TRACKING WITH HYBRID-DRIVE ACTIVE VISION

†Orson Sutherland, ‡Sebastien Rougeaux, †Samer Abdallah & ‡Alex Zelinsky

Robotic Systems Laboratory
†Department of Engineering, FEIT
‡Department of Systems Engineering, RSISE
Australian National University
Canberra, ACT 0200 Australia
http://syseng.anu.edu.au/rsl/

Abstract

The work in this paper demonstrates effective tracking in an everyday environment with hybrid-drive active vision. We base our approach on the assumption that to efficiently perceive an unknown, unconstrained, potentially evolving environment, the observing system must be able to process and act on gathered visual information as it happens in real-time. Moreover, we endow the robot's tracking system with two behavioural subgroups, saccade and smooth pursuit and show that saccade time and motion smoothness can be optimised through the use of a single trapezoidal profile motion (TPM) algorithm. In this way, the simplicity and compactness of the controller's design is improved. To demonstrate this, we present a 4 degree of freedom binocular active vision platform, nicknamed HyDrA, that uses cable-drive technology to minimise mass-in-motion and hence improve performance. By combining template matching and TPM control, we show that HyDrA can detect and track the motion of an arbitrary object as it moves within a cluttered scene.

1 Introduction

This paper presents an active sensor for use in human/robot interaction and autonomous robot navigation called HyDrA (for Hybrid Drive active vision). It takes the view that the most important sensory faculty that can be endowed upon a robot destined for interaction with humans and human environments, be that for gesture recognition or navigation, is visual perception. The work undertaken on HyDrA described here is to prove that cable drive and TPM control are relevant and capable technologies in Active Vision systems. It will briefly outline the Active Vision concept and then introduce cable drive and TPM. Finally, it will present experiments in which a human head is tracked in a cluttered environment.

In the late eighties a new paradigm was born on the back of increased computer processing power, known as Active [Aloimonos et al., 1988] or Animate [Ballard, 1991] Vision. It was founded on the notion of dynamic perception, where the act of sensing was intimately intertwined with action [Brooks, 1992]. More precisely it is the process of extracting meaningful information about a scene through the strategic movement of vision sensors. It is in stark conceptual contrast to the passive approach which attempts to reconstruct 3D models of a scene from static images. The advantages of the active approach over the passive include full panoramic range of view, foveal vision, stereo fusion and the ability to obtain moving images of static scenes [Rougeaux, 1999].

Without a doubt, Active Vision has been largely in-

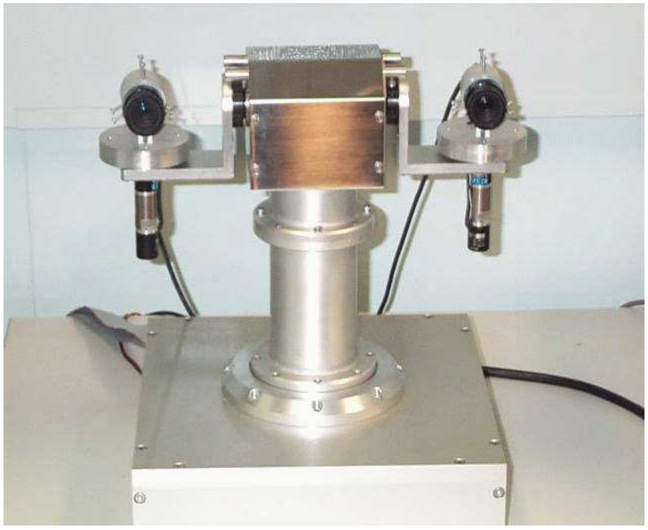

Figure 1: HyDrA

spired by the principles displayed in biological vision systems. The concepts it has laid claim to in the field of robotics are common to many biological organisms ranging from insects to mammals. Nevertheless, perhaps out of some perceived introspective knowledge, it is the human ocular system that the vast majority of platforms have attempted to mimic. These are characterised by a pair of cameras that can rotate either symmetrically for control of vergence [Ferrier and Clark, 1993], or independently to achieve both vergence and version [Sharkey et al., 1993] with-in an independently controlled tilt plane. For more information on other leading platforms see [Pahlavan et al., 1996; Maki, 1996; Sharkey et al., 1993; Reid and Beardsley, 1995; Reid and Murray, 1996; Murray et al., 1995; Murray et al., 1997; Rougeaux and Kuniyoshi, 1998]

