International Journal of Computer Vision, 11:2, 147-164 (1993) 147

Real-Time Binocular Smooth Pursuit

DAVID COOMBS*
NIST, Robot Systems Division,†Gaithersburg, MD 20899 david.coombs@nist.gov

CHRISTOPHER BROWN
University of Rochester, Computer Science Department, Rochester, NY 14627 brown@cs.rochester.edu

Received October 1, 1992; Revised May 21, 1993

Abstract

This paper examines the problem of a moving robot tracking a moving object with its cameras, without requiring the ability to recognize the target to distinguish it from distracting surroundings. A novel aspect of the approach taken is the use of controlled camera movements to simplify the visual processing necessary to keep the cameras locked on the target. A gaze holding system implemented on a robot's binocular head demonstrates this approach. Even while the robot is moving, the cameras are able to track an object that rotates and moves in three dimensions.

The central idea is that localizing attention in 3-D space makes precategorical visual processing sufficient to hold gaze. Visual fixation can help separate the target object from distracting surroundings. Converged cameras produce a horopter (surface of zero stereo disparity) in the scene. Binocular features with no disparity can be located with a simple filter, showing the object's location in the image. Similarly, an object that is being tracked is imaged near the center of the field of view, so spatially-localized processing helps concentrate visual attention on the target. Instead of requiring a way to recognize the target, the system relies on active control of camera movements and binocular fixation segmentation to locate the target.

*This work was largely completed while Coombs was affiliated with the University of Rochester.

†Correspondence should be addressed to David Coombs, NIST, Robot Systems Division, Building 220, Room B-124, Gaithersburg, MD 20899, `david.coombs@nist.gov`.

Product endorsement disclaimer: references to specific brands, equipment, or trade names in this document are made to facilitate understanding and do not imply endorsement by the National Institute of Standards and Technology.

1 Introduction

The importance of eye movements to biological visual systems is obvious from their ubiquity. Even insects exhibit eye movements, although they are accomplished by head movements [Land, 1975]. In contrast, controlled camera movements have played a relatively small role in computer vision research. However, there is growing interest in the role of camera movements in robotic visual perception, and some lessons for computer vision systems may be learned by studying biological vision systems. Both have limits on resolution and field of view, and they must therefore direct their visual sensors toward areas of the environment that are of interest. Also, both animal and robot visual systems exist to provide visual perception of the dynamic environments in which their owners operate. Animals and robots move through a world of still and moving objects, and they must recognize threats and opportunities; they must be able to fix their gaze on objects in order to do so.

The larger problem of gaze control can be broken down functionally into the subproblems of *holding gaze* on a visual target and *shifting gaze* between objects. Gaze shifts, called *saccades*, transfer fixation rapidly from one visual target to another.

The goal of *smooth pursuit* camera movement is to follow an object with the cameras, holding gaze fixed steadily upon it. This contrasts with computer vision's traditional (passive) tracking task. In passive tracking, the object's image is tracked in the views of the cameras; the cameras move without regard to the motion of the target object. For instance, the cameras on a mobile robot may point straight ahead like automobile headlights. The optical flow observed by the robot will result from the three dimensional structure of the scene and the robot's motion. Further, the target will move about in the cameras' images. In contrast, during active visual following the cameras rotate to follow the target. Consequently, the target's retinal[1] slip (slip of the target's image on the imaging surfaces) is minimal, and the image of the surrounding scene slips instead. In addition, the target's image is held near the center of the field of view.

In order to be as general as possible, the pursuit system should be able to follow a moving object without necessarily recognizing it first. A robot that must recognize an object to be able to follow it will only be able to use that facility in domains where every object is known to it and in which it has a means to recognize all objects. Such a robot's applicability will clearly be limited. In its most general form, distinguishing the visual target is the notorious segmentation problem, and it is not clear how to extract this information from the visual signal in all cases. This work explores the use of *precategorical* visual cues (*i.e.*, prior to object recognition) in order to distinguish the visual target from distracting objects and the surrounding scene.

