

# Predictive Gaze Control

Christopher Brown

University of Rochester, NY 14627

---

*Animate vision systems, biological or robotic, employ gaze control systems to acquire, fixate, and stabilize images. Our goal is to guild robust gaze control behaviour from cooperating lower-level visual reflexes. Predictive control strategies can cope with time delayed, multi-rate, and interacting controls. Solutions are explored through simulation incorporating ten primitive gaze control capabilities, more or less comparable to subsystems in primate gaze and head control. Versions of several of the subsystems have been implemented on a binocular robot. Smith prediction is the basic paradigm, using kinematic simulation of the agent and optimal filtering to predict world state.*

---

One of the goals of artificial *animate vision* [2,3] is to design a systems architecture in which multiple objectives (such as moving and observing) can proceed in parallel. One common premise is that cognitive processes at high abstraction levels rely on a hierarchy of lower-level "skills", "reflexes", and active vision capabilities that autonomously keep the agent out of trouble and perform generally useful vision computations [7]. Another premise is that active control over and perception of an agent's own state (*proprioception*) makes many problems in perception, planning, and acting easier. A *gaze control* system manages several basic, interacting head, eye, and even body motion capabilities, with the aim of supporting purposeful (or default) activity. One basic activity is the visual *acquisition* of an object. The action can be reflexive, in response to a stimulus deemed interesting, or under control of a higher level engaged in planning or acting. Another basic ability is to *pursue* or track an object moving relative to the observer: Stabilization of the image on the sensor is necessary for high resolution imaging, and the resulting *proprioception* (i.e. motor commands that effected the tracking) provides information about object motion. From the point of view of control theory, a gaze control system has two main *technical problems*: the *interaction of component subcontrols* and *delay*. Both biological and robotic systems can easily have delays that are of the same order of magnitude as the timescale of the actions.

Gaze control mechanisms have long been studied in biological systems (there are extensive references in [12,5,19]). Much of the work concentrates on how biological systems solve the two basic technical problems mentioned above. This paper investigates predictive mechanisms as a solution for the problems primarily in a robotic context, but occasionally relates the results to some findings and theories from primate gaze control.

One way to cope with delays is to use strictly open loop control. The other approach, more common in engineering applications, is to use *predictive* and modeling techniques to anticipate the state of the plant, its input, and indeed the world [14], thus coping with both delays and interactions.

*Smith's principle* [21,22] is the basic tenet that the desired output from a controlled system with delay  $T$  is the same as that desired from the delay-free system, only delayed by  $T$ . The principle leads to several techniques for controlling delayed systems. Smith's principle may be coupled with *signal synthesis adaptive control* [1], which predicts object motion to allow more accurate responses. Kinematic and dynamic models for plant prediction can be known apriori or derived from learning and used to replace feedback [13]. The solution described here uses Smith prediction to integrate multiple controls with delays. It uses signal synthesis adaptive control with flexible and general techniques of *kinematic simulation* to predict the state of the plant and variance-minimizing *optimal filtering* to predict the state of the world. At least in simulation, the resulting predictions have three effects. *Delays are overcome, interactions are overcome, and performance is improved.* Predictive techniques seem to form a sound basis for the design of integrated, high performance sensorimotor systems.

## THE ROCHESTER ROBOT

Research in artificial *animate vision* has been made possible by recent technical advances in real-time computer vision and control. The area still includes a wide spectrum of problems from hardware design through software for parallel systems support and applications, to the integration of heterogeneous computers, sensors, and effectors into behaving systems. Many laboratories are developing similar systems to investigate these issues and to take advantage of the new technologies, and it seems that complex visuo-motor systems will be the rule in the robotics laboratory of the near future. The usual current setup has controllable sensors (often bi- or trinocular TV cameras, perhaps sonar) and powerful parallel computation, including frame-rate image analysis hardware. Sometimes the sensors are mounted on a roving cart.

Rochester's binocular robot head is mounted on a six degree of freedom arm. The two cameras are on a common tilt platform, and have independent pan axes. The hardware is capable of motions comparable to primate performance (about 1 m/sec head velocity with less than 1 mm. positioning accuracy, and 300 degrees/sec camera rotations with .14 degree positioning accuracy). The camera controllers are capable of supporting full-speed gaze-shifts to random directions at a rate of 5/s. The camera and robot controllers support several types of motion, but the frequency responses of system components under various forms of control have not been measured. The aperture, focal length, and focus of the cameras are not yet controllable. The video output is processed by a Datacube MaxVideo pipeline-parallel image processing system that can do many low and intermediate-level vision operations at 30Hz (video frame rate). The host computer for the system is currently a Sun/3 computer but

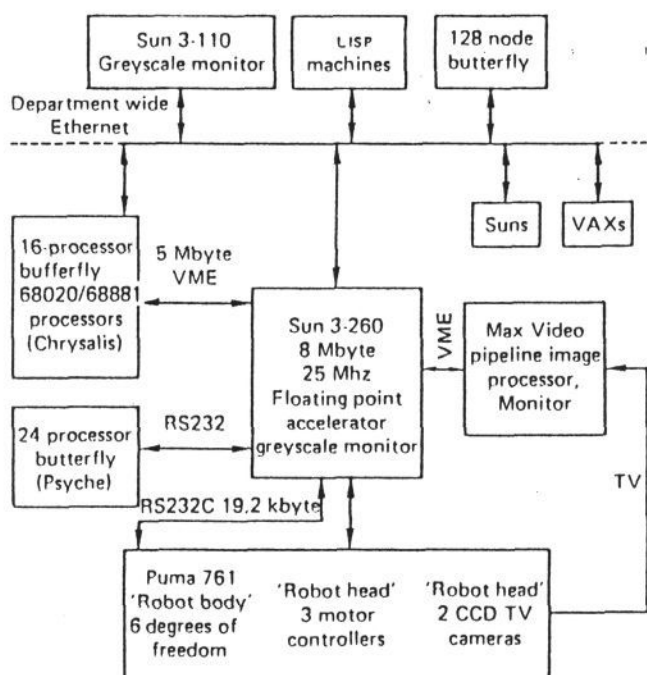


Figure 1: The laboratory computer organization. The 24 node Butterfly will replace the host Sun, and faster interfaces to the Puma and MaxVideo are being installed.

will soon be a 24-node Butterfly Parallel Processor (BPP). Each node of the BPP is an M68020 processor with M68881 floating point coprocessor and 4MB of memory. There is fast switch allowing the processors to share memory. The plan is to implement the control algorithms in the BPP, using multiple nodes as necessary for speed, with the Datacube furnishing real-time input. A real-time package for the Psyche operating system will support the applications. LISP planning programs can communicate with robot applications code over the ethernet. Fig. 1 shows the current hardware organization.

## BASIC CONTROLS

This paper deals, unless otherwise explicitly noted, with simulated capabilities and occasional psychophysical and physiological data. Early work [8] introduced five control systems, four of them resembling capabilities implemented on the Rochester Robot, and functionally reminiscent of primate capabilities of the same names.

**Saccadic:** an eye control to produce rapid gaze shifts, "sampled" at a relatively coarse interval with respect to the discrete simulation time increment, and driven from positional retinal error (distance from retinal origin). At each activation it produced a sequence of maximum-speed pan and tilt velocity commands that was calculated from approximate kinematics to center the image in the retina.

**Smooth pursuit:** a proportional, integral, derivative (PID) eye control driven by the retinal *positional* error of the object's image, to follow objects in motion relative to the eyes. This and the following controls are "continuous" in that they are computed at every simulation time increment – we use the quoted "continuous" and "sampled" in the sequel to make

this distinction in a discrete (sampled-data) implementation. **Vergence:** a PID eye control to reduce disparity between left and right images.

**Vestibulo-ocular reflex (VOR):** a proportional-gain eye control to oppose head motion with contrary eye rotation. Head rotation is easily canceled, and this capability also includes a version of "otolith-ocular reflex" that compensates for translational head motions using information about object range [6,11]. Its input is not sensory, but an "efferent copy" of the head translation and rotation command.

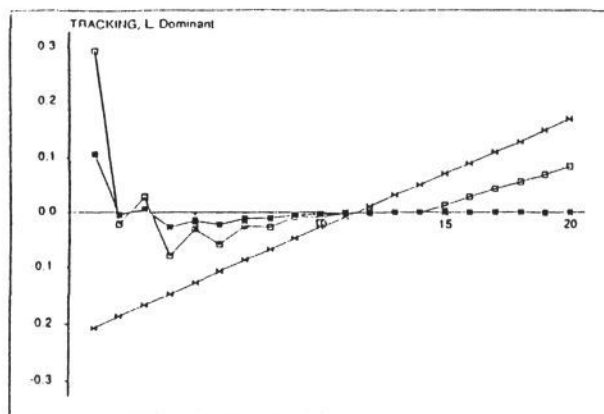
**Head compensation:** a proportional-gain head-control system driven by head-relative eye position that rotates the head in the direction of eye rotation to keep the eyes centered and away from their mechanical stops. This capability may not mirror any named biological one, but it does reflect a widespread primate tendency to move the head, if allowed, during pursuit and eye saccades.

Fig. 2 illustrates the cumulative effect of simply superimposing control capabilities: each operates independently and their outputs are simply summed at the effectors. Delays are zero, latencies (except for saccades) unity. Fig. 3 shows the dramatic effects of control delay on the system. The smallest delays, applied uniformly or to just one control, destabilize the system seriously.

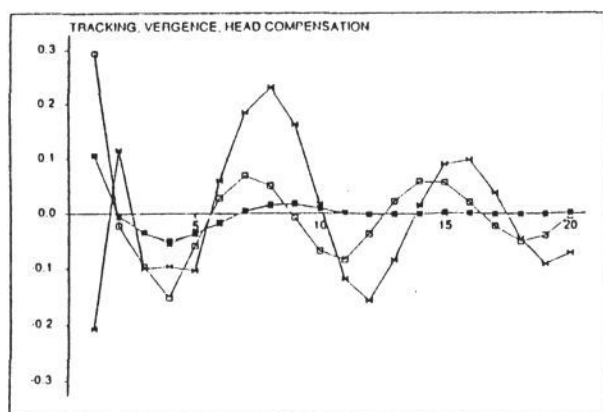
## ENHANCED CONTROLS

In later work some improvements were made to the controls. This work is more fully described in [9]. All controls originally operated from retinal coordinates. However, predictions of object position in head or LAB (laboratory) coordinates are needed to predict retinal images. Head rotations, pans, and tilts all induce camera origin translations due to the geometry of the head, and non-retinal representations are more robust (as when the object temporarily is lost). *Versions of pursuit, vergence, and eye saccadic capabilities were added that use spatial information.* There was no capability for estimating the state of objects moving in LAB (relative motion was produced with a static object and observer motion.) *A pipeline of object state descriptions is maintained using variance-minimizing filters to predict object state from observations.* The head compensation reflex was the only head control. *The system was given "head saccade" and "head pursuit" controls.* The eye saccade control algorithm was unsophisticated, and there was no significant head and eye cooperation for quick gaze shifts. *Four fast gaze-shift algorithms involving both head and eyes were investigated.* The two modes of operation (pursuit and saccadic) were simplistically assumed to correspond to inflexible combinations of lower level capabilities. *All controls can now be activated and deactivated independently.*

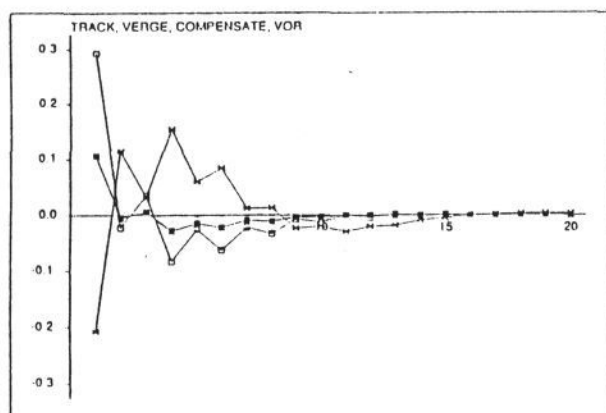
The later work is not committed on several issues, such as whether the eye saccadic system should control one or both eyes. Binocular eye saccades are easy to implement, but there are many other ways that the eyes can cooperate during gaze shifts, depending on technical considerations such as the possibility of visual computation during the shift, how the vergence control is specified, etc. We likewise shall not address issues of orchestrating the smooth transition between saccadic and pursuit (or other) tasks. Considerable work is still needed on the topic of smooth blending of subcontrols, which forms the foundation of motor skills.



(a)



(b)



(c)

Figure 2: Increasingly effective delay-free control results from superposition of noninteracting controllers. (a) Tracking only: The left (dominant) eye pans and tilts, inducing tilt in the right eye. The tracker uses a position error signal. The right eye gets no pan signal, and its horizontal error accrues from target motion. The left eye tracks successfully until it hits mechanical stop at tick 14. (b) Add vergence and head compensation: This control is to keep eyes from hitting mechanical stops by turning the head in the same direction as the tracking motion. A less-desirable effect is to amplify the tracking signal, overcompensating and destabilize the tracking. (c) Add VOR, which effectively compensates the head rotation with eye rotations.

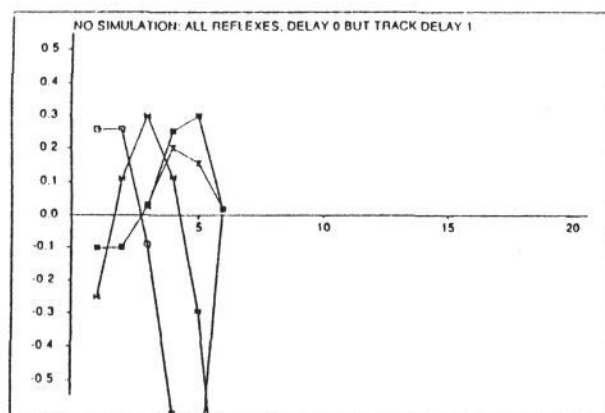


Figure 3: The no-delay controller applied with zero delay in all controls except tracking, which has a delay of one tick. Ideally this graph should be a delayed version of Fig. 2(c).

## DEALING WITH DELAYS

The Smith predictor incorporating the object state predictions is shown in Fig. 4. Its derivation appears, for instance, in [8,14]. The basic idea is to have a zero-delay feedback (path C) based on simulation. The model-delayed simulation data (path B) is compared with the actually-delayed data from the plant (path A): the difference (at D) is zero for perfect simulation, so it provides information about the simulation adequacy and (if slowly varying compared to the control delays) can compensate for inaccurate modeling. A more satisfactory approach (not yet implemented) is parameter adaptive control: compute sensitivity derivatives and use the signals from the plant (path A) and the difference between predicted and actual behavior (path D) to *adapt* the model. This is the dual problem of model-reference adaptive control, which forces the system to *conform* to the model.

The work described here uses the following *interacting controls* algorithm [8], which assumes each controller knows its own delay  $T$ , and the delays of all the other controllers in the set  $S$  that share an output with it. *Look ahead the maximum delay  $M$  of any controller in  $\{S\}$  and retrieve the predicted robot and control states for that time. Apply the control appropriate for these future states at (possibly future) time  $M-T$ .*

In early work, only the kinematic simulation existed, and so in dynamic vision situations, prediction could only be accurate if the simulated robot moved while the object remained stationary. Fig. 5 shows how the interacting controls algorithm successfully stabilizes performance (compare Fig. fig:node14) even when several subcontrols are at work, all with different delays. Prediction fails if knowledge of the world is inaccurate: Fig. 6 illustrates the need for prediction of 3-D object motion. The need is addressed with dynamic models and optimal filter prediction for moving objects.

## OBJECT PREDICTION

In contrast with the explicit kinematic simulation used to predict the system state, standard optimal (i.e. variance-minimizing) filtering techniques are used to predict the po-





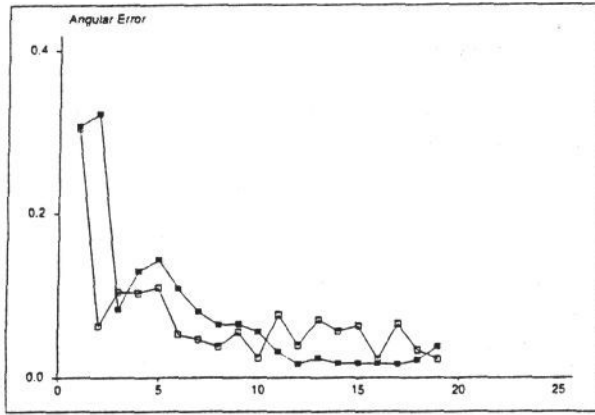


Figure 7: The angular pointing error of the left eye while pursuing a moving object (initially off center in the visual field), measured to the true object position. Light squares: results with current noisy image data. Dark squares: results with output of the  $\alpha - \beta$  filter.

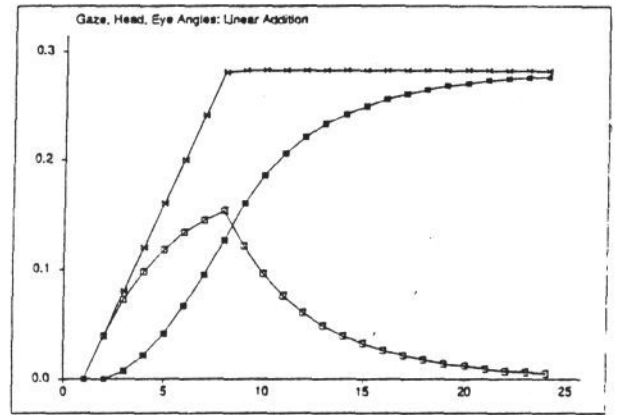
$\hat{x}(k+1|k)$  predicted from the last estimate to be  $F\hat{x}(k|k)$  and the innovation, or difference between a predicted measurement and the actual measurement. The predicted measurement  $\hat{z}(k+1|k)$  is produced by applying (here a trivial) measurement function to the predicted state.

The  $\alpha - \beta$  filter is a special case of the Kalman filter. For our assumptions, the optimal values of  $\alpha$  and  $\beta$  can be derived (see [4], for example) and depend only on the ratio of the process noise standard deviation and the measurement noise standard deviation. This ratio is called the object's *maneuvering index*  $\lambda$ . The constant covariances assumed by time-invariant filters may suffice for sensors stationary in LAB, but a head-mounted depth sensor would very likely require the full power of a Kalman filter.

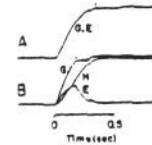
The incorporation of an optimal filter into the control loop was motivated by the necessity of predicting object state for the Smith predictor. The filter, however, also has a noticeable and beneficial effect on pursuit performance (Fig. 7), and indeed it seems that humans may use stochastic prediction [16,18].

## EYE-HEAD SACCADES

This section describes four versions of cooperating eye-head motions. A first example of an eye-head saccade is provided by simply running the "sampled" eye saccades simultaneously with the "continuous" VOR and head-compensation reflexes ([8]). The result (Fig. 8(a)) is similar to that which once was claimed (under the "linear summation hypothesis") to hold in primates, *viz.* that their eye rotational velocity measured in LAB is kept constant (eye saccadic velocity is decreased by head saccadic velocity) [17]. However, in primates, the linear summation hypothesis has been contradicted by findings indicating eye-head gaze shifts outperform eye saccades alone (Fig. 8(b)) [20]. Certainly in a robotic context a speed improvement is attainable, and the following three versions are intended to outperform the first one, and are more fully described in [9].



(a)



(b)

Figure 8: (a) Robot gaze (butterflies), eye (light squares), and head (dark squares) angles showing linear summation effect arising from reflex interaction. Here the eye saccade control is active together with with head-compensation and VOR reflexes. The combination of eye saccades with head compensation and VOR forces the gaze velocity to be the velocity of the eye saccade acting by itself. Time units are arbitrary and chosen to illustrate the behavior. (b) Human gaze (G), eye (E), and head (H) movements during 80 degree gaze shift. A has no head motion, B has 80 degree head movement, and gaze shift is completed twice as fast.

The *Gaze Feedback* version is inspired by "local feedback" physiological theories, and is based on "continuous" head and eye position-feedback controls. A head saccade is simply implemented as a pursuit operation, and cooperating eye-head gaze-shifts are simultaneous head and eye pursuit. For various reasons this version fails to achieve a speed improvement. The *Sampled Optimal* version is a closed-form "sampled" solution for full-speed eye and head movements. It depends on an open-loop calculation depending on known constant head velocities, and so is not robust against variations or disturbances.

The *Simulated Optimal* version uses the familiar pipeline of predicted future states to maintain optimal speed and add the "continuous" ability to deal with arbitrary velocity profiles as well as disturbances. The algorithm first computes a head saccade by determining, for the  $x$  and  $y$  dimensions, the direction to rotate and then driving at maximum velocity, stopping after there is a zero-crossing in predicted object position (say at time  $k$ ). The head position at time  $k-1$  may be closer to the object: the closest position is chosen. The sequence of head-saccade commands is inserted into the pipeline and the resulting changes to head and eye position computed (these values overwrite existing values, as the saccade is taken to replace previous control). Then the eye saccade is computed exactly like the head saccade, except eye position is computed by adding the head-relative eye velocity to the head velocity. To cope with disturbances, and to implement successive ("catch up") saccades, the simulation is placed in a loop that runs until the saccade is successfully completed, checking if at any time a new calculation must be performed. Fig. 9(b) shows a result from this control scheme.

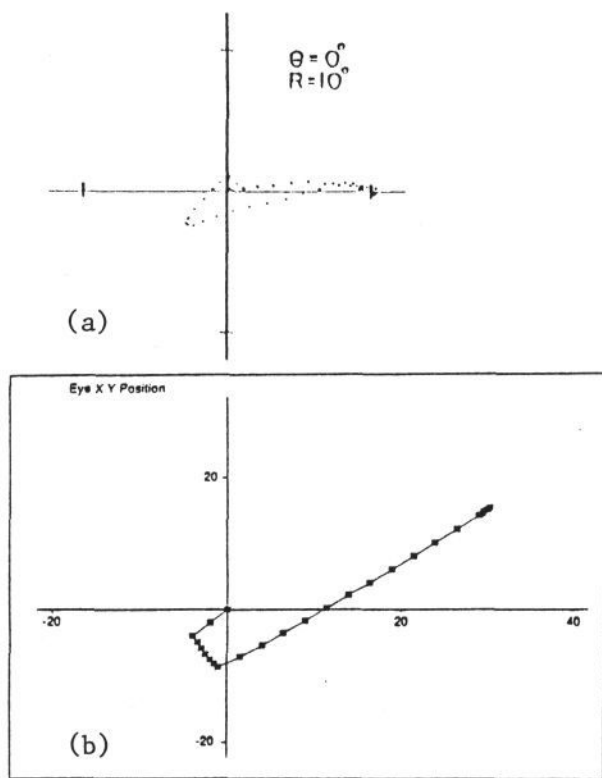


Figure 9: (a) Data from primate saccades with eye velocities: undisturbed and disturbed (eyes dragged down to left) prior to saccade but after stimulus has vanished. (b) Robot camera (x, y) position for eye-only saccade to moving object. with disturbance during  $3 \leq k \leq 10$  adding negative (x, y) velocity before and during first part of saccade.

After reaching the object, the eye continues tracking the moving object. When macaque monkeys have their eye velocities electrically disturbed after the target stimulus has vanished, they successfully correct the saccade (Fig. 9(a)). The recalculation can sometimes be done during the latency period of the original saccade, so the correction does not delay saccade onset [15]. In the simulated system the control delay is unavoidable, but correction occurs with both eye and eye-head saccades.

## References

- [1] A. T. Bahill and J. D. McDonald. Adaptive control models for saccadic and smooth pursuit eye movements. In A. F. Fuchs and W. Becker, editors, *Progress in Oculomotor Research*, Elsevier, 1981.
- [2] D. H. Ballard. Animate vision. In *International Joint Conference on AI-89*, August 1989.
- [3] D. H. Ballard. Behavioural constraints on animate vision. *Image and Vision Computing*, 7(1):3-9, 1989.
- [4] Y. Bar-Shalom and T. E. Fortmann. *Tracking and Data Association*. Academic Press, 1988.
- [5] A. Berthoz and G. Melvill Jones. *Adaptive Mechanisms in Gaze Control: Facts and Theories*. Elsevier, 1985.
- [6] B. Biguer and C. Prablanc. Modulation of the vestibulo-ocular reflex in eye-head orientation as a function of target distance in man. In A. F. Fuchs and W. Becker, editors, *Progress in Oculomotor Research*, Elsevier, 1981.
- [7] R. A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA-2:14-23, 1986.
- [8] C. M. Brown. Gaze controls with interactions and delays. In *DARPA Image Understanding Workshop. Submitted IEEE-TSMC*, pages 200-218, May 1989.
- [9] C. M. Brown. *Prediction in gaze and saccade control*. Technical Report, Oxford University OUEL 1771, U. Rochester Comp. Sci. Dept. TR-295, May 1989.
- [10] C. M. Brown, H. Durrant-Whyte, J. Leonard, and B. Rao. Centralized and noncentralized kalman filtering techniques for tracking and control. In *DARPA Image Understanding Workshop*, pages 651-675, May 1989.
- [11] A. Buizza, R. Schmid, and J. Droulez. Influence of linear acceleration on oculomotor control. In A. F. Fuchs and W. Becker, editors, *Progress in Oculomotor Research*, Elsevier, 1981.
- [12] A. F. Fuchs and W. Becker. *Progress in Oculomotor Research*. Elsevier, 1981.
- [13] A. Kawato, K. Furukawa, and R. Suzuki. A hierarchical neural network model for control and learning of voluntary movement. *Biological Cybernetics*, 57:169-185, 1987.
- [14] J. E. Marshall. *Control of Time-Delay Systems*. Peter Peregrinus Ltd., 1979.
- [15] L. Mays and D. Sparks. The localization of saccade targets using a combination of retinal and eye position information. In A. F. Fuchs and W. Becker, editors, *Progress in Oculomotor Research*, Elsevier, 1981.
- [16] J. A. Michael and G. Melvill Jones. Dependence of visual tracking capability upon stimulus predictability. *Vision Research*, 6:707-716, 1966.
- [17] P. Morasso, E. Bizzi, and J. Dichgans. Adjustment of saccade characteristics during head movements. *Experimental Brain Research*, 16:492-500, 1973.
- [18] P. D. Neilson, N. J. O'Dwyer, and M. D. Neilson. Stochastic prediction in pursuit tracking. *Biological Cybernetics*, 58:113-122, 1988.
- [19] B. W. Peterson and F. J. Richmond. *Control of Head Movement*. Oxford University Press, 1988.
- [20] D. A. Robinson and D. S. Zee. Theoretical considerations of the function and circuitry of various rapid eye movements. In A. F. Fuchs and W. Becker, editors, *Progress in Oculomotor Research*, Elsevier, 1981.
- [21] O. J. M. Smith. Closer control of loops with dead time. *Chemical Engg. Prog. Trans.*, 53(5):217-219, 1957.
- [22] O. J. M. Smith. *Feedback Control Systems*. McGraw-Hill, 1958.