

# Gaze Control for a Two-Eyed Robot Head

Fenglei Du Michael Brady David Murray

Robotics Research Group

Engineering Science Department

Oxford University

19 Parks Road Oxford OX1 3PJ

## Abstract

As a first step towards building a two-eyed active vision system, gaze control is discussed in this paper. Instead of using separate subcontrollers for each of the subfunctions, loosely corresponding to saccade, pursuit, vergence, VOR(vestibulo-ocular reflex), and OKR(opto-kinetic reflex), that most previous work has done, a potentially parallel gaze controller is proposed whose structure is supervising-planning-adaptation. Based on simulation, the cooperations and interactions in gaze control and the consequence of delays due to image processing and the local controllers are discussed.

## 1 Introduction

As a first step towards integrating image processing, attentional control and gaze control into a system which can be used in real tasks, gaze control is discussed in this paper. Our goal is two fold. First, we want to study the interactions between perception and action. We want to know that, if the behavior of the system can be actively controlled by perception, to what extent (and how) can perception be improved. This is an essential question which we must address before we can possibly realize *intelligent machines*. Second, by studying a computer vision system which is similar in some aspect to the primate visual system, we hope to provide some hints to physiologists to get a better understanding of primate vision.

Following the definitions used in physiology[10, 9], gaze control can loosely be divided into:

- saccade: quickly shift the head to fixate to a new area of interest
- pursuit: tracking the object of interest
- vergence: to reduce the retinal disparity
- focusing: keeping the camera to focus on the object of interest
- VOR: stabilize the images when the head is moving
- OKR: using optic flow(retinal slip) to stabilize the images

Gaze control has been studied by several groups[13, 7, 1, 2, 3]. But the proposed methods either too simple for real task [7] or too complicated to implement[1, 2, 3]. Here we propose a new potentially parallel gaze controller.

## 2 The head

The proposed mechanism for our active vision system is shown in figure 1. It has four degrees of freedom and is driven by DC servo motors. Its details and kinematics are discussed in [11].

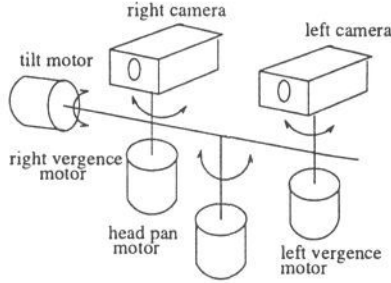


Figure 1: The mechanism for the active vision system

## 3 Gaze Control

The functional structure of the system is shown in figure 2. There are three control loops. Loop A consists of image processing and perception. For the gaze controller, it is a feedback loop. So far as the gaze controller is concerned, this loop provides delayed feedback. Since we cannot find an equivalent delay-free system[8] for this structure, the Smith predictor cannot be used to deal with the time delays in this loop. In order to reduce the effects of time delay and noise(of image processing) in this loop, a filter and a predictor must be used in the gaze controller so that the motion parameters of the object of interest used for gaze control is nearly noise and delay free.

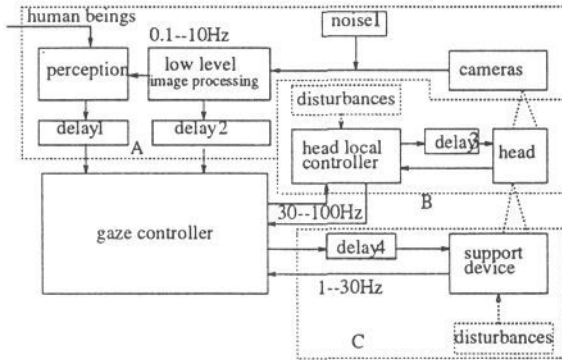


Figure 2: The structure of the active vision system with time delay and noise

Loop B is the main control loop of our gaze controller. *delay3* represents the time that the head package needs to carry out an instruction from the gaze controller. Obviously, this delay is determined by the type of instructions and the properties

of the local controllers. In our case, it is not difficult to design a good path planner so that  $delay_3$  is small. For a “sluggish” system with a small time delay, Smith’s method is not optimal[8] (the joint control of our head can be expect to be “sluggish”). The best controller for this kind of system is the conventional PID controller. This is mainly because it is difficult to acquire the exact model of the plant. The mismatch between the plant model and the real plant decreases the performance of a Smith predictor. So in this control loop too, a Smith predictor is not appropriate.

The third loop is  $C$ , which is not tightly coupled to the gaze controller. When the head is mounted on some device and the control of this device is necessary for gaze, this loop is involved. Here we are mainly concerned with the gaze control of the head so the details of this loop will not be discussed further in this paper.

