. Bottom: enhanced image sequences. The drop outs between 40 and 43 are caused by failures in the sign detection phase.

5.2. Results and verification

The enhanced image is an improvement from the original up-sized image, as shown in Fig. 10. The efficacy of the original and enhanced image in speed classification was then tested by correlating a template image of ‘40’, ‘60’ and ‘80’ signs with the images. The template images consisted of only the number on the sign, not the bounding circle. Fig. 11 shows the correlation results for a ‘60’ sign sequence. In all cases trialled the enhanced image sequence showed a consistent improvement in reliability over the original image. Both the original and enhanced sequences show the expected upward trend in correlation value over time as the sign becomes larger. The relative differences between the templates and the consistency over time justify the expectation of better classification. The correlation drops slightly as the sign approaches the edge of the field of view as motion blur introduces repeated strongly non-Gaussian noise in the observations.

To verify our enhancement technique we implemented a recent super-resolution algorithm based on global optimisation and compared the results. The method used was the MAP algorithm used by Capel and Zisserman [4]. In this method a penalty function is used to influence the result based on the prior $p(\mathbf{I})$. A suitable text prior/penalty is implemented as a function of the gradient magnitude of the image. For small gradient magnitudes the penalty is a quadratic ($f(I'^2)$); as the gradient magnitude increases and crosses the threshold α the penalty has linear ‘tails’ ($f(|I'|)$). Our implementation used the Matlab *fmincon()* function with a scalar error composed of the sum of the squared differences plus the weighting (λ) of the penalty contribution. Best results were obtained with $\lambda = 0.025$ and $\alpha = 40$. Please refer to [4] for a full description of the implementation. Fig. 12 shows the result of the minimisation. The off-line image has more consistent intensity within the foreground and background regions but has lost some contrast overall. While an improvement on the temporal mean image is achieved, the tuning of λ and α that would provide a significant benefit across all the test image sequences proved difficult. The real-time result seemed similar to the off-line technique though exhibited more artifacts from the latter end of the image sequence, which is to be expected with the incremental update approach.

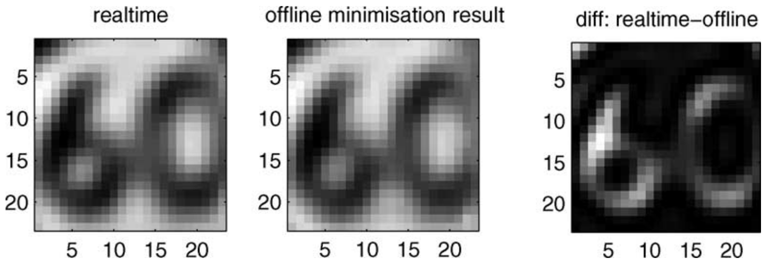


Fig. 12. Result from off-line non-linear minimisation.

6. Conclusion

This paper has presented a context sensitive driver assistance system. By not only monitoring the driver's actions, but also the driver's observations we were able to infer whether the driver was likely to have seen a sign. If an important sign was detected the system checked whether (a) the driver had looked at the sign, and (b) whether the state of the vehicle was compliant with the sign. If the driver has not seen the sign, and the car's state is not compliant with the sign, the driver can be informed with high priority. If, however, the driver appears to be aware of the sign, or the vehicle is not in an incorrect state, the information can be made available to the driver in a more passive manner. Automatic sign classification was improved significantly by online image enhancement of the sequences of approaching signs.

Acknowledgements

The support of the STINT foundation through the KTH-ANU grant IG2001-2011-03 is gratefully acknowledged, as is the support of the Centre for Accident Research and Road Safety, Queensland (CARRS-Q). National ICT Australia is funded by the Australian Department of Communications, Information Technology and the Arts and the Australian Research Council through *Backing Australia's ability* and the ICT Centre of Excellence Program.

References

[1] N. Apostoloff, A. Zelinsky, Vision in and out of vehicles: integrated driver and road scene monitoring, Int. J. Robotics Res. (2003) 5–28.

[2] S. Baker, T. Kanade, Limits on super-resolution and how to break them, in: Proceedings of the International Conference on Pattern Recognition, 2000, pp. 372–379.

[3] N. Barnes, A. Zelinsky, Real-time radial symmetry for speed sign detection, in: Proceedings of the IEEE Intelligent Vehicles Symposium, 2004.

[4] D. Capel, A. Zisserman, Super-resolution enhancement of text image sequences, in: Proceedings of the International Conference on Pattern Recognition, 2000, pp. 600–605.

[5] J.D. Crisman, C.E. Thorpe, SCARF: A color vision system that tracks roads and intersections, IEEE Trans. Robotics Autom. 9 (1) (1993).

[6] F. Dellaert, C. Thorpe, S. Thrun, Super-resolved texture tracking of planar surface patches, in: Proceedings of the IEEE International Conference on Intelligent Robotic Systems, 1998, pp. 197–203.

[7] E.D. Dickmanns, V. Graefe, Applications of dynamic monocular machine vision, Mach. Vision Appl. 1 (4) (1988) 241–261.

[8] E.D. Dickmanns, V. Graefe, Dynamic monocular machine vision, Mach. Vision Appl. 1 (4) (1988) 223–240.

[9] C.Y. Fang, C.S. Fuh, S.W. Chen, P.S. Yen, A road sign recognition system based on dynamic visual model, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2003, pp. 750–755.

[10] D.A. Gordon, Perceptual basis of vehicular guidance, Public Roads 34 (3) (1966) 53–68.

[11] N.L. Haworth, T.J. Triggs, 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.