The goal of this work has been to build a robot gaze holding system whose only knowledge of the target is essentially that the cameras are initially pointed at it. The situation is depicted in Figure 1. The gaze holding problem is to maintain fixation on a moving visual target from a moving platform. To do this, the errors in camera orientation must be determined, so the location of the target's image on the retina must be found. The approach taken in this work exploits binocular cues and the fact that the cameras are actively following the target to avoid requiring object recognition.

It is important to note that it is easier to detect the tracking signals for active visual following than for tracking an object in passive stereo-motion image sequences. First, motion blur emphasizes the signal of target over the background. In passive visual following, the target's image slips across the retina and may thus be degraded by motion blur. During active pursuit, however, the eyes move to follow the target and stabilize the retinal image. Thus the image of the surrounding scene rather than the target moves across the retina and suffers from motion blur. The result is that image of the target is emphasized over the image of the background. Second, maintaining vergence isolates the target by disparity filtering. Holding vergence on the target enables the object to be isolated by simple zero-disparity filtering that detects objects at the fixation distance. The simplest implementation is just the logical AND of stereo vertical edge images. Thus, maintaining vergence on the target makes it possible to locate the target for pursuit control with simple precategorical visual processing. Third, active visual following also enables localized visual processing. The target's retinal location is roughly known because the pursuit system is keeping it near the center of view. This permits spatially localized visual processing, as illustrated in Figure 2.

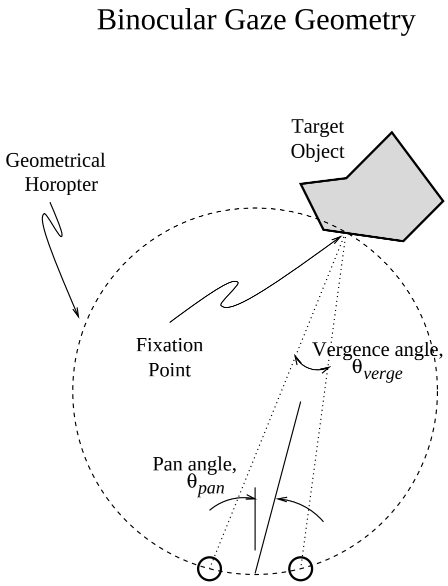


Figure 1: Top view of binocular gaze geometry. The goal of gaze holding is to keep the eyes or cameras fixed on a common world point or visual target. The gaze vector, $\vec{\theta}$, consists of the gaze pan and tilt angles and vergence angle. In order to keep the world point fixated, the gaze holding system must generate gaze and vergence angles that keep the cameras directed toward the target.

2 Background

Holding gaze is a fundamental capability of biological visual systems. There are several visual perception and behavioral motivations for holding gaze in robotic as well as biological systems, including foveal vision, and motion blur.

Foveal Vision: Foveal visual systems have small areas of high resolution. To get a high resolution image of an object, the fovea must be directed

[1]The *retina* is the back of the eye, where the photoreceptors are located. We will also use the term "retina" to refer to the receptor array of a camera.

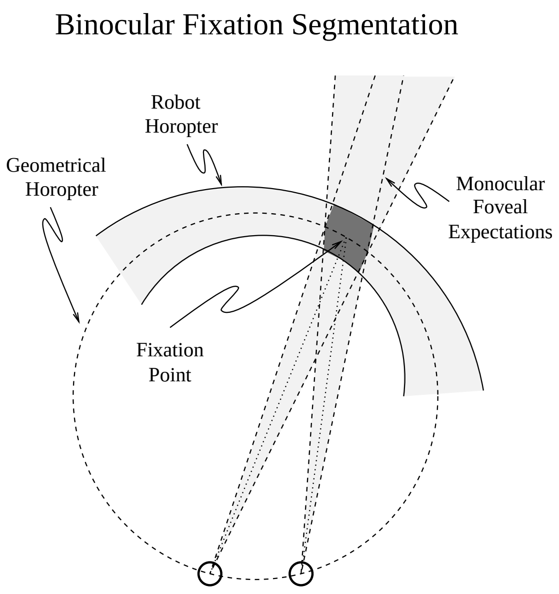


Figure 2: The result of fixating a target is that the object lies near the horopter, which is the set of world points whose disparity is zero. The stereo images of an object that lies near the horopter have a narrow range of disparities. In this top view of binocular fixation, the lightly shaded areas are the regions of space that are highlighted by foveal and disparity filtering. The intersection of these areas is shaded darker, and it corresponds to the area around the fixation point in which objects can be segmented.