Combining all the above, the details of the gaze controller are shown in figure 3. A potentially parallel structure is observed. There are three main parts. The supervision and decision making is mainly used to decide the behavior of the gaze controller. It includes filtering, prediction and decision making. The planning is used to transform the desired action to the trajectories of the joints according to the state and characteristics of the head. Instead of adjusting the control parameters of local controllers to deal with the nonlinearities and disturbance in the plant, the signal synthesis mechanism tries to synthesize the best instructions to the local controllers according to the desired trajectory, the state of head, and the joint models. Actually, since there are no control parameters to be adjusted in real time in our local controllers, the signal synthesis adaptation is the *only* way to improve the performance of our gaze control.

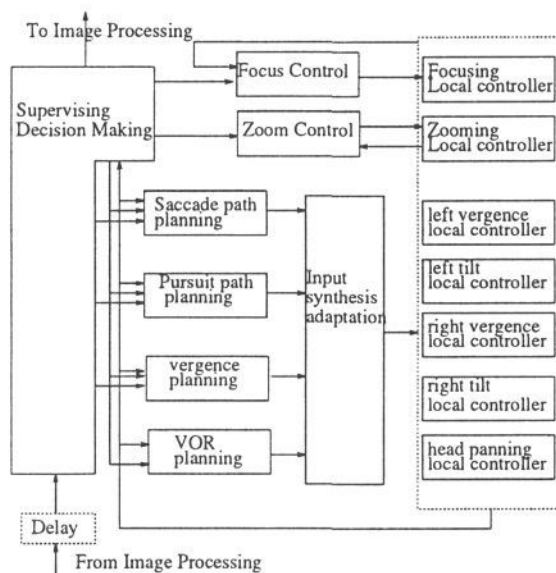


Figure 3: The structure of gaze controller

The path planning is discussed in [12]. The other parts of gaze controller is discussed in the following sections.

## 4 Supervising and Decision Making

Decision making determines how the head should respond to a changing situation. Supervising deals with time delays and noise of image processing, and maintaining the higher level controller's estimate of the state of the head. Here only the major part of supervising, filtering and prediction, are discussed.

### 4.1 Filtering and prediction

For filtering and prediction, linear Kalman filters are used.

For the estimation of motion parameters of the object of interest, the state equation is:

$$\mathcal{X}(k+1) = F(k+1) * \mathcal{X}(k) + \mathcal{V}(k+1) \quad (1)$$

where:

$$F(k) = \begin{bmatrix} I & I * \delta T_k & I * \frac{\delta T_k^2}{2} \\ O & I & I * \delta T_k \\ O & O & I \end{bmatrix} \quad \mathcal{X}(k) = \begin{bmatrix} X(k) \\ V(k) \\ A(k) \end{bmatrix} \quad \mathcal{V}(k) = \begin{bmatrix} \nu_p(k) \\ \nu_v(k) \\ \nu_a(k) \end{bmatrix} \quad (2)$$

where the position, velocity, and acceleration of the object of interest in 3-D at time  $k$  are  $X(k) = [x(k), y(k), z(k)]^T$ ,  $V(k) = [v_x(k), v_y(k), v_z(k)]^T$ , and  $A(k) = [a_x(k), a_y(k), a_z(k)]^T$  respectively, the time interval from time  $k$  to time  $k+1$  is  $\delta T_{k+1}$ , and  $\nu_p$ ,  $\nu_v$ , and  $\nu_a$  are Gaussian, identically distributed, temporally uncorrelated noise with zero mean.  $I$  and  $O$  are  $3 \times 3$  unit and zero matrices respectively.

The measurement equation is:

$$Z(k+1) = [I, O, O] * \mathcal{X}(k+1) + \omega(k) = H * \mathcal{X}(k+1) + \omega(k+1) \quad (3)$$

Similarly, the equations for disparity can be got.

### 4.2 Decision Making

A potentially parallel decision making process is listed in table 1. Where the VOR has been redefined as the process of adjusting the "eye" to its most comfortable position.