In addition to the mechatronic design, the control strategies employed in Active Vision are also inspired by biological systems. Particularly, *saccade* and *smooth pursuit* are two very important ocular reflexes that active platforms have adopted. Saccades involve rapid point to point motion between areas of interest in a scene such as when a new target is fixated. To smoothly follow a moving object however, a different strategy is employed. Smooth pursuit is the act of tracking the movement of such an object. This involves moving the "eyes" to keep the target centered in the field of view as it moves through the observed scene.

Attention is now turned to HyDrA's mechanical and control implementation.

2 HyDrA

HyDrA is a high-performance binocular head developed in our laboratory at the Australian National University and exhibits similar properties to the human vision system in that it has:

- **Binocular Vision.** HyDrA carries two cameras for stereo fusion.
- **Four mechanical degrees of freedom.** HyDrA can control the cameras in vergence and version as well as tilt. A neck allows for pan.
- **High Dynamic Performance.** HyDrA's rotational velocities and accelerations are comparable to human vision system capabilities and allow for fast target acquisition and saccade. In particular its vergence and tilt joints can achieve velocities in excess of $450°.s^{-1}$ at accelerations of $20000°.s^{-2}$.
- **High Precision.** Speed and position can be controlled to a high degree of precision for saccade, smooth pursuit and fixation.

The fundamentals of the design are heavily influenced by the mechanics of the human vision system which is by definition binocular, light weight and agile. This agility is attributable to the advantageous inertial properties of the eye, the considerable acceleration of the ocular muscles and the layout of these muscles around the eye which allow for each degree of freedom to be driven independently and unburdened by the extra weight of the co-members in the system. To help simulate these properties HyDrA makes use of cable-drive mechanisms to reduce mass-in-motion and hence enhance the inertial properties of the head in addition to improving response and precision.

2.1 Cable Drive

Brooks *et al.* (1997) reported that by relocating the joint motors to a fixed off-board location, high-performance could be achieved through a small low-inertia 'eye', without incurring the high cost, complexity or weight of more traditional approaches. The idea was that joints were linked to their respective motors by light-weight high-tensile cables so that no motors were required to be moved about any axis.

Cable drive has proven to be an effective means of motive force in robotic applications such as manipulators [Townsend and Salisbury, 1993]. In principle it involves the transferral of power from a motor to a final moving component via a series of cable wrapped pinions and pulleys. The cables are fixed at both ends and in the case of the bevel gear used in HyDrA relies on the terminators to apply torque to the drive-train (see Fig. 2). Obviously the point-to-point angle of rotation is restricted

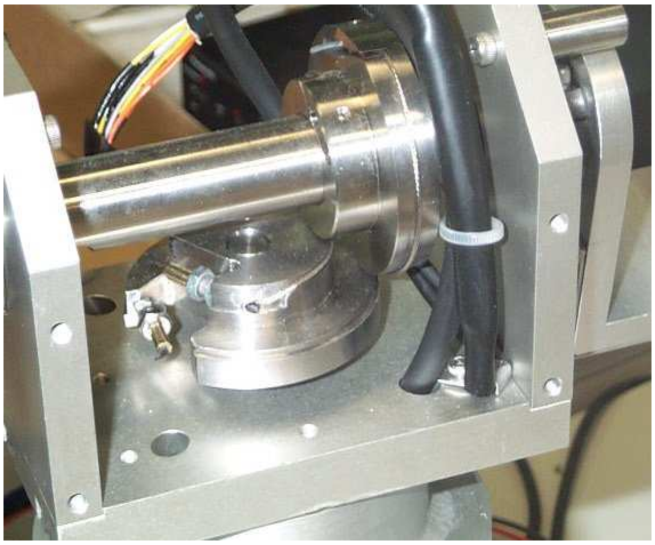

Figure 2: Picture of HyDrA's cable drive bevel gear, with terminators

by the length of the cable, but in light of simulating humanoid vision this is not a disadvantage. In addition to being comparatively light-weight, several benefits are bequeathed to cable driven transmissions. Joints incur:

