Steering and Navigation Behaviours using Fixation

David W. Murray, Ian D. Reid and Andrew J. Davison
Department of Engineering Science, University of Oxford,
Parks Road, Oxford, OX1 3PJ, U.K.
email: [dwm, ian, ajd]@robots.ox.ac.uk

Abstract

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 demonstrate a similar use of 'active' fixation in the autonomous navigation of a robot vehicle around an obstacle, and show how the control law devised for steering in the robotic example is applicable to the observed human performance data. We discuss the merits of fixation for mobile robot localization.

1 Introduction

In active machine vision [1, 2], 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 thence to enable them to interact (eg [4]).

In this paper we first describe a visual behaviour which steers a robot vehicle around an obstacle, using a rule derived from the angle between the direction of gaze and the direction of translation of the vehicle.

The example is illustrative for three reasons. First, it provides a clear-cut demonstration of the way that active fixation can obviate the need for relatively complicated visual processing — here we need only perform a simple correspondence at the image centre instead of having to compute optical flow. Secondly, it indicates how the loss in visual information is replaced by proprioceptive data from encoders on the head. Thirdly, a natural example of the same visual behaviour has been observed by Land and Lee [6], who measured the relationship between a human driver's gaze direction and steering response while negotiating a twisting road. Land and Lee found that driving around a bend involved repeated periods of fixation along tangents to the inside kerb, and that the angle between the heading and the direction of gaze was highly correlated with the steering response, the latter measured by the angle to which the steering wheel was turned. We return to their data in discussion.

In the penultimate section we describe progress on a second behaviour using active fixation to localize a vehicle.

2   Method

The electromechanical stereo head used in the work is mounted at the front of a motorized and steerable vehicle (Figure 1). 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 $20,000^\circ s^{-2}$ and smooth tracking speeds ranging down from $400^\circ s^{-1}$ to $0.03^\circ s^{-1}$. 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.
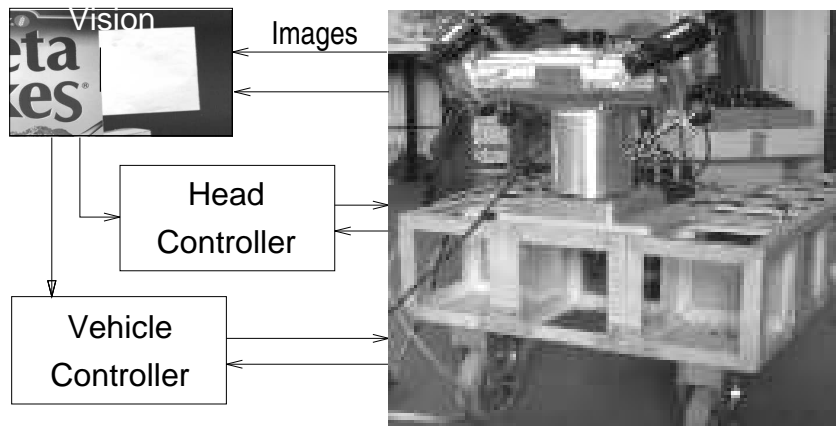

Figure 1: The apparatus and visuo-control scheme. The vision module comprise a vertical extended edge detector run independently in left and right image streams.

The method of steering uses fixation on a scene feature which the vehicle should approach then turn around. For indoor navigation, the feature most often encountered is a vertical edge (Figure 1). Every 40 ms as a new image is captured a feature detector running in each camera determines the angular displacement of the feature from the image centre and communicates it 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 ($\theta_L = \theta_R$ in Figure 2). 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.