at the target. Obviously, high resolution stereoscopic vision can only be achieved with both foveas directed at the point of interest simultaneously. An argument can be made for the ultimate necessity of non-uniform resolution, in order to provide both high resolution and a wide field of view [Tsotsos, 1988]. Thus future robot systems may be equipped with foveas. If so they will require pursuit and vergence systems for the same reasons that natural foveal visual systems require them. Work on spatially-variant visual sensors has begun [Van der Spiegel *et al.*, 1989; Rojer and Schwartz, 1990], so we can expect to use camera systems with foveas in the near future.

Motion Blur: Motion blur degrades the resolution of moving images. An image that moves over the retina will be degraded by motion blur, according to the integration time of the receptors (which have limited spatiotemporal resolution). It is easy to demonstrate this to yourself by moving your finger in front of some text and following your finger with your eyes. The background blurs as a result. If you fixate on the background, your moving finger will appear blurry. You can clearly read the text behind your finger.

Facilitating Stereo Fusion: Active vergence control can facilitate stereo fusion. By definition, the fixation point has a stereoscopic disparity of zero, and points nearby tend to have small disparities. This makes it possible to use stereo algorithms that accept only a limited range of disparities. Such systems can be very fast and are amenable to hardware implementation [Nishihara, 1984; Olson, 1991].

Indexical Behavior: A consequence of actively following the target with the sensor is that simple visual sensing techniques that are uniquely available in this situation can be used to segment the target. This opens the possibility of developing a suite of *indexical* behaviors [Agre and Chapman, 1987; Whitehead and Ballard, 1990] that the robot uses to interact with "the object of regard", such as `go-to-the-fixated-object`, and `pick-up-the-fixated-object`.

Gaze holding has several aspects that are accomplished in animals by a set of specialized controls that are mentioned here and sometimes referred to by analogy later. The *smooth pursuit* system tracks continuous target motion. The *vestibulo-ocular reflex (VOR)* and *otolith-ocular reflex* systems rotate the eyes to compensate for head motion that is detected by the vestibular and otolith organs (accelerometers) in the inner ear. The *opto-kinetic reflex (OKR)* and other *visual following* mechanisms stabilize the eyes on stationary scenes. The *vergence* system rotates the eyes in opposite directions, controlling the distance of the fixation point (at which the optic axes intersect). The *saccadic* system provides fast gaze shifts. In animals these systems cooperate in complex ways [Carpenter, 1988].

Vision alone provides the measure of how well gaze is being held steady on an object. Non-visual cues (*e.g.*, head motion) can be used to stabilize gaze on a fixed point in the world. Sometimes such cues are available with low latency and minimal sensing and processing, leading to higher performance from stabilization systems that use them. However, only vision can provide cues to the motion of a moving object. This work focuses on the problem of using visual cues alone to hold gaze on a moving object from a moving platform.

2.1 The Pursuit Problem and Related Work

First, consider taking a hint from primate pursuit models that might help us build a robot pursuit system. Krauzlis and Lisberger [1989] present a recent model of the smooth pursuit system. However, like most such models, their model does not address the extraction of the retinal slip of the the target. Lisberger, *et al.* [Lisberger *et al.*, 1987], argue that retinal slip and retinal image acceleration must be among the signals to which the pursuit system responds, based on analysis of the behavior of the system. However, the means by which the retinal slip of the target is distinguished from other optical flows experienced by the visual system remains unclear.

In the computer vision literature, a few attempts have been made to pursue moving objects based on optical flow, and most of them have used simple approaches to motion segmentation. *E.g.*, Lee and Wohn [1988] use stop-and-look (or static-look-and-move) tracking in which the target is expected to be the only moving object. Allen [1989] uses spatio-temporal motion energy to servo a robot arm-mounted camera on a target moving in front of a blank background. Tölg [1992] follows the traditional pursuit model [Young, 1971], with internal positive feedback and catch-up saccades. The motion-based target segmentation assumes the object is moving coherently. Jenkin [1991] uses stereomotion channels to estimate the target's motion. However, it is not clear how the trajectory detectors distinguish the motion of the target from the motion of the surrounding scene.