- **No backlash.** Standard, low-cost gear mechanisms suffer from significant backlash relative to the desired angular resolution in active vision systems. Cable drive removes the need for performance compromise or expensive solutions to backlash problems.
- **No slippage.** They are free from problems of compliance found in belt-drive and certain gear systems.
- **No need for lubrication.** They are free of friction arising from tooth meshing and are thus suitable for rapid motion with out the need for lubrication.
- **Very little loss.** Cable wrapped joints only incur rolling friction which is high efficiency. As a result transmission efficiency is very high (typically 96%).

In the case of HyDrA both cable and harmonic drive transmissions are used. In particular its pan and tilt joints are cable driven and its vergence joints carry harmonic drive motors: hence hybrid drive (see figure 1).

2.2 HyDrA's mechanics and performance specifications

HyDrA has 4 degrees of freedom where the cameras are implemented in the Helmholtz configuration, sharing a common tilt plane but mounted on separate vergence axes. In this way, they are free to move independently of each other with-in the tilt plane allowing for both vergence and version control. To further increase the potential field of view a neck has been implemented to allow for pan.

HyDrA can achieve peak velocities and accelerations comparable to the human eye, where high-speed degrees of freedom are simulated by the tilt and vergence axes. Each joint is powered by a DC motor which allows for velocities in excess of $450^{\circ}.\mathrm{s}^{-1}$ for tilt and vergence at accelerations of up to $20000^{\circ}.\mathrm{s}^{-2}$. In addition high resolution encoders allow for very precise position and velocity control. Table 2.1 summarises HyDrA's mechanical performance parameters.

Axis	Pan	Tilt	Vergence
Max Velocity	$350^{\circ}.s^{-}1$	$470^{\circ}.s^{-}1$	$450^{\circ}.s^{-}1$
Max Accel	$16000^{\circ}.s^{-}2$	$16000^{\circ}.s^{-}2$	$20000^{\circ}.s^{-}2$
Angular Res	0.0018°	0.0018°	0.0018°
Axis Range	180°	180°	180°

Table 1: Key performance parameters for HyDrA

3 Control

HyDrA's controller is an extension of preliminary work undertaken by Murray *et al.* (1992) on TPM. In particular our approach allows for the implementation of a single algorithm for both saccade and smooth pursuit, enhancing the simplicity and compactness of the controller's design.

In theory, TPM can provide both the quickest and the smoothest possible state to state motion given the system constraints. The controller requires only the current head velocity, the predicted target velocity during the move, the distance of the target from the center of the head's field of view and the time allowed for the move. To see how the controller achieves these behaviours we consider the TPM algorithm it implements in the following.

3.1 Trapezoidal Profile Motion

The essence of the TPM problem, also known as the bang-coast-bang model, is to catch a target initially a distance x_0 from the image center either in the shortest time possible in the case of saccade or in the smoothest way possible in the case of smooth pursuit. Both the joints' and the target's starting velocities are potentially non-zero and disparate. Specifically it causes an axis to