Table 1: The subcontrol activate conditions for the gaze control

control	condition for being activated
saccade	the object of interest is "far" from the head fixation point
pursuit	the object of interest is moving and "near" the fixation point
VOR	being enabled and the head is at a "uncomfortable" position
OKR	being enabled and there exists optical flow in the image planes
vergence	being enabled and the object of interest is "stationary"
focusing	being enabled

## 5 Adaptation

The local controllers of the head package are position error driven control systems. If the instructions to the local controllers are given by the speed of the joints, there must exist some delays. This is what we must avoid. Generally, if we want a PID control system for a mechanical joint with no static error to a constant speed input, the following control law should be adopted:

$$C = K_p(X_d - X) + K_v(\dot{X}_d - \dot{X}) \quad (4)$$

where  $K_p$  and  $K_v$  are the position gain and velocity gain,  $C$  is the control,  $X_d$  and  $\dot{X}_d$  is the desired motion,  $X$  and  $\dot{X}$  is the real motion. Generally,  $K_v$  is selected such that the system is critically damped. For our system, in order to achieve the above control law, we can give the instruction as follows:

$$E = (X_d - X) + \frac{K_v}{K_p}(\dot{X}_d - \dot{X}) \quad (5)$$

To date we have not used adaptive techniques in our gaze controller. If the disturbances and nonlinearities in the local controllers are significant, adaptive control techniques can easily be applied.

## 6 Simulation

Based on the above discussion, a simulator of the gaze controller has been implemented. In order to show the effects of the delays of local controllers, three different types of models of local controllers are used (corresponding to large, moderate, and small delay). Their responses to different inputs are shown in figure 4. The workspace of the head is shown in figure 5.

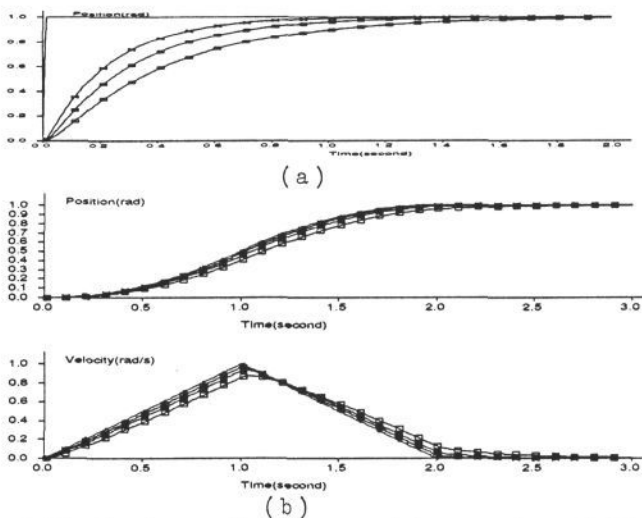


Figure 4: The local controller responses (a) to step input (b) bang-bang control

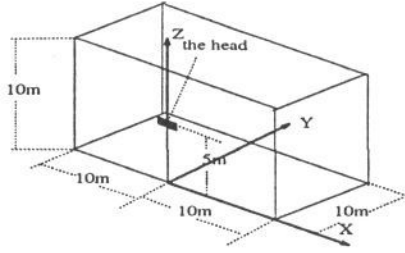


Figure 5: The work space

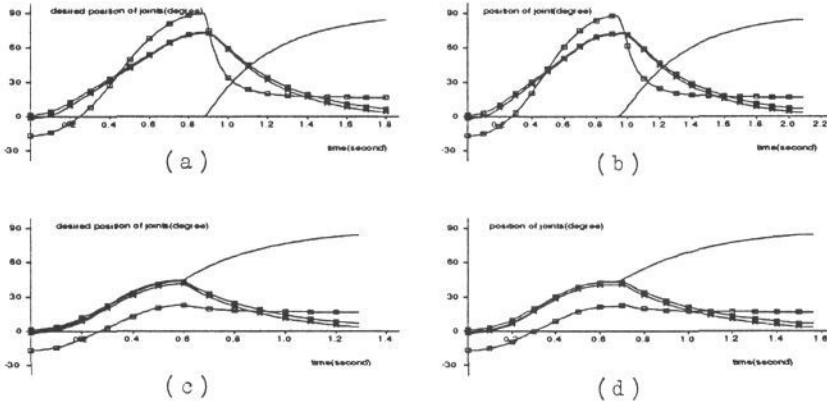


Figure 6: The cooperation and interaction of saccade and VOR. (a) Saccade while holding head panning stationary, planned motion, a large VOR motion is needed. (b) Saccade while holding head panning stationary, real motion, and VOR. (c) Full motion saccade, planned motion, VOR is relatively small. (d) Full motion saccade, real motion, and VOR. solid line: pan; the line with hollow squares: tilt; the line with black squares: left vergence joint; and the other is for the right vergence joint

Examples of the interactions and cooperations between saccade and VOR is shown in figure 6. All the controls adopt a saccade then VOR scheme. The saccade is from an initial point  $(0.0, 10.0, 2.0)$  to an end point  $(-10.0, 0.0, 8.0)$  in the work space.