Referring to Figure 2, to move around the fixated point O at some safe radius R requires a change in translational heading of

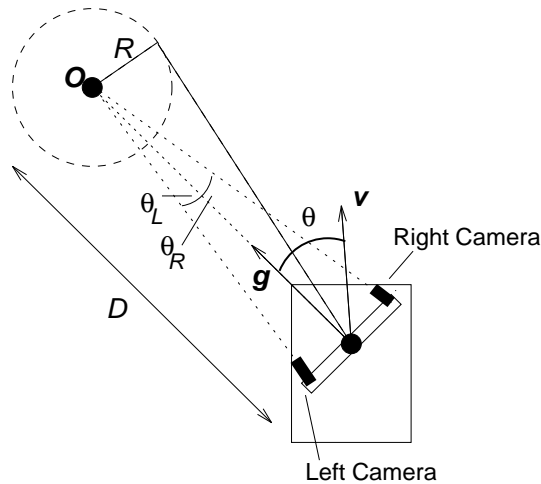$$h(t) = \theta(t) - \sin^{-1}\left(R/D(t)\right)$$

Figure 2: The scene, robot and head geometry viewed from above. The active head fixates O and the robot is required to steer into an orbit of radius R around the point.

where D is the distance to the obstacle. The sign of R determines the direction of turn about the fixated point: a counter-clockwise turn (viewed from above as in Figure 2) is generated if $R > 0$ and a clockwise turn if $R < 0$.

Two methods of recovering D are possible. The one implemented here uses the angle of convergence of the stereo head. Under symmetric convergence $D = (I/2) \cot \theta_{\mathrm{L}}$, where I is inter-camera separation. The second method utilizes the ratio of rate of change of angle to speed, $D = (v/\dot\theta) \sin \theta$, and is thus achievable monocularly (although further odometry is required from the vehicle because its rotational motion must be derived). Once $h(t)$ is found, a steering demand

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

is sent to the vehicle controller. The gain in both live and simulated experiments is $\kappa = 0.5$.

3   Results

As the simulation of Figure 3 shows, the rule's effect is to steer the vehicle into a circular orbit of radius $|R|$ around the fixated point. Here $R > 0$, and the orbit is counter-clockwise sense.

The same sense is chosen in the experiment shown in Figure 4, where several frames cut from a video show the robot steering around a vertical pole. 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.

Although the experiment demonstrates the steering rule applied to a fixed position, the rule is equally useful when the fixation point moves. This is illustrated in Figure 5, where the fixation point marked with an × is constantly updated as
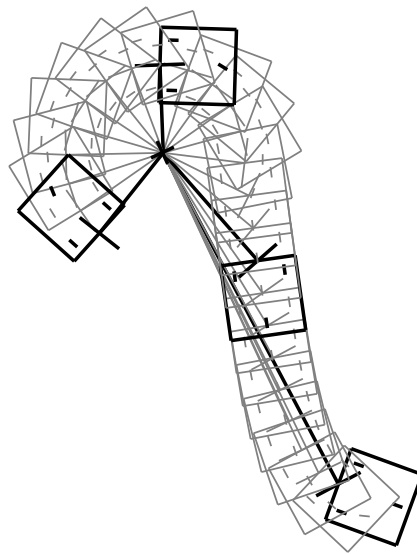
Figure 3: A simulation of the effect of the simple steering rule derived in the text.
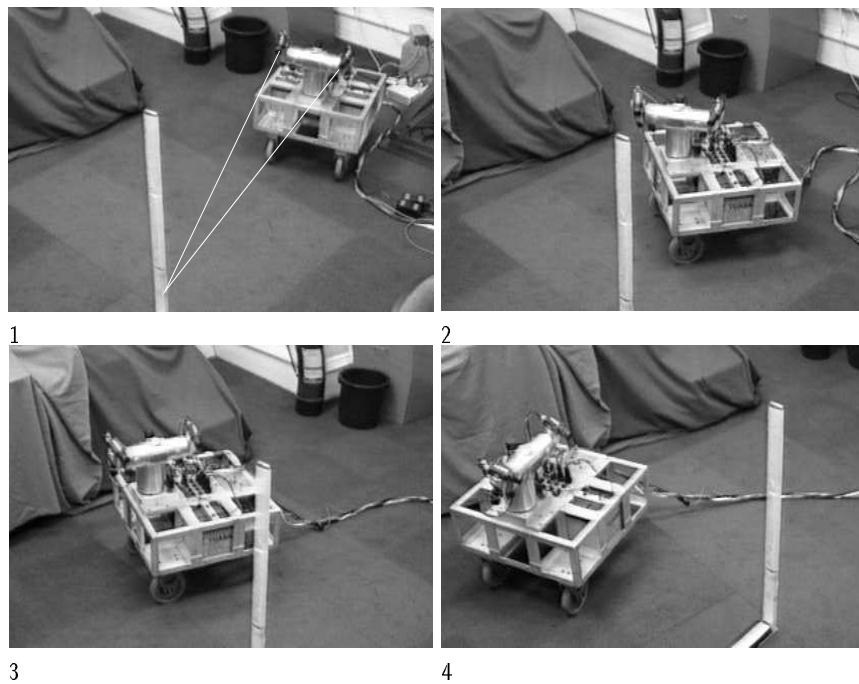

1

2

3

4

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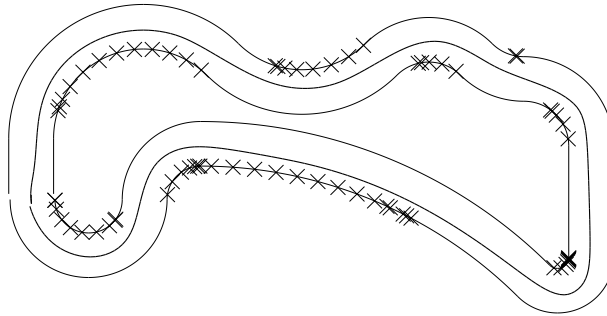
Figure 5: A simulation of the control law driving a vehicle clockwise around a track with relatively high curvatures, using fixation on the furthest visible tangent point. The crosses *times* show successive positions of the fixated tangent point as it moves ahead of the vehicle.

the vehicle moves to be the most distant visible tangent point, in the spirit of the observations of Land and Lee [6]. R, the safe radius, is now the desired distance from the kerb. The central line shows the track followed by the vehicle steering in this way. The top right of the figure shows how gentle corners are 'cut' in a way that would not occur with a kerb- or centre-hugging algorithm.

4   Discussion

Figure 6 reproduces a portion of the data accumulated by Land and Lee [6] whilst steering a vehicle around a winding one-way road. The steering angle response has been advanced in time by 0.75s, compensating for processing delay, and scaled down uniformly by a factor of about 3. Overall, the response is linearly proportional to the gaze angle, though we note an apparent asymmetry in the data for left- and right-handed turns.

They discussed 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,

$$C = 1/(R\cos\theta) - 1/R \; , \qquad (2)$$

where R is the distance between vehicle and kerb. The implication is that the steering angle s is set to a fixed value that would take the vehicle in an arc of constant curvature up to the tangent point. When θ and s are small (meaning a bend in the road is gentle), this model makes C proportional to θ^2. In the same limit, the curvature of the vehicle's trajectory C is proportional to steering input s. That is, s would be proportional to θ^2 — counter to the striking feature of their data in Figure 6 which is that s is linearly related to θ over extended periods.