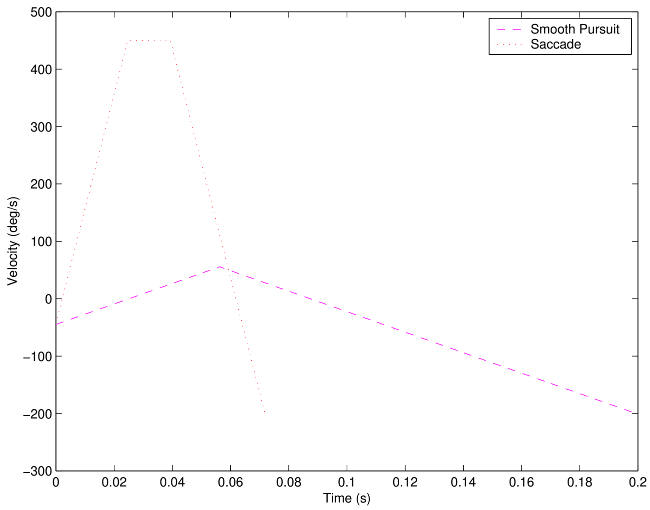

Figure 3: Velocity profiles during trapezoidal motion for saccade and smooth pursuit

accelerate at a constant acceleration to a pre-calculated ceiling velocity[1], coast at this velocity for a given period and then decelerate at the same constant rate as the initial acceleration until the target velocity is reached (see figure 3). From a mathematical perspective it is a 4 dimensional problem, where the bang acceleration a, ceiling velocity v, move time T and total distance travelled x are the unknowns. The initial joint velocity v_1, the target velocity v_2 and the target's initial distance from the image center x_0 are the givens.

[1]The maximum absolute velocity of the TPM trajectory

The Algorithm

Since the acceleration a is assumed to be constant, the time taken by the head to accelerate from its initial velocity v_1 to the ceiling velocity v is

$$T_a = \frac{s \cdot v - v_1}{s \cdot a}, \tag{1}$$

where s is positive if the head accelerates from v_1 to v and negative if it decelerates.

Similarly the time taken to decelerate from the ceiling velocity to the target velocity v_2 is

$$T_d = \frac{s \cdot v - v_2}{s \cdot a}. \tag{2}$$

Note that the rate of deceleration is equal to the rate of acceleration.

If T_c is the time spent coasting at the ceiling velocity, the total time of the trapezoidal profile motion is

$$T = T_a + T_c + T_d. \tag{3}$$

The distance traveled by the head during the action is

$$x = \frac{s \cdot v + v_1}{2} T_a + s \cdot v T_c + \frac{s \cdot v + v_2}{2} T_d, \tag{4}$$

but can also be considered as the sum of the initial distance of the target from the foveal center x_0 and the distance travelled by the target during the move

$$x = x_o + T v_2. \tag{5}$$

With these general TPM equations the specifics of saccade and smooth pursuit can now be developed.

Saccade

Saccades involve changing the head's current position and velocity state to that of the target, as inferred by its previous states, in the shortest time possible (see Fig. 4). Motion smoothness is not a concern and hence acceleration is set to its maximum possible magnitude. Two cases can arise:

- The ceiling velocity required for the action is less than the maximum allowed velocity and hence no time is spent coasting.
- The theoretical ceiling velocity required for the action is greater than the maximum allowed velocity and hence some time must be spent coasting (see figure 3)

It is useful to assume T_c as initially being zero so that (1), (2), (4) and (5) yield

$$s \cdot v = v_2 \pm \frac{1}{2}\sqrt{4 s x_0 a - 2(v_1^2 + v_2^2) - 4 v_1 v_2}, \tag{6}$$

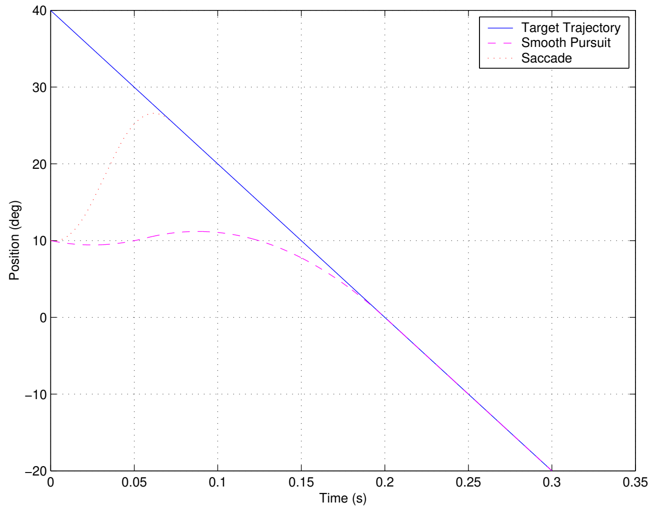

Figure 4: Position profiles during trapezoidal motion for saccade and smooth pursuit