[12] S.-H. Hsu, C.-L. Huang, Road sign detection and recognition using matching pursuit method, Image Vision Comput. 19 (2001) 119–129.

[13] T. Jochem, D. Pomerleau, No Hands Across America Official Press Release, Carnegie Mellon University, 1995.

[14] B. Johansson. Road sign recognition from a moving vehicle. M. Phil. Thesis, Centre for Image Analysis, Sweedish University of Agricultural Sciences, 2002.

[15] L. Fletcher, L. Petersson, A. Zelinsky, Driver assistance systems based on vision in and out of vehicles, in: Proceedings of the IEEE Symposium on Intelligent Vehicles, 2003.

[16] M. Land, D. Lee, Where we look when we steer, Nature (1994) 742–744.

[17] G. Loy, N. Barnes, Fast shape-based road sign detection for a driver assistance system, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai, Japan, 2004.

[18] G. Loy, A. Zelinsky, Fast radial symmetry for detecting points of interest, IEEE Trans. Pattern Anal. Mach. Intel. 25 (8) (2003) 959–973.

[19] J. Miura, T. Kanda, Y. Shirai, An active vision system for real-time traffic sign recogntition, in: Proceedings of the 2000 IEEE International Conference on Vehicles Symposium, 2002, pp. 52–57.

[20] P. Paclik, J. Novovicova, P. Somol, P. Pudil, Road sign classification using the Laplace Kernel classifier, Pattern Recogn. Lett. 21 (2000) 1165–1173.

[21] L. Petersson, L. Fletcher, N. Barnes, A. Zelinsky, An interactive driver assistance system monitoring the scene in and out of the vehicle, in: Proceedings of the IEEE International Conference on Robotics and Automation, 2004, pp. 3475–3481.

[22] G. Piccioli, E.De. Micheli, P. Parodi, M. Campani, Robust method for road sign detection and recognition, Image Vision Comput. 14 (3) (1996) 209–223.

[23] L. Priese, J. Klieber, R. Lakmann, V. Rehrmann, R. Schian, New results on traffic sign recognition, in: Proceedings of the Intelligent Vehicles Symposium, IEEE Press, Paris, 1994, pp. 249–254.

[24] Seeing Machines, FaceLAB face and eye tracking system, http://www.seeingmachines.com, 2004.

[25] D.G. Shaposhnikov, L.N. Podladchikova, A.V. Golovan, N.A. Shevtsova, A road sign recognition system based on dynamic visual model, in: Proceedings of the 15th International Conference on Vision Interface, 2002.

[26] T. Ishikawa, S. Baker, I. Matthews, T. Kanade, Passive driver gaze tracking with active appearance models, in: Proceedings of the 11th World Congress on Intelligent Transportation Systems, 2004.

[27] K. Takemura, J. Ido, Y. Matsumoto, T. Ogasawara, Driver monitoring system based on non-contact measurement system of driver's focus of visual attention, in: Proceedings of the IEEE Symposium on Intelligent Vehicles, 2003, pp. 581–586.

[28] R.Y. Tsai, T.S. Huang, Multiframe image restoration and registration, Adv. Comput. Vision Registration 1 (1984) 317–339.

[29] B.A. Wandell, Foundations of Vision, Sinauer Associates, Sunderland, Mass, USA, 1995.

Luke Fletcher is completing a PhD at the Department of Information Engineering, Research School of Information Sciences and Engineering, Australian National University. In 1996 Luke completed BE (Hons.), BSc degrees at the University of Melbourne. His research interests include: robust & real-time computer vision, Driver Assistance Systems, perception and visual interfaces.

Gareth Loy received his PhD in robotics from the Australian National University (2004) for his research on computer vision for human computer interaction. Prior to this he completed his BE in systems engineering (1999) and BSc mathematics (1997) at the Australian National University. During his PhD, he spent several months on complementary research projects at the University of Western Australia and the Humanoid Interaction Lab at AIST, Japan, undertook consulting work as a research scientist for Seeing Machines, and wrote an undergraduate lecture course in computer vision. In 2003 he joined the Royal Institute of Technology (KTH) in Stockholm, Sweden, as a post-doctoral researcher where he continues to pursue his interests in human tracking, feature detection and real-time computer vision algorithms.

Nick Barnes received his BSc and PhD from the Department of Computer Science and Software Engineering at the University of Melbourne. Since 2003 he has been a researcher with the Autonomous Systems and Sensing Technologies Programme at National ICT Australia. Prior to this he was a lecturer in the Department of Computer Science and Software Engineering at the University of Melbourne for more than three years, as well as spending six months as a post-doctoral fellow with the LIRA-Lab at the University of Genoa, Italy. He also spent time with Accenture, working in the telecommunications industry. His research interests focus on perception and action, principally in developing effective perceptual systems to facilitate the action of robots, as well as for diagnosis with medical image analysis. He also has a strong interest in general robotics, and perception in biology.

Dr. Alex Zelinsky is Director of the Information and Communication Technologies (ICT) Centre within the Commonwealth Scientific Industrial Research Organisation (CSIRO) in Australia (www.ict.csiro.au). Prior to joining CSIRO Dr. Zelinsky Professor of Systems Engineering at the Australian National University (1996–2004). He received his Bachelor (1983) and PhD (1991) degrees in computer science and electrical engineering from the University of Wollongong, Australia. Dr. Zelinsky was CEO and co-founder of Seeing Machines Pty Ltd (2000–2004) (www.seeingmachines.com). Seeing Machines is a world-leader in vision-based human-machine interfaces. Dr. Zelinsky is widely published in the robotics and computer vision fields. His areas of research include; mobile robotics, human-machine interaction, intelligent transportation systems and real-time computer vision systems.