It appears that the natural data are more easily reconciled with our scheme for robot navigation. If the curvature is always modest (unlike the course in Figure 5), the fixated tangent point will always be far enough ahead to make $\sin^{-1}(R/D)$ small. From the data in Figure 6, the maximum $|\dot\theta| \approx 15°\text{s}^{-1}$ when $\theta \approx 20°$ and
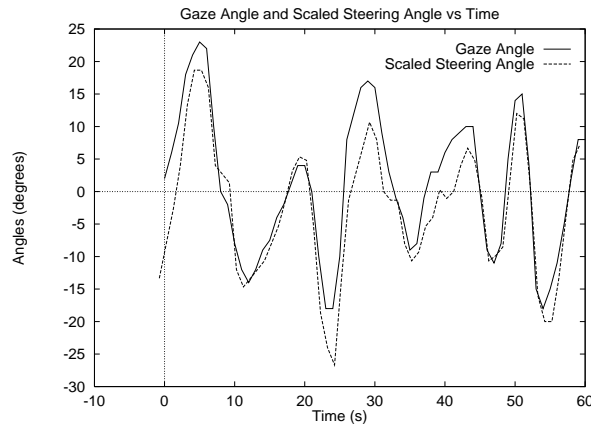
Figure 6: A sampled portion of the data of Land and Lee. The gaze angle is shown to scale, but the steering response has been advanced in time by 0.75s and scaled down by a factor of approximately 3.

$v = 12.5\text{ms}^{-1}$, giving $D \approx 17\text{m}$. If we take R as 1m, 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 [5] that indicate that drivers perform best when the fixation distance D is around 1s ahead of the vehicle.)

Thus for the driving conditions explored by Land and coworkers [6, 5] we should expect little deviation from linear proportionality between the instantaneous steering response and gaze angle. Proportionality also appears more appropriate when the behaviour is regarded within the framework of feedback control.

5 Towards recovering localization with an active head

The steering behaviour described is closed-loop and requires no knowledge of the vehicle's position to operate, and so is suited to tactical navigation without a map. It assumes that if the camera can see it, the robot can reach it. Obviously such a behaviour must accompanied by one for obstacle avoidance.

For navigation with a map, localization is required. If we assume that fixation alone is to be used, visual sensing will return only the range and direction relative to the vehicle of the currently fixated feature — exactly the sort of sensor data available from, say, sonar. But whereas sonar sensors are cheap enough to have several in a ring around a vehicle which can be multiplexed electronically, a fixating camera platform can address only one point in the scene, and there is a time cost associated with moving to look at another point. One approach that can be taken is for the vehicle to stop when localization is needed and use the active head to make sequential observations of several reference features. This technique, and its tie-in with familiar structure from motion methods in computer vision, has
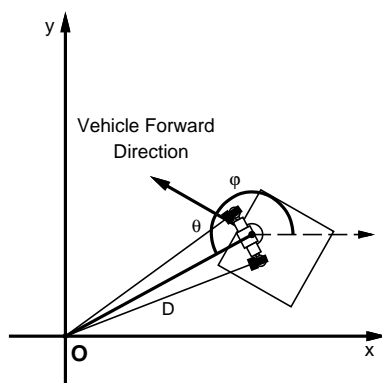
Figure 7: The geometry of the localization system. The robot vehicle at (x, y) and with orientation ϕ relative to the x-axis observes a reference point at the origin O and recovers a gaze angle θ and distance D.

been investigated in previous work [3], and will certainly be of value in some circumstances — for instance, for re-orientation after the robot has become lost or distracted from its current task. However, for efficient continuous localization while the vehicle is moving, the robot must extract whatever information is available from observation of a single feature at a time. What kind of information can be obtained, and where should it look to best improve knowledge of its location?

To begin to explore this problem we have implemented in simulation a location algorithm similar to one used in sonar navigation work by Leonard and Durrant-Whyte [7]. Much of this work on directed sonar sensing is relevant to navigation using active vision. One difference between the visual head and a sonar sensor is that with vision we are able to measure angles very accurately, but obtain quite large errors in measurements of depth. For sonar the reverse is true.

5.1   Localization with the Extended Kalman Filter

The situation in consideration is depicted in Figure 7. The reference point under consideration is used to define the origin of an xy coordinate frame. The orientation of the axes is determined by the initial position of the vehicle: we define the y-axis to pass through this point. The position and orientation of the vehicle at time t are hence expressed with three components (x, y, ϕ). Two input parameters control the vehicle motion: the velocity v at which it is driven, and the angle s to which the steerable wheel is turned.

At time intervals Δt a measurement is obtained from the active head of the distance D and gaze angle θ to the reference point. The Extended Kalman Filter provides a method for making the best use of these measurements to calculate the current position of the vehicle, by producing an estimate and estimate variance which depend not only on the most recently acquired data but implicitly on all the previous ones along the trajectory.

The state vector $\mathbf{x}$, control vector $\mathbf{u}$ and measurement vector $\mathbf{z}$ used in the
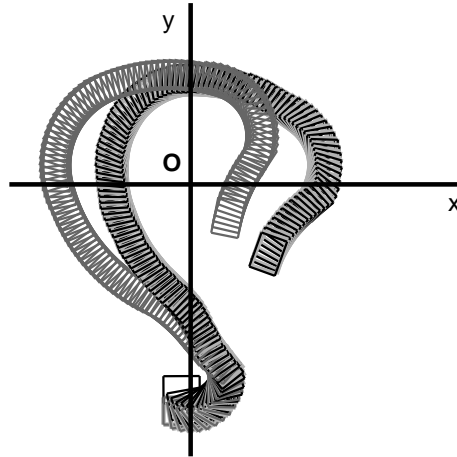
Figure 8: The light grey true vehicle trajectory is closely followed by the filtered estimate of position (shown in black), while an estimate based just on odometry data quickly diverges. The scale of the diagram represents 10 metres in the simulation.

EKF are therefore:

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ \phi \end{pmatrix} , \qquad \mathbf{u} = \begin{pmatrix} v \\ s \end{pmatrix} , \qquad \mathbf{z} = \begin{pmatrix} D \\ \theta \end{pmatrix} . \qquad (3)$$

The state transition equation describes how the system being modeled changes from one time step k to the next. In typical notation,

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k), \mathbf{u}(k)) + \mathbf{v}(k) , \qquad (4)$$

where $\mathbf{v}(k)$ is Gaussian process noise. For our system, the transition function $\mathbf{f}$ is

$$\mathbf{f}(\mathbf{x}(k), \mathbf{u}(k)) = \begin{pmatrix} x(k) + R(k)\left(\cos\phi(k)\sin K(k) + \sin\phi(k)\left(\cos K(k) - 1\right)\right) \\ y(k) + R(k)\left(\sin\phi(k)\sin K(k) + \cos\phi(k)\left(1 - \cos K(k)\right)\right) \\ \phi(k) + K(k) \end{pmatrix}$$

$$(5)$$

where $R(k) = L/\tan s(k)$ and $K(k) = v(k)\Delta t/R(k)$. L is the constant wheelbase of the vehicle, and Δt is the time-step. $R(k)$ has the physical interpretation of being the signed radius of curvature of the trajectory between points k and $k+1$, and $K(k)$ is the incremental angle through which the vehicle turns.

The measurement equation, written as

$$\mathbf{z}(k+1) = \mathbf{h}(\mathbf{x}(k+1)) + \mathbf{w}(k+1) , \qquad (6)$$

where $\mathbf{w}(k+1)$ is Gaussian measurement noise, relates measurements to the current state of the system. In our case,

$$\mathbf{h}(\mathbf{x}(k+1)) = \begin{pmatrix} \sqrt{x(k+1)^2 + y(k+1)^2} \\ \tan^{-1}\frac{-y(k+1)}{-x(k+1)} - \phi(k+1) \end{pmatrix} . \qquad (7)$$
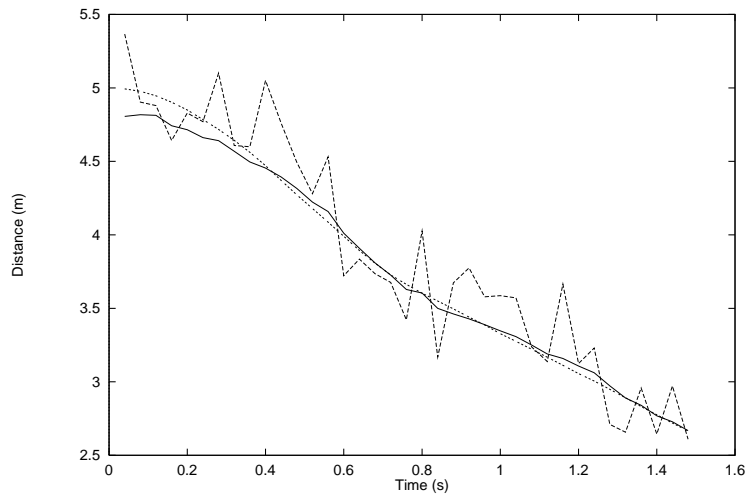
Figure 9: Distance of the vehicle from the reference point O over time: the smooth trajectory (solid line) output by the filter closely follows the true value (dotted line), and contrasts with the jagged dashed line representing the measurements.

Both types of noise **v** and **w** were modeled, in terms of potential errors in the vehicle control and head measurements respectively.

5.2 Testing the Localization Algorithm

Since the simulation provides us with ground truth for the vehicle motion, we can evaluate the performance of the filtering method for localization. Figure 8 shows how when the control parameters are manipulated to make the vehicle follow a winding course around the reference feature, the filtered estimation of position remains locked onto the true trajectory (to within a few centimetres typically). A second estimate, based solely on integration of the control parameters fed to the vehicle, which represents the information available from non-visual odometry, is seen to quickly diverge.

Figure 9 demonstrates clearly the smoothing effect of the filter on the recovered trajectory. It compares the noisy range measurements of the distance of the vehicle from the reference point (as a trajectory similar to the first stages of that in Figure 8 is followed) with the true value of this distance and the value calculated from the filter.

We deduce from these early results that the EKF does provide a very promising way to estimate localization from fixation on a single reference point. The next stage of our work will be to implement this technique in the real world for further experimentation.

6   Conclusions

We have demonstrated that a very simple rule linking fixation angle and steering angle can be used to guide a robot 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 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 rule proposed here appears to provide a simpler explanation of the natural observations of human driver performance made by Land and Lee [6], avoiding the need for the driver to estimate the average curvature of the road ahead to the next tangent point.

It would be of interest to measure human driver gaze angles and responses on sharper bends where $\sin^{-1}(R/D)$ 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 suggests 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.

We have shown that information very useful for localization can be obtained from active fixation on a world feature. Which features will be the best to fixate on and when should a transition be made from one to another? These are questions we hope to address soon, but instinct suggests that given an uncertainty in location, which can be thought of as an ellipse surrounding the current estimate, an observation at right angles to the major axis would make the best use of the accurate angular information available and reduce the uncertainty the most.

Acknowledgments

References

[1] R. Bajcsy. Active perception. *Proc IEEE*, 76:996–1005, 1988.

[2] D.H. Ballard. Animate vision. *Artificial Intelligence*, 48:57–86, 1991.

[3] A. J. Davison, I. D. Reid, and D. W. Murray. The active camera as a projective pointing device. In *Proc. 6th British Machine Vision Conf., Birimingham*, pages 453–462, 1995.

[4] J. Kŏsecká, R. Bajcsy, and M. Mintz. Control of visually guided behaviours. In *Real-time Computer Vision*. Cambridge University Press, Cambridge UK, 1994.

[5] M.F. Land and J. Horwood. Which parts of the road guide steering? *Nature*, 377:339–340, 28 September 1995.

[6] M.F. Land and D.N. Lee. Where look when we steer. *Nature*, 369:742–744, 30 June 1994.

[7] J. J. Leonard and H. F. Durrant-Whyte. *Directed Sonar Navigation*. Kluwer Academic Press, 1992.