where the smaller of the two values is taken if v_2 is greater than v_1 and vice-versa. But if both of these values are in excess of the maximum allowed velocity, acceleration and velocity in (1), (2), (4) and (5) are replaced by their respective maxima from which

$$T_c = \frac{1}{v_2 - s \cdot v}\left(\left(\frac{s \cdot v + v_1}{2} - v_1\right) T_a + \left(\frac{s \cdot v + v_2}{2} - v_2\right) T_a - x_0\right), \tag{7}$$

is calculated and hence T is deduced.

Equation (6) also defines the value of s. In particular it must be such that the operand of the radical is greater or equal to zero, hence:

$$\begin{cases} \text{s = +1} & \text{if } (v_1 - v_2)^2 + 4 x_0 a \geq 0 \\ \text{s = -1} & \text{otherwise} \end{cases}$$

Smooth Pursuit

In smooth pursuit we wish to move from one position and velocity state to the next in a given amount of time with the optimal smoothness (see Fig. 5) . To achieve this the acceleration in moving to and from the ceiling velocity must be as small as possible. Again both the coasting and no-coasting cases mentioned above are relevant and again we start by assuming that the coasting time is zero initially, from which (1), (2), (4) and (5) yield:

$$v = \frac{x}{T} \pm \frac{1}{2T}\sqrt{4x^2 - 4Tx(v_1 + v_2) + 2T^2(v_1^2 + v_2^2)}. \tag{8}$$

If these values are in excess of the maximum allowable velocity of the head, the time constraint is unrealiseable. In this eventuality there are two possibilities:

- The move time is replaced by the smallest possible time given the constraints (amounting to a saccade)

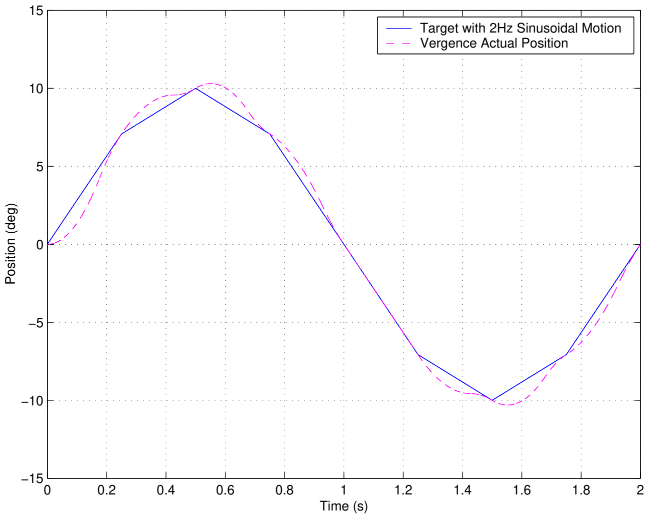

Figure 5: Tracking a sinusoidal target with trapezoidal profile motion at 8Hz

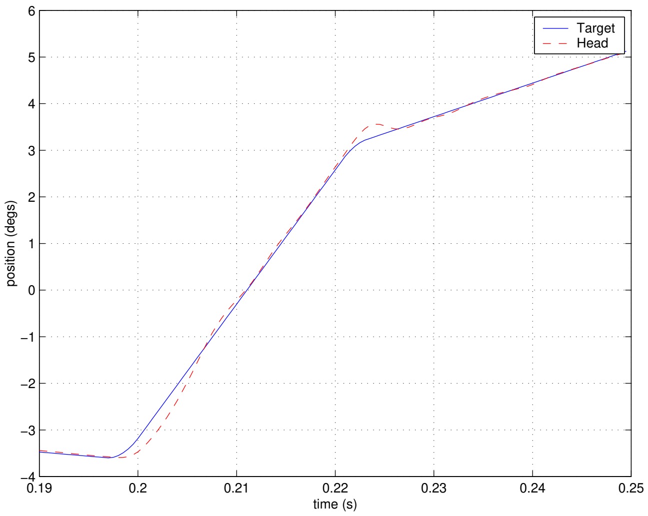

Figure 6: Sequence of two real-time experimental saccades

and the move is completed. In addition an error code is returned to the controller.