Other approaches assume the object is preselected and rely on view-dependent features; therefore they may be prone to slip off a rotating or deforming target. Corke and Paul [1989] smoothly follow the centroid of a blob that is translating in a fronto-parallel plane, by translating the camera in a fronto-parallel plane. The moments of the blob are computed on a binary image obtained by thresholding the greyscale input image. (The values of the moments that identify the target are known *a priori.*) This approach is likely to be brittle to changes in lighting, point of view, etc. Papanikolopoulos *et al.* [1991] smoothly tracks pre-selected features (image patches). It is not known how to select such features automatically, although this problem has been and will doubtlessly continue to be a subject of investigation [Matthies *et al.*, 1989; Thorpe, 1983]. Waxman *et al.* [1988] saccadically track a set of features whose spatial relationship is pre-selected, and the object is therefore identifiable. Pahlavan *et al.* [1992] integrate dual independent monocular stabilization with vergence and focus accommodation. The vergence system helps keep the monocular systems looking at the same fixation target. Each of these systems searches for matches of intensity patches so each component is susceptible to the problems related to view-dependent features, and the synergy of the systems must overcome this obstacle.

Clark and Ferrier [1988] present a binocular tracking system that saccadically tracks the appropriate conjunction of features that locate an object whose properties are known *a priori.* Their system has demonstrated saccades and position-servoed vergence to binocularly fixate a target as it moves in three dimensions. The left and right images are processed independently, and the location of maximum "salience" (the desired combination of features determined to be relevant) in each image is taken to belong to the target object. However, neither is there a guarantee that the feature values will be invariant to point of view, nor are the maximum salience values necessarily unique. Consequently, the correspondence problem may not be trivial enough to yield to salience images.

Wavering *et al.* [Wavering *et al.*, 1993] describe binocular smooth pursuit on a high performance head. High speed pursuit is achieved for frame-rate visual processing with careful calibration of the visuomotor system. The target's future position is estimated and forward kinematic models are used to follow the target accurately at ocular rotational speeds up to 6 rad/s ($345°/s$). To accomplish the visual processing at frame rates, only centroids of thresholded images were used to locate the target image.

3 Demonstration System

Demonstrations of vergence and binocular pursuit systems have been implemented on the Rochester robot head, which has three degrees of freedom (DOF), as shown in the photo of Figure 3. The head is mounted on the end of an industrial Puma six DOF robot arm with a of about $2m$ in radius. The cameras can move at speeds exceeding 7 rad/s ($400°/s$) so they can approximate human saccade speeds. The cameras tilt up and down together on a common platform, and pan independently from side to side, driven by three motors, one for the tilt platform and twin pan motors (Figure 4). A mechanical advantage of the Rochester

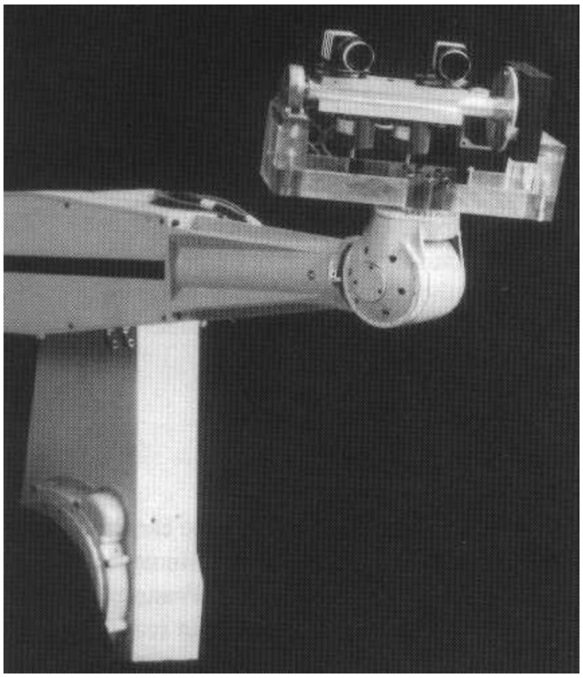


Figure 3: Portrait of the Rochester robot head.

