Steering without Representation using Active Fixation

D W Murray, I D Reid and A J Davison
Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

Abstract. This paper demonstrates the use of active fixation on both fixed and moving fixation points to guide a robot vehicle using a steering rule which, at large distances, sets the steering angle directly proportional to the deviation of gaze direction from translation direction. Steering a motor vehicle around a winding but otherwise uncluttered road has been observed by Land and Lee (1994) to involve repeated periods of visual fixation upon the tangent point of the inside of each bend. We suggest that proportional rule devised for steering in the robotic example appears applicable to the observed human performance data, providing an alternative explanation to the quadratic rule proposed by Land and Lee.

1 Introduction

In active machine vision (Bajcsy and Allen 1984; Bajcsy 1988; Ballard 1991; Aloimonos *et al* 1988), visual feedback is used to control not only the physical parameters of a camera or cameras — most importantly their direction of gaze or focus of attention — but also how the resulting imagery is processed from frame to frame. The aim is to construct a set of visual behaviours in which sensing and perception are tightly coupled to specific robotic actions, and then to embed these in a framework which allows constructive interaction (for example, Kŏsecká *et al* 1994).

In recent papers, we have described the design and realization of a highly agile electromechanical camera platform and a visuo-control system for active vision research (Sharkey *et al* 1993; Sharkey and Murray 1996) and have demonstrated a number of reactive and purposive visual behaviours running on it. The reactive behaviours have included rapid saccadic redirections of the cameras in response to unexpected optical flow — both "capture" saccades when the resultant projected image motion is largely translational, and "panic" saccades when the image motion is predominantly divergent (Murray *et al* 1995). The more purposive responses include monocular smooth pursuit using optical flow (Bradshaw *et al* 1994; Bradshaw *et al* 1997), and monocular and stereoscopic tracking and structure recovery using feature clusters (Reid and Murray 1996; Fairley *et al* 1995). Our work is carried out in an engineering context, but a question which is often raised is whether such active heads might prove useful adjuncts to the computer alone in computational studies of biological visual

behaviours.

In this paper we give an illustrative example where the use of active cameras in machine vision might bear upon natural measurements. Active fixation is used to steer a robot vehicle around an obstacle, using a steering rule derived from the angle between the instantaneous directions of gaze and of translation. For relatively large distances, the rule is one of direct proportionality between the angle of steering and the angle between the directions of gaze and instantaneous translation.

The example is illustrative in several ways of the merits of the active vision paradigm. First, it provides a clear-cut demonstration of the way that active fixation can obviate the need for relatively costly visual processing — here we need only perform a simple correspondence at the image centre instead of having to compute visual motion. The example also indicates how the loss in visual richness is compensated for by proprioceptive data from encoders on the camera platform. Furthermore a natural example of the same visual behaviour, though not necessarily the same mechanism, has been observed by Land and Lee (1994). They measured the relationship between a human driver's gaze direction and steering response while negotiating a twisting road, and found that the task involved repeated periods of fixation along tangents to the inside kerb, and that the angle between the vehicle's heading and the direction of gaze was highly correlated with the steering response. We return to their data in discussion.

2 Representation-free steering of a robot around a near point

A widely-used strategy in robot navigation using visual or sonar sensors is to sense the 3D environment to establish free, navigable, space and then to choose a safe "middle way" through the space using, for example, potential field methods (for example, Cameron and Probert 1994). Augmenting this pre-planned navigation module are reactive routines which devise local excursions around initially undetected obstacles.

The need to establish and update dynamically such a 3D map of the environment arises when no attempt is made to determine which part of the scene geometry is important for task in hand. In many situations, however, the obvious focus points are points of a range discontinuity in the navigable region, which occur, for example, at the inside of the "next bend", as shown in Figure 1(a). Such points need not necessarily be detected within a 3D range map, as they also appear in the image as discontinuities in visual motion and disparity.

Our first aim then is to direct a robot vehicle equipped with an active camera towards and around a focus point at some safe distance R without representing the robot's surroundings.

As sketched in Figure 1(b), the robot vehicle has a controllable camera platform mounted above the front unsteered wheel, and is able to use a closed loop vision process to maintain its direction of gaze $\mathbf{g}$ towards O. Odometers on the platform allow the determination of the angle between this vector and the back-front axis of the vehicle. Odometers on the vehicle's wheels and steering column enable the instantaneous translational velocity $\mathbf{v}$ of the head platform to be determined, again referred to the front-back axis of the vehicle. Depending on the vehicle's kinematics and the position of the camera platform on the vehicle, $\mathbf{v}$ will not point directly forward when the vehicle is turning. However, on our particular vehicle the platform is mounted directly above the centre of the unsteered and undriven axle which is the centre of rotation. In this case $\mathbf{v}$ does point along the vehicle axis and has magnitude $v = V \cos s$, where V is the steered wheel's axis speed and s the steering angle.

We define the gaze angle to be the angle $\theta(t)$ between $\mathbf{v}$ and $\mathbf{g}$ at any instant. To steer the vehicle towards an orbit of radius R around O then requires a change in the direction of translational velocity of

$$h(t) = \theta(t) - \sin^{-1}\left(R/D(t)\right)$$

where D is the distance to the obstacle. To effect this in a gradual way, a proportional steering demand derived from $h(t)$

$$s(t) = \kappa h(t) \tag{1}$$

is sent to the vehicle controller. The gain κ used in both live and simulated experiments is 0.5.

When the robot is close to the point, two methods of recovering D are possible. The first uses the angle of convergence of the stereo head. The gaze controller is programmed to provide symmetric convergence, $\theta_R = \theta_L$ and then $D = (I/2)\cot\theta_{\mathrm{L}}$, where I is inter-camera separation. The second method is achievable monocularly and is the fixated analogue of optical flow. It utilizes the ratio of rate of change of gaze angle to speed, $D = (v/\dot{\theta})\sin\theta$. Note that further odometry is required from the vehicle because its rotational motion must be derived to recover $\dot{\theta}$ relative to the vehicle's body, and because the magnitude of the translational velocity is required. Although this method is robust enough when using a camera moved by an accurately calibrated robot arm, we found the relatively low quality of vehicle odometry rendered the monocular method inaccurate.

As the simulation of Figure 2 shows, the rule's effect is to steer the vehicle into an circular orbit around the fixated point, in a counter-clockwise sense if $R > 0$ and a clockwise sense if $R < 0$. The radius of the orbit will be greater than $|R|$ is the distance moved by the vehicle

is other than infinitesimal between samples. In the practical realization therefore we prefer to detect the time at which the steering demand increases rapidly — that is, the time at which D approaches $|R|$ and θ approaches $+90°$ for $R > 0$ or $-90°$ for $R < 0$ — and to switch to an open-loop steering manoeuvre. For our vehicle this requires $s = \tan^{-1}(L/R)$ where L is the distance between the front axle and the rear steered wheel.

2.1 A practical realization

The electromechanical stereo head used in the work is mounted at the front of a motorized and steerable vehicle (Figure 3(a)). The head alters the directions of gaze of its two cameras using four degrees of rotational freedom: a central pan (P) or neck axis, left (L) and right (R) vergence axes, and an elevation axis (E). Each axis is driven by a DC servo-motor fitted with an harmonic drive gearbox, giving minimal backlash, and is capable of accelerations in excess of $20,000°\text{s}^{-2}$ and smooth tracking speeds ranging down from $400°\text{s}^{-1}$ to $0.03°\text{s}^{-1}$. An encoder attached to the motor side of the gearbox feeds back information to the servo-controller at several hundred Hz, allowing precise control of position and velocity even without using visual feedback.

For navigation in the built environment, a commonly available feature to fixate upon is a vertical edge. We use a template edge which, once initialized, searches perpendicular to its direction to find the new position of the edge as illustrated in Figure 3(b). Its angular displacement from the image centre is sent to the head controller which in turn powers the head's P, L and R axis motors to re-centre the feature. The kinematic redundancy between P, L and R is eliminated in this work by requiring symmetric convergence ($\theta_\text{L} = \theta_\text{R}$ in Figure 1). As the vehicle moves, the resulting time-varying angle θ between the cyclopean direction of gaze $\mathbf{g}$ of the head and the instantaneous translational velocity $\mathbf{v}$ of the vehicle is derived from odometric information from encoders on the head axes and used to derive the steering commands as described below. The vision and control computations are performed on transputers which communicate directly with the vehicle's servo-controller and communicate via a PC with the head's servo-controller.

Figure 4 shows four frames cut from a video of the robot steering around a vertical pole. The safe radius was set to be $R = +1\text{m}$, resulting in counter-clockwise rotation. The stereo head can be seen fixating on the pole, maintaining symmetrical convergence. The direction of gaze approaches perpendicular to the direction of motion as the vehicle steers into an orbit

3 Moving and distant fixation points

Because the gaze direction is continually updated, the method is immediately applicable to a moving fixation point. An obvious application here is to road following, where the fixation moves along the road ahead of the vehicle. Raviv and Herman (1993) suggested using the tangent to the road edge as a fixation point as it slips along ahead of the moving vehicle, and fixation on the tangent has also been found to be important for human steering (Land and Lee 1994). Raviv and Herman however proposed a steering rule based based on analysis of image motion, using it to recover changes in steering direction. Here we propose using fixation more directly, removing the need to compute image motion and using gaze direction to generate an absolute steering demand. Raviv and Herman also proposed that the camera direction changes be locked to the steering direction, a constraint which seems wholly unnecessary.

Figure 5 shows a twisting track. As the vehicle moves clockwise around the circuit, the fixation point is constantly updated to be the most distant visible tangent point on either side of the track. Visible here means that the line of sight $\mathbf{g}$ to a tangent must not intersect a nearer track edge. Every 10th update, we have indicated the fixation point ($\times$) and, as expected, more time is spent fixating on points of higher curvature. Note, for example at the top right of the figure, that corners are 'cut' in a way that would not occur with a kerb- or centre-hugging algorithm.

An apparent difficulty in applying the "correction" term $q = \sin^{-1}(R/D)$ is the visual determination of D when D is large. However, when $D \gg |R|$, where $|R|$ is now desired minimum distance from the kerb, q can be neglected. We illustrate the effect of various piecewise approximations to q in Figure 6. The simulations model a vehicle travelling at 50 kph along a 4 m wide road with bend with radii of curvature decreasing from 40 m to 10 m. The three sets use the (i) the full correction term, (ii) a piecewise approximation

$$q = \begin{cases} \sin^{-1}(R/D) & \text{for } 0.20 < R/D \\ \sin^{-1}(0.1) & \text{for } 0.07 < R/D < 0.20 \\ 0 & \text{otherwise,} \end{cases}$$

and (iii) no correction ($q = 0$), and within each set a range of delays in feedback is considered. The gain is 0.5 and the sampling rate is 25 Hz. Instability can of course always be removed by slowing down: reducing the gain κ usually results in disaster from understeer.

3.1 A robot realization

We have implemented the track-following behaviour using our robot vehicle. To simplify visual processing (which is not the issue here) a winding white road was laid out on the floor

of the laboratory, from which strong edges were detected and linked. Tangent points on the road's kerbs were determined (appearing as vertical edge sections in our images since the head was moved in such a way as not to introduce cyclotorsion) and sorted to find the most distant visible tangent. A demand was then sent to the active head to fixate the chosen point by moving the pan axis. The elevation angle was also controlled such that the fixated point always appeared near to the bottom of image — this ensured that maximum "look-ahead" was available to detect new tangent points appearing in the distance. Figure 7 shows examples of tangent points detected and centrally fixated by the robot head.

Once tracking of the first tangent point was established, the vehicle was set in motion at a constant velocity. Its steering angle s was then controlled using the rule in Equation 1. Image processing and updates of the robot and head parameters were carried out at about 1Hz on a 100 MHz Pentium (the edge detection process taking up the majority of processing time). Transfer of the fixation point was automatically carried out when motion of the vehicle made visible a new most distant tangent point.

A selection of still pictures cut from a video of the vehicle following the road is shown in Figure 8. It should be noted that the constraint of the small size of our laboratory meant that the bends were relatively tight in this experiment, and performance was improved when the correction term $q = \sin^{-1}(R/D)$ was incorporated in the vehicle control equation.

4 A connection with human performance?

Figure 9 reproduces a portion of the data on steering angle response and direction of gaze accumulated by Land and Lee (1994) whilst observing a driver steering a vehicle around a winding one-way road. The steering angle response has been advanced in time by 1s to compensate for the processing delay (a figure found by cross-correlation) and scaled down uniformly by a factor of about 3.

To see whether the steering rule used here could account for the data, we first consider whether the tangent point is sufficiently far ahead for the $\sin^{-1}(R/D)$ term to be negligible. From the data in Figure 9, the maximum $|\dot{\theta}| \approx 15°\mathrm{s}^{-1}$ when $\theta \approx 20°$ and $v = 12.5\mathrm{ms}^{-1}$, giving $D \approx 17$m. If we estimate R as 1 m, then $\sin^{-1}(R/D)$ is only some $3°$ — small compared with the gaze angle of $\theta = 20°$. These numbers are compatible with more recent observations of Land and Horwood (1995) that indicate that drivers perform best when the fixation distance D is around 1 s ahead of the vehicle. Thus if we were to use the steering law proposed here to "explain" the driving conditions explored by Land and Lee (1994), and indeed by Land and Horwood (1995), we might expect little deviation from linear proportionality between the instantaneous steering response and gaze angle.

Land and Lee gave a geometrical discussion of their observations in terms of the relationship between average curvature C of the road between the vehicle and the fixated point, and the angle θ between gaze and heading directions

$$C = \left(\frac{1}{R\cos\theta} - \frac{1}{R} \right) , \tag{2}$$

where R is the distance between vehicle and kerb. The implication is that the steering angle is set to a fixed value that would take the vehicle in an arc of constant curvature. The geometry is sketched in Figure 10, and it is clear that

$$s = \tan^{-1}\left(\frac{L}{r+R} \right)$$

and

$$\frac{1}{R+r} = \frac{1}{R}(1-\cos\theta) ,$$

so that, expanding in powers of θ

$$s = \frac{L}{2R}\theta^2 + O(\theta^4) + \ldots \quad .$$

This explanation models a quadratic dependence of s on θ, rather than a linear dependence used in our work.

Is there evidence supporting one or the other in the natural data? Figure 11(a) shows the variation of steering response s with gaze angle θ for the temporal sequence, together with the least-squares fits to linear and quadratic functions. Given the data scatter, especially for positive gaze angles, it is hard to be unequivocal about one or the other, although the linear fit one has a smaller χ^2 value.

There is however an asymmetry apparent in this data set, both in scatter and overall gain between data with positive and negative gaze angles. Plotting and fitting these separately, we find that the scattered positive data (Figure 11c) is marginally better fitted by the quadratic, but that less scattered data with negative gaze angles (b) is certainly best fitted by a straight line.

5 Conclusions and conjectures

We have demonstrated both in simulation and in real-time robotic experiments that a very simple rule linking fixation angle and steering angle can be used to guide an autonomous vehicle fitted with an active stereo head around a given obstacle. Provided proprioceptive information is available from the head, visual processing is reduced to a search (in our case

in only one dimension) around the centre of the image in order to maintain fixation. We have shown in simulation and experiment that the same steering rule would guide a vehicle along a road, were the fixation point to transfer from tangent point to tangent point as the road ahead unfolded.

The linear dependence of steering response versus gaze angle provides an alternative plausible explanation of the response observed in human drivers by Land and Lee (1994), who implicitly proposed a quadratic relationship to account for the data. In terms of geometry and data-fitting, the rule proposed here certainly provides a simpler explanation, and one which avoids the need for the driver to estimate somehow the average curvature of the road ahead to the next tangent point. Furthermore, proportionality of response to error signal is attractive when viewed from the perspective of feedback control. We stress however that from robotic experiments we can conclude nothing about the correctness of either explanation. Indeed, the observation of Land and Horwood (1995) that drivers also utilize information from the near road edge is likely to confound any decision between models.

It would perhaps be of interest to measure human driver gaze angles and responses on sharper bends where the term $\sin^{-1}(R/D)$ is substantially greater than 0.1 — perhaps particularly on hairpin bends where the tangent point is obscured from the driver's view by the vehicle's body. In the latter case, our experience, albeit unquantified, is that the eye and head are drawn towards the stationary centre of curvature. One drives by looking out the side window at a fixed point with gaze angles approaching 90°, as does the robot vehicle when steering around a stationary fixation point.

A second issue that might warrant further investigation is that of asymmetric gain (Figure 11(b,c)) which appeared despite the road's being one-way. Would drivers accustomed to driving in opposite handed vehicles on the opposite side of the road perform differently?

References

Bajcsy R (1988) "Active perception" *Proc IEEE* **76** 996–1005

Bajcsy R and Allen P (1984) "Sensing Strategies" *Proc US-France Robotics Workshop, Philadelphia, PA* 96–102

Ballard D H (1991) "Animate vision" *Artificial Intelligence* **48** 57–86

Bradshaw K J, McLauchlan P F, Reid I D, and Murray D W (1994) "Saccade and pursuit on an active head/eye platform" *Image and Vision Computing* **12**(3) 155–163

Bradshaw K J, Reid I D, and Murray D W (1997) "The active recovery of 3D motion trajectories and their use in prediction" *IEEE Trans on Pattern Analysis and Machine Intelligence* **19**(3) 219–234

Cameron S and Probert P, editors (1994) *Advanced Guided Vehicles*. (World Scientific:Singapore)

Fairley S M, Reid I D and Murray D W (1995) "Transfer of fixation for an active stereo platform via affine structure recovery" *Proc 5th Int Conf on Computer Vision, Boston MA* 1100–1105 (IEEE Computer Society Press:Los Alamitos CA)

Aloimonos J, Weiss I and Bandyopadhyay A (1988) "Active vision" *International Journal of Computer*

Vision **1**(4) 333–356

Kŏsecká J, Bajcsy R and Mintz M (1994) "Control of visually guided behaviours" In *Real-time Computer Vision* (Cambridge University Press: Cambridge UK)

Land M F and Horwood J (1995) "Which parts of the road guide steering?" *Nature* **377** 339–340

Land M F and Lee D N (1994) "Where look when we steer" *Nature* **369** 742–744,

Murray D W, Bradshaw K J, McLauchlan P F, Reid I D, and Sharkey P M (1995) "Driving saccade to pursuit using image motion" *International Journal of Computer Vision* **16** 205–228

Raviv D and Herman M (1993) "Visual servoing from 2-d image cues" In *Active Perception* ed Aloimonos Y (Lawrence Erlbaum Associates: New Jersey)

Reid I D and Murray D W (1996) "Active tracking of foveated feature clusters using affine structure" *International Journal of Computer Vision* **18**(1) 1–20

Sharkey P M, Murray D W, Vandevelde S, Reid I D, and McLauchlan P F (1993) "A modular head/eye platform for real-time reactive vision" *Mechatronics* **3**(4) 517–535

Sharkey P M and Murray D W (1996) "Delays vs performance of visually guided systems" *IEE Proceedings on Control Theory Applications* **143**(5) 436–447

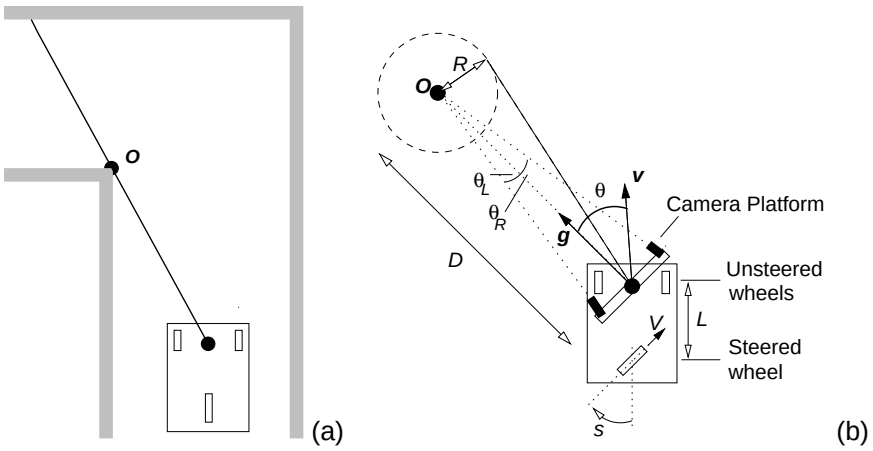


Figure 1: The scene, robot and head geometry viewed from above. (a) Focus points occur where the range within the navigable region is discontinuous. (b) The active head fixates the point about which the robot is required to steer at a safe radius of R.

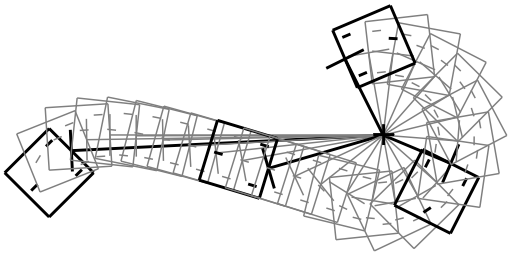


Figure 2: A simulation of the effect of the simple steering rule derived in the text.


(a)

(b)

Figure 3: (a) The stereo robot head used in this work. (b) The extended vertical edge detector. A search is made for the edge in the new image by examining pixel values either side of the old position and perpendicular to the old orientation.

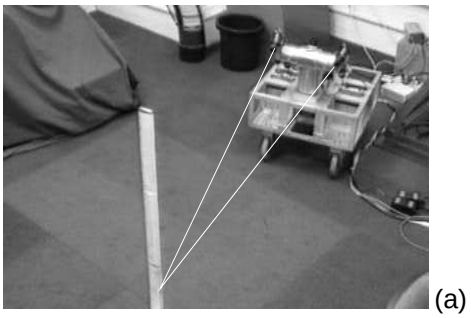
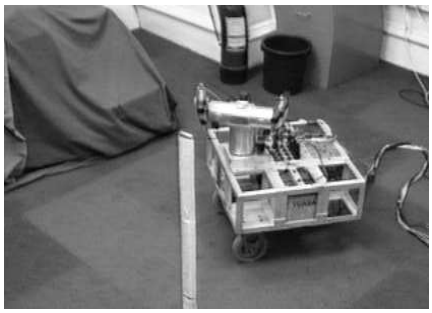
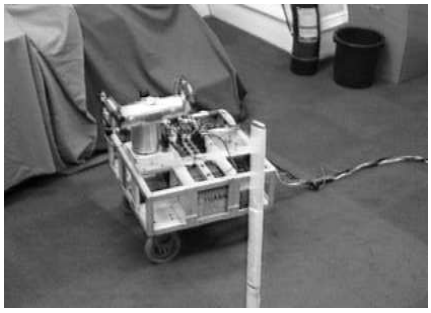
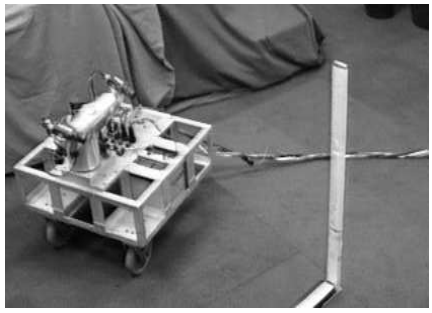

(a)

(b)

(c)

(d)

Figure 4: The stereo head fixates on the vertical pole maintaining symmetrical convergence, and the vehicle steers into an orbit, here an counter-clockwise one. The complete sequence in MPEG format can be obtained from the World Wide Web at http://www.robots.ox.ac.uk/∼lav/.

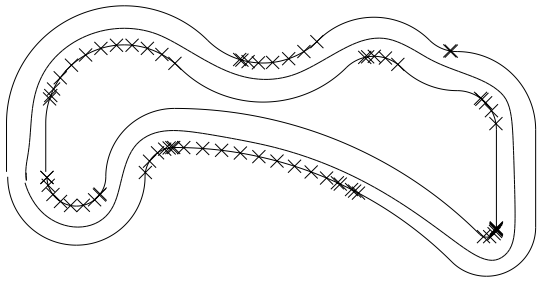


Figure 5: A simulation of our control law driving a vehicle clockwise around a track with high curvatures, using fixation on the furthest visible tangent point. The crosses $\times$ show successive positions of the fixated tangent point as it moves ahead of the vehicle.

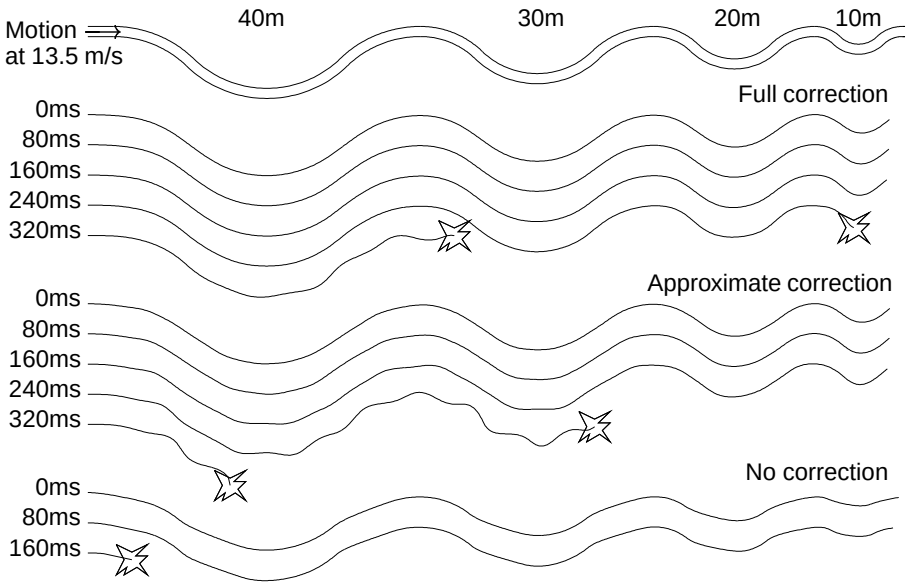


Figure 6: The onset of instability for a range of feedback delay for a particular road comprising chicanes with radii of curvature decreasing from 40m to 10m. The simulation stops when the vehicles hits the road sides.

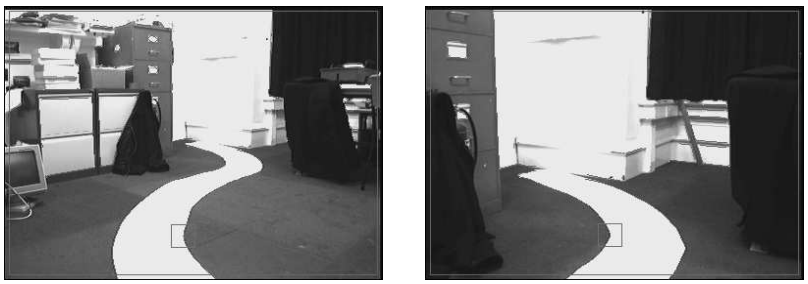


Figure 7: Two examples of the most distant visible tangent point detected and fixated in views from the robot's camera.

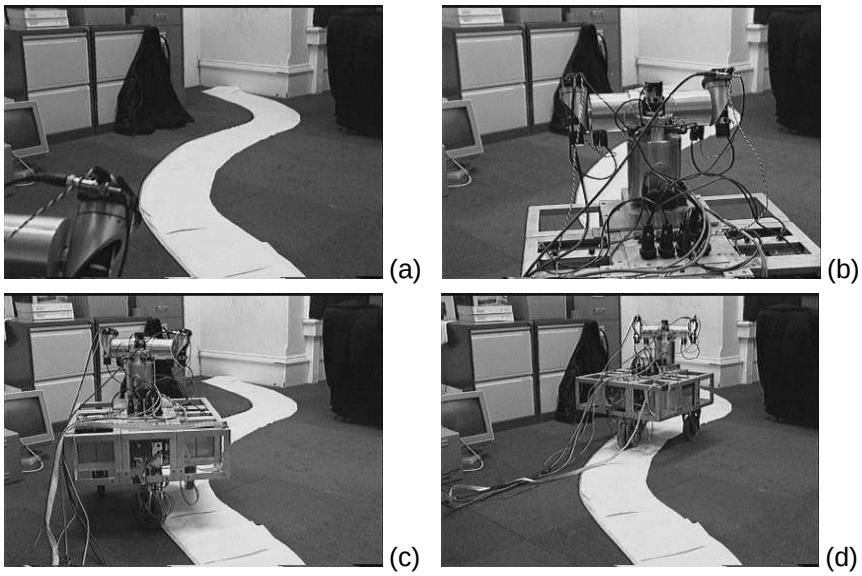


Figure 8: A series of stills cut from a video sequence of the robot vehicle following a track in the laboratory. The complete sequence is at http://www.robots.ox.ac.uk/∼lav/.

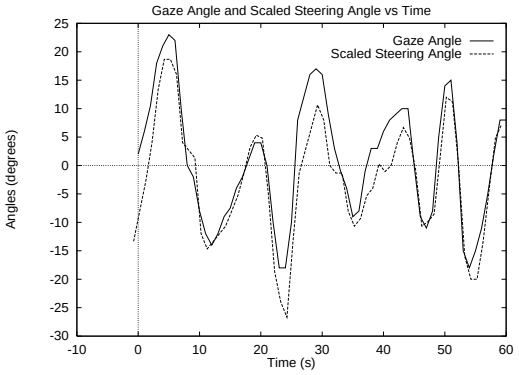


Figure 9: A temporal portion of the data of Land and Lee. The gaze angle is shown to scale, but the steering response has been advanced in time by 1s and scaled down by a factor of approximately 3.

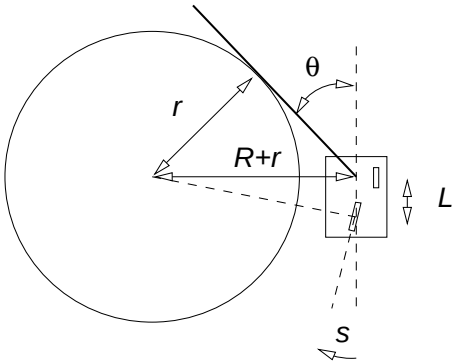


Figure 10: The geometry of turning a bend of constant curvature $C = 1/r$.

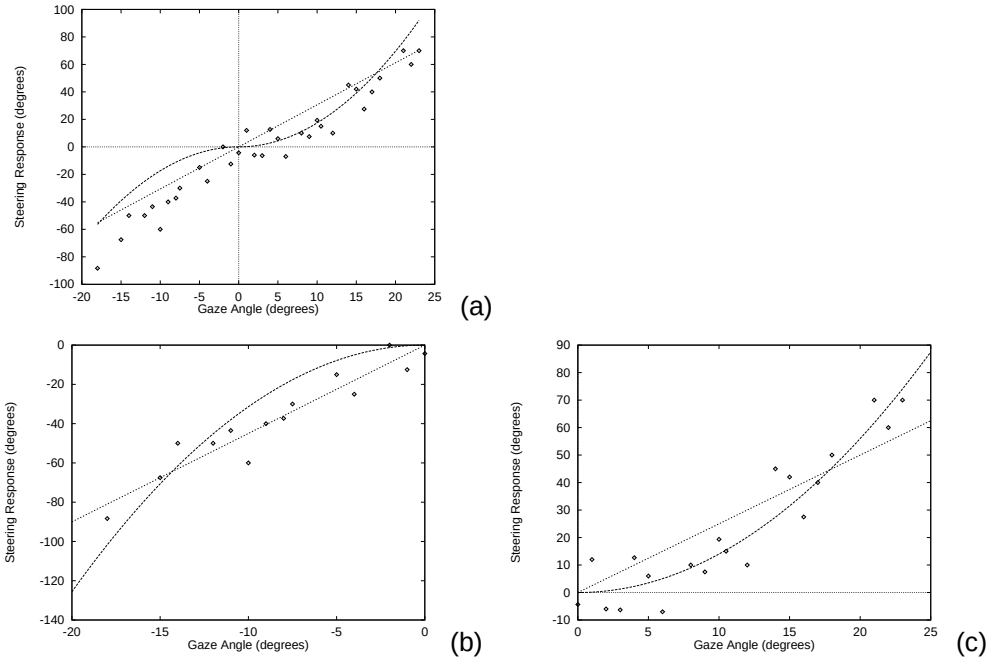


Figure 11: The steering response (not scaled) as a function of gaze angle with best linear and quadratic fi ts for (a) all, (b) negative and (c) positive gaze angles.