- The move is aborted and an error code is returned to the controller.

In the case of HyDrA, the controller has been implemented to initiate a saccade if the minimum possible ceiling velocity is above the physical limits of the head. If both the values in (8) are feasible, the smaller of the two values is taken if v_2 is greater than v_1 and vice-versa. Finally, the operand of the radical in (8) is never negative and so the s term can be neglected.

4 Implementation and Results

In principle a TPM controller is simple and elegant. Being able to perform both saccades and smooth pursuit (with optimal smoothness), it is also both compact and functional. Figure 7 shows the flow diagram for a complete tracking system using template matching in conjunction with TPM control. The controller identifies

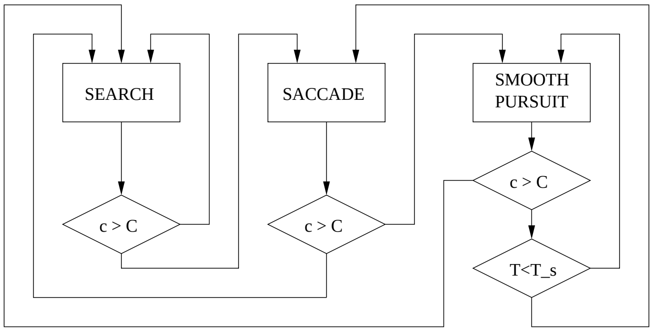

Figure 7: Flow chart for the gaze controller.

three major states: *Search, Saccade and Smooth Pursuit.* In the search phase the point with the highest correlation value above some limit C is identified as the location of the target in the image. The value of C is determined empirically. Since the target can potentially be anywhere in the field of view, the full image is used as the search window. Once located, the controller then tracks the target statically for three sample periods in order to predict its velocity and position in the next frame. With this information it instigates a saccade to center the object as quickly as possible. To test whether the saccade was successful, a 32x32 pixel search window around the image center is cross-correlated with the template. If the peak result is less than C, the target is considered lost and the search state is resumed. Otherwise the saccade is deemed successful and the controller commences smooth pursuit.

Figure 8 shows the position, error and velocity curves for a sequence involving the search, fixation and smooth pursuit of a human head using an ear template. A move time of 30ms was used to ensure processor delays did not cause any mismatch between predicted target position and actual target position during the 30Hz loop cycle. The initial instability is due to template projection variance subsequent to saccade. We propose to overcome this by investigating the use of alternative vision processing techniques such as colour matching, zero disparity filtering and optic flows.

5 Conclusion

The mechanical design of the head has proven reliable and capable. As well as reducing mass-in-motion, cable

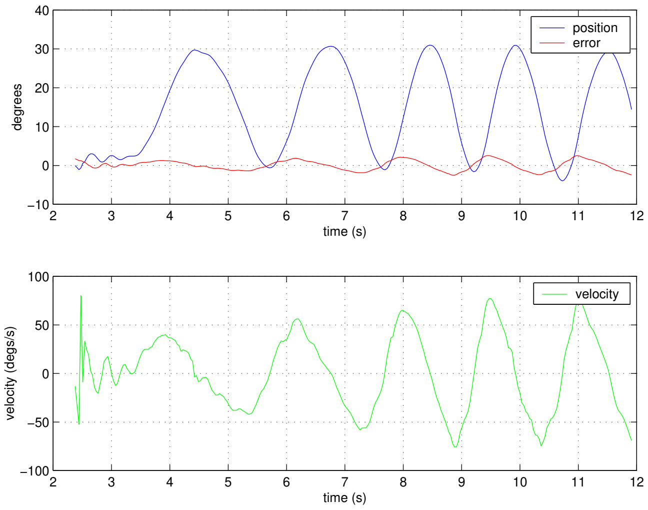

Figure 8: Search, saccade and tracking an oscillating target with a 30ms move time.

driven joints have zero backlash allowing for high resolution and faster response. In addition we have shown that TPM can be used to implement a controller capable of both saccades and smooth pursuit. A common algorithm for both saccades and smooth pursuit is advantageous in its simplicity and compactness of design. To demonstrate the effectiveness of the controller template matching was employed in conjunction with TPM to track a human head. However, template matching is too simplistic a vision strategy and hence future work will investigate, colour matching, optic flows and zero disparity filtering.

References

[Aloimonos et al., 1988] Aloimonos, J., Weiss, I., and Bandopadhay, A. (1988). Active vision. *International Journal on Active Vision 1*, pages 333–356.

[Ballard, 1991] Ballard, D. (1991). Animate vision. *Artificial Intelligence 48*, pages 57–86.

[Brooks et al., 1997] Brooks, A., Dickens, G., Zelinsky, A., Keiffer, J., and Abdallah, S. (1997). A high-performance camera platform for real-time active vision. In *Proceedings of the First International Conference on Field and Service Robotics*, pages 559–564, Canberra, Australia.

[Brooks, 1992] Brooks, R. (1992). *Forward. Active Vision.* MIT Press.

[Ferrier and Clark, 1993] Ferrier, N. J. and Clark, J. J. (1993). Camera heads and model based navigation and reactive control. *Active Robot Vision*, 6:9–31.

[Maki, 1996] Maki, A. (1996). *Stereo Vision in Attentive Scene Analysis.* PhD thesis, Computer Vision and Active Perception Laboratory, Stockholm University.

[Murray et al., 1995] 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 Conference of Computer Vision 16(3)*, pages 205–228.

[Murray et al., 1992] Murray, D. W., Du, F., McLauchlan, P. F., Reid, I. D., Sharkey, P. M., and Brady, M. (1992). Design of stereo heads. In Blake, A. and Yuille, A., editors, *Active Vision*, pages 155–174. MIT Press.

[Murray et al., 1997] Murray, D. W., Reid, I. D., and Davison, A. J. (1997). Steering without representation using active fixation. *Perception 26*, pages 1519–1528.

[Pahlavan et al., 1996] Pahlavan, K., Uhlin, T., and Eklundh, J.-O. (1996). Dynamic fixation and active perception. *International Journal of Computer Vision 17*, pages 113–135.

[Reid and Beardsley, 1995] Reid, I. D. and Beardsley, P. (1995). Self-alignment of a binocular robot. In *British Machine Vision Conference.*

[Reid and Murray, 1996] Reid, I. D. and Murray, D. W. (1996). Active tracking of foveated feature clusters uing affine structure. *International Journal of Computer Vision 18(1)*, pages 41–60.

[Rougeaux, 1999] Rougeaux, S. (1999). *Real-Time Active Vision for Versatile Interaction.* PhD thesis, Universite d'Evry Centre d'Etude Mechanique d'Ile de France and Humanoid Interaction Laboratory Electrotechnical Laboratory.

[Rougeaux and Kuniyoshi, 1998] Rougeaux, S. and Kuniyoshi, Y. (1998). Robust tracking by a humanoid vision system. In *First Int. Workshop Humanoid and Human Friendly Robotics, Tsukuba, Japan.*

[Sharkey et al., 1993] Sharkey, P. M., Murray, D. W., Vandevelde, S., Reid, I. D., and McLauchlin, P. F. (1993). A modular head/eye platform for real-time reactive vision. *Mechatronics 3(4)*, pages 517–535.

[Townsend and Salisbury, 1993] Townsend, W. T. and Salisbury, J. K. (1993). Mechanical design for whole-arm manipulation. *Robots and Biological Systems: Towards a New Bionics?*, pages 153–164.