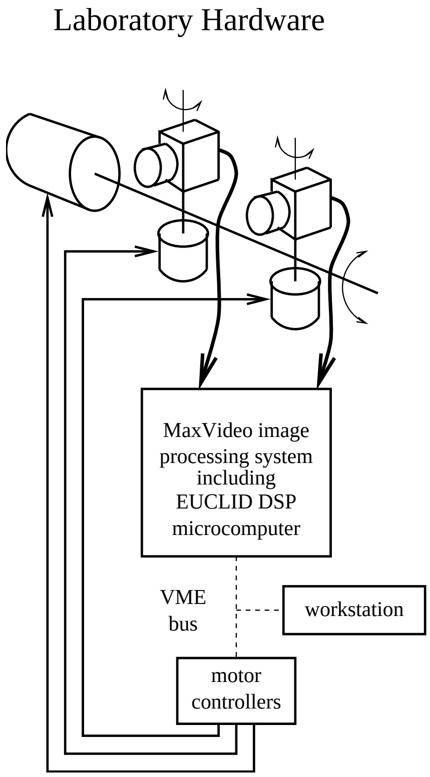


Figure 4: Hardware configuration used for pursuit and vergence control experiments on the Rochester Robot.

head design is its simplicity: the compact mechanism and fairly direct linkages facilitate rapid saccades.

Figure 4 describes the configuration of laboratory hardware on which the demonstration systems are implemented. The MaxVideo image processing system captures images from the cameras and performs the bulk of the image processing. It is coordinated by a Sun workstation, which interprets the visual signals to produce the appropriate motor commands for the motor controllers.

The workstation sends commands to the motors via intelligent stepping motor controllers that allow the control program to issue commands in terms of absolute position, relative position, velocity or velocity profile. The ability to issue buffered velocity commands enables the control program to generate smooth movements without paying constant attention to the motors.

4 Vergence

The vergence system is half of the binocular pursuit system. This section describes the vergence system previously reported in [Olson and Coombs, 1991]. It is described briefly here for completeness. Vergence of eyes or cameras that share a common tilt plane results in the optic axes intersecting at some point in the tilt plane. The vergence angle of a binocular system is the angle between the optic axes of its cameras. By analogy with primate visual systems having a central high-resolution fovea, we say a camera foveates a target if the target is at the center of the visual field.

The vergence system can be thought to control the distance from the cameras to the fixation point along some specified gaze direction. The vergence problem can be defined as that of controlling the vergence angle to keep the fixation distance appropriate for the current gaze target. Since the desired vergence angle is directly related to target distance, any sensory cue to depth or depth changes may be useful to the vergence system. In humans, there is a strong link between the accommodation (focusing) and vergence systems. The most obvious direct cue is binocular disparity, but other depth cues (such as motion, texture, and shading) can also be used, as can information about depth changes (*e.g.* measured or predicted self motions, dilations or contractions of the visual field).

4.1 Vergence Error

Binocular disparity is one of the most powerful visual cues to vergence error, and the mapping from disparity to vergence error is simple. Disparity measurement has been studied extensively in the context of stereo depth reconstruction. Unfortunately most of the disparity estimators used for stereopsis are too powerful and slow for use in real-time vergence application. We use a correlation measure between sub-sampled left and right images, or "foveal" windows of such images.

The *power cepstrum* of a signal is the Fourier transform of the log of its power spectrum. It was introduced in [Bogert *et al.*, 1963] as a tool for analyzing signals that contain echoes. In this application the right and left images are concatenated into a single rectangular image and the power cepstrum of the concatenated image is computed using an discrete Fourier transform on a digital signal processing chip.

The filter yields a disparity histogram, and the vergence error is measured as the (x, y) location of the highest peak in disparity space—it should be brought to $(0, 0)$. The cepstral filter has the advantage of separating the disparity peaks from the central correlation peak and (more deeply) of attenuating low-bandwidth signals, hence acting like a combination of "interest operator" and matched filter correlator [Olson and Coombs, 1991].