The relationships between gaze angle, head angle, and eye angle during the above saccade are shown in figure 7. Compared with the saccades of human beings[1], The response of this system is slightly slower. This is mainly because the maximum joint acceleration in the simulation is quite small.

An example of the cooperation and interaction of tracking, saccade and VOR are shown in figure 8. The trajectory is a circle with diameter  $10m$  on a wall at  $Y = 10.0m$ . The initial fixation point of the head is at  $(0.0, 10.0, 5.0)$  (the center of the circle). The time delay of image processing is  $0.1ms$ . The added noise is Gaussian with standard deviation  $3mm$ .

The interactions of different delays on pursuit is shown in figure 9. We can clearly see that the delays of local controllers have only a little effect on the tracking error. The delay of image processing is the main error sources.

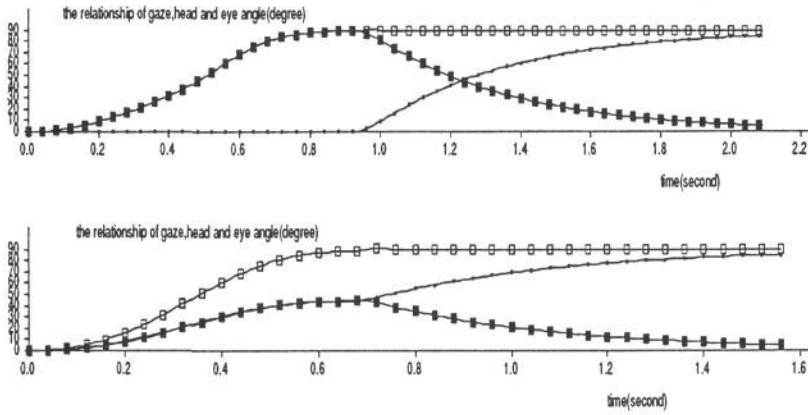


Figure 7: The angular relationship of gaze, head and eye motion during saccade. Solid line describes the head angle. The line with hollowed squares represent the gaze angle. The eye angle is described by the line with solid squares.

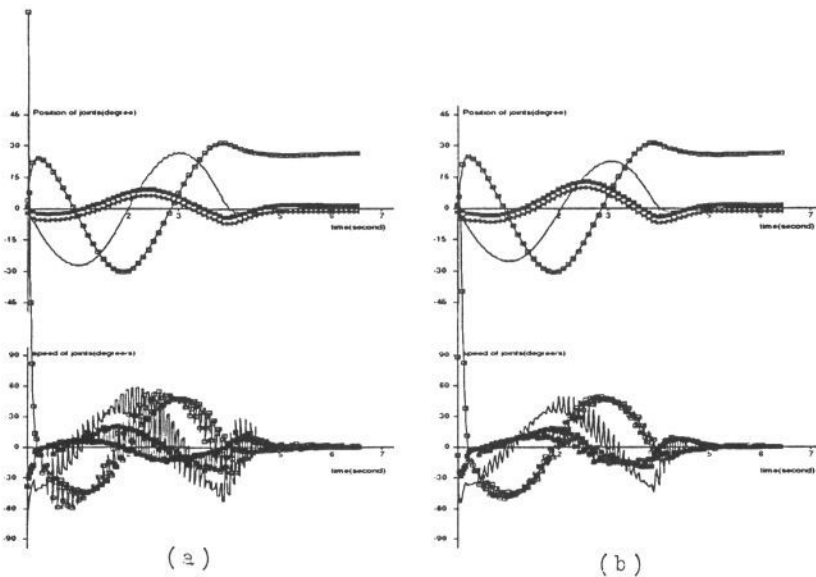


Figure 8: The cooperation and interaction of pursuit, saccade and VOR (a) the planned motion for the head (b) the real motion of the head

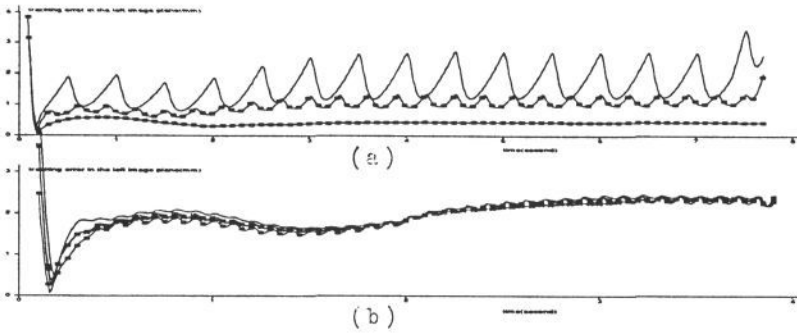


Figure 9: The tracking error in the left image plane along the trajectory ( $f = 28mm$ ). (a) The effect of image processing delay. The object speed is  $4m/s$ . The solid line: delay  $0.5s$ , the line with hollowed squares: delay  $0.3s$ ; the line with solid squares: delay  $0.1s$ . (b) The effect of local controller delay. The speed of object along the circle is  $8m/s$ . Image processing delay is  $0.1s$ . Solid line: large delay; the line with hollow squares: moderate delay; The solid square line: small delay; The line with butterflies: mixed delay.

## 7 Discussion

The cooperations of the subcontrols in gaze control are quite complicated, especially when decision making is included. Saccade will work cooperatively with VOR, vergence, pursuit and focusing. Generally, there is no interaction between VOR, vergence and saccade if the limitations of joint angles are ignored. If the joint angles' limits are considered, the interaction between saccade and VOR must be considered.

For smooth tracking, pursuit must work cooperatively with saccade, VOR, vergence and focusing. The VOR and focusing act like the "dog", they will check the state of the head at every sample time. If necessary, the control is activated. We have not introduced vergence into our simulator yet, so the interactions between vergence and other controls have not been studied. The interactions and cooperations between pursuit and saccade is an essential relationship in smooth tracking. If the tracking error is not small the saccade control must be used to quickly reduce the tracking error. If the tracking error is smaller than a given constant the smooth tracking must be activated.

In all the simulations we have performed, the gaze controller behaves properly even though the interactions and cooperations of the subcontrol are quite complicated and frequent. No stability problem arises.

Combining the simulation results up to now we make the following conclusions:

- Designing a subcontroller for each subfunctions in gaze control to build up the gaze controller is not appropriate.
- Using a supervising-planning-adaptation structure to design the gaze controller appears to be more reasonable.
- The performance of the gaze controller is mainly determined by the delay due to image processing.
- To some extent, the delays of the local controllers has only a very small effect



on the performance of gaze control.

- The sample rate of the gaze control must be high enough so as to handle the frequent initialization of rapid attention shift.

## Acknowledgments

Thanks to Paul Beardsley and the others in the group for their help during the work. Fenglei Du thanks to Jesus College of Oxford University for the Ning Po scholarship. Partially support from ESPRIT grant VOILA is acknowledged.

## References

- [1] C. Brown. Prediction and Cooperation in Gaze Control. In *Biological Cybernetics*, 63:61-70, 1990.
- [2] C. Brown. Gaze Control with Interactions and Delays. *IEEE Trans Syst Man Cybern IEEE-TSMC20(2)*, March, 1990.
- [3] C. Brown. Prediction in Robotic Gaze and Saccade Control. April 26 1989.
- [4] J.(Y.) Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. In *First International Conference on Computer Vision, (London, England, June 8-11, 1987)*, pages 35-54, Washington, DC., 1987. IEEE Computer Society Press.
- [5] R. Bajcsy. Active Perception. In *P-IEEE* 76:996-1005, 1988.
- [6] C. Brown. Gaze controls with interactions and delays. In *Image Understanding Workshop (Palo Alto, CA, May 23-26, 1989)*, pages 200-218, San Mateo, CA, 1989. Defense Advanced Research Projects Agency, Morgan Kaufmann.
- [7] J.J. Clark and N.J. Ferrier. Modal control of an attentive vision system. In *Second International Conference on Computer Vision (Tampa, FL, December 5-8, 1988)*, pages 514-523, Washington, DC., 1988. Computer Society Press.
- [8] J.E. Marshall. Control of Time-delay Systems. Peter Peregrinus Ltd., 1979.
- [9] R.H.S. Carpenter. *Movements of the Eyes*. Pion, London, 1988
- [10] D.A. Robinson. The Oculomotor Control System: A Review. In *Proceeding of IEEE*, Vol.56, pp 1032-1049, 1968.
- [11] Fenglei Du. The Fundamentals for a Two-Eyed Active Vision System. First Year Report, Robot Research Group, Oxford University, 1991.
- [12] Fenglei Du, Michael Brady and David Murray. The Kinematics and Eye Movements of a Two-Eyed Robot Head. In *Proceeding of British Machine Vision Conference 91*, Glasgow, 1991.
- [13] K. Pahlavan and J.O. Eklundh. A Head-Eye System for Active, Purposeful computer Vision. TRITA-NA-P9031, CVAP 80, Royal Institute of technology, Stockholm, Sweden, 1990.