Since this error measure is based on *what* disparities are present but not *where* they are, the target should dominate the scene (be associated with the most common disparity). This in turn can be accomplished by spatial windowing and pursuit in dynamic scenes with many distractors (we describe this approach in Section 6). However, another more general approach is to track the disparity peak of interest regardless of its size, keeping it at zero disparity.

4.2 Vergence Control

The workstation finds the maximum peak in the cepstral filter output, converts the pixel disparity to angular coordinates, and applies the control law to issue identical and opposite (symmetric) vergence velocity commands to the camera motors. Symmetric vergence allows us to decouple the vergence and tracking controls for simplicity, but it is not necessary. The workstation issues the motor commands *after* initiating the next digitization in order to allow digitization to proceed concurrently with motor control. This causes a slight delay in issuing the motor commands, but permits a substantially higher overall sampling rate. The loop consistently takes three frame times to complete. Thus, the system achieves a servo rate of 10 Hz.

A proportional-integral-derivative (PID) controller (*e.g.* see [Dorf, 1980]) is used in cascade with the camera motor to generate oculomotor responses to reduce the estimated disparity. The controller gains were chosen empirically to obtain slightly underdamped response, resulting in a small overshoot in the step response. The demonstration system's response to a sinusoidal input is shown in Figure 12. The response to sinusoidal stimuli of frequencies up to 2 Hz suggests that the system has second order characteristics, and its constant time delay produces the expected linear phase shift.

5 Zero Disparity Filtering

With the vergence system keeping the cameras verged on the target, the fixation target's retinotopic location can be found by disparity filtering. Features that have no stereo disparity can be detected in real time using a disparity filter. The region of space that contains objects that project onto the retinas with no stereo disparity is called the *horopter*. Ideally, disparity filtering can thus be used to isolate a target at a given range from its foreground and background.

Several definitions for the horopter have been proposed, but we will use one adapted from [Reading, 1983, p. 88]: the horopter is the set of points in three dimensional space which project to corresponding points in both eyes (or cameras). Note that it is uniquely identified with the fixation point (intersection of the two primary lines of sight) and it is the boundary between crossed and uncrossed disparity. *I.e.*, for every possible fixation point in space, there is a unique horopter, and the binocular disparity of every point on the current horopter is zero.

The 3-D shape of the horopter is at least rather complicated (and at worst may not be well-defined), since it involves vertical, as well as horizontal, disparities. However, the horopter's 2-D shape in the tilt plane has received considerable attention (see Figure 5). The *geometrical horopter* is the circle (also known as the Vieth-Müller circle) passing through the two nodal points of the cameras and the fixation point. This is the shape of the horopter that would result if the corresponding point in each retina represented an equal angle with respect to the fixation point. The *human em-*

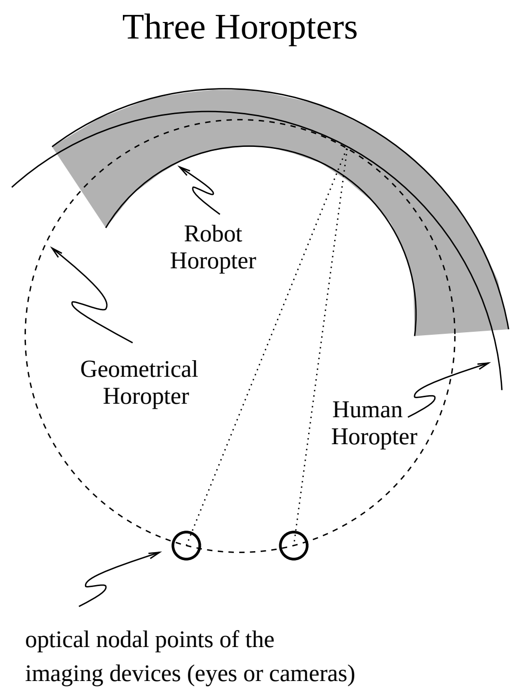


Figure 5: Horopters. The geometrical horopter, also called the Vieth-Müller circle, is the circle that passes through the optical nodal points of the cameras and the fixation point. The arc labeled human horopter depicts the empirical horopter, which is determined experimentally. Several factors have been identified that can influence this deviation. In contrast to the human horopter, the robot's horopter is defined by the points that stimulate the zero disparity filter (ZDF), in accordance with our adopted definition of horopter. The thickness of the robot horopter depends on the tightness of the tuning of the ZDF. This depiction is intended to be qualitative only; it should be noted that rigorous efforts have not been made to precisely determine the shape of the robot's empirical horopter.

pirical horopter is measured by several techniques with varying results. The empirical horopter is consistently found to deviate from the geometrical horopter in approximately the way that is depicted in Figure 5. The reasons for the difference include the spacing of "corresponding" points in the retina and the visual cortex, optical distortions, and distortions of the shape of the eye when it is rotated to different eccentricities. (See [Howard, 1982] for a concise summary. More extensive surveys can be found in [Reading, 1983; Shinkman *et al.*, 1985; Tyler and Scott, 1979].)

A simple edge-based zero disparity filter (ZDF) was designed and implemented to run at frame rates on MaxVideo image processing hardware [Coombs and Brown, 1991]. Essentially, it is the logical AND of stereo vertical edge images. The features used by the filter are vertical edges since they are easily detected by convolution and they can give useful information about horizontal disparity. (Clearly horizontal edges provide no helpful information about horizontal disparity, since long horizontal edges can match over much of their length even with substantial disparity. Only their *ends* can be compared to find horizontal disparity.) The results of simple edge-based filtering can be seen in Figure 6. The details of this algorithm and others that are intended to reduce susceptibility to aliasing are described in [Coombs *et al.*, 1992], and more sophisticated frequency-based disparity filtering is described in [Olson and Lockwood, 1992]. The windowed output of the ZDF is the input to the smooth pursuit system. Thus, the pursuit system is given the retinotopic location of the target object when the system is properly verged on it.

6 Smooth Pursuit Control

There are two natural and obvious measures of target following performance: position error and velocity mismatch. Restated, a pursuit system could attempt to center the target image, or at the opposite extreme, the system could try to stabilize the target's image on the retina (reducing slippage) by matching the target's velocity without regard to the retinal position of the target's image.

Note that the goals of image-centering and slip-minimizing conflict in a linear control system since smooth camera movements can only improve one of these measures at the expense of the other. Suppose the system is following with a positional lag and a velocity mismatch. If the image is to be stabilized by matching velocity, any accumulated positional error must be tolerated since an effort to correct the position with smooth movements must slip the image across the retina. Similarly, if the image of the target is to be held in a desired location with smooth movements, image stabilization must be sacrificed because a mismatch in velocity is required to change the image position.

One approach to this problem is give precedence to one of these goals. Thus one simple pursuit system could attempt to keep the target image centered without regard to the image slip required. Another simple system could try to minimize target slip. However, it is generally believed that the primate pursuit behavior consists of a combination

(a) (b) (c)

Figure 6: Zero disparity filtering of the scene shown in stereo images (a) and (b) yields output shown in (c).

of both smooth servo control that matches velocity to minimize slip and catch-up saccades that recenter the target image when it deviates too far from the fovea. A more realistic model of the smooth following component includes a small response to position error in addition to retinal slip.

This is a clever nonlinear solution to the dilemma of how to minimize both velocity and position error simultaneously. As previously noted, smooth movements alone cannot achieve both goals, and certainly saccadic movements cannot reduce motion blur since they don't match velocity. The catch-up saccades perform the important function of correcting accumulated position error while introducing minimal target slip, and target slip is minimized by the smooth component that matches the target velocity with the camera velocity.

Unfortunately, implementing saccades in a binocular system requires careful attention, especially to the vergence angle. The various complications involved in coordinating saccades have led us to employ the simple strategy of using only smooth camera movements in the demonstration system.

One of the most challenging problems faced by visuomotor control systems is coping with the delays in the system. Delays can cause a system to be unstable. For instance, if a feedback system can respond more quickly than it can measure the error, it may respond to old error signals and over-react. Consider driving a car on a slippery road. The driver turns the wheel, but the car does not immediately change its course, so the driver turns the wheel further and consequently over-steers the car. Since cameras are quickly maneuverable and visual processing is slow, visuomotor systems face a similar control problem due especially to the delays of visual processing. There are a couple of obvious ways to prevent overshooting responses. One approach is to model the delays present in the system and anticipate the response of the system with an internal model of the the visuomotor system. Some simple predictive control is described here, and some simulation investigations of predictive gaze control are described in [McDonald *et al.*, 1983; Brown, 1990; Brown and Coombs, 1991]. Experiments using prediction to improve pursuit performance are reported in [Wavering *et al.*, 1993] A simpler approach reduces overshooting simply by reducing the responsiveness of the system enough to make it stable. A combination of these methods is used in the pursuit system demonstration.

The demonstration pursuit system uses independent PID controls for pan and tilt, and we have experimented with $\alpha-\beta-\gamma$ prediction (a constant acceleration model in world coordinates, not image coordinates) [Bar-Shalom and Fortmann, 1988]. The $\alpha-\beta-\gamma$ filter is used to apply a simple linear model to the target in order to predict its future state. If the error signal can be successfully predicted, the effect of processing latency can be mitigated by controlling the system with predicted error signals. Once the visual signal is processed, the observed error is compared with the predicted error. The predictive filter is also useful to cope with signal dropout: the target dynamics can be run forward and the predicted position of the target can be tracked until the signal reappears or the system gives up.

Using the $\alpha-\beta-\gamma$ predictor in the loop to predict the delayed signal can produce more accurate tracking. Figure 7(a) shows the camera movement and tracking error as the camera tracks the image of a dark object in approximate harmonic motion with a period of about five seconds. The object is rotating in a plane and thus its distance from the camera varies and its velocity is not purely sinusoidal. The error is measured as the off-axis angle of the centroid of the object's image. There is approximately 80 ms delay in the system. The small phase difference between the sinusoidal waveforms of the target and camera motion induces a surprisingly large error. In Figure 7(a) an $\alpha - \beta - \gamma$ filter is used ($\lambda = 1$) with no predictive advance, so the tracking signal is smoothed somewhat. The target signal's *maneuvering index*, λ, is the ratio of the standard deviations of process noise and measurement noise. Choosing $\lambda = 1$ expresses equal confidence in the quality of the observed data and the prediction. In Figure 7(b) the filter extrapolates the signal 50 ms into the future. The result is livelier tracking, but error is reduced. In fact, more advance destabilizes tracking, even though the advance of 50 ms does not account for the entire 80 ms latency in the system.

7 Binocular Pursuit: Combining Pursuit and Vergence

Figure 8(a) shows how vergence, zero disparity filtering, and tracking work together. Vergence and tracking use foveally-processed visual signals.

The visual processing for the gaze holding system is implemented in real time almost entirely on a Datacube MaxVideo image processing system (Figure 8(b)). The processing begins by digitizing a stereo pair of images from the robot head's cameras. The cameras are synchronized so the images are taken simultaneously.

The goal of the binocular pursuit system is to generate smooth camera movements that correct both gaze angle and vergence errors. The binocular pursuit control loop consists of three stages: digitization, error estimation (both gaze angle and vergence), and error correction (of both gaze and vergence angles). The timing of the system is sketched in Figure 9. Digitization is done under control of the workstation using the MaxVideo digitizers, convolvers and frame stores (one each per camera). In addition, the zero-disparity image is thresholded by the FeatureMax and the list of above-threshold features is stored in its memory. It takes between one and two RS-170 frame times (33 to 67 milliseconds), depending on how much time remains in the current video frame when the command to acquire the next frame is issued. The workstation is free to do other things during digitization. Once the images are available in the frame store, the workstation signals EUCLID to extract the images from the frame buffers and estimate the disparity. This process takes approximately 59 milliseconds, after which EUCLID places the disparity estimate in a known location in shared memory and issues an interrupt to signal completion. The workstation converts the pixel disparity to angular coordinates by multiplying it by an empirically determined constant. While EUCLID is estimating the disparity, the workstation computes the centroid of the zero-disparity features. The length of time required depends on the number of zero-disparity features. Once the gaze and vergence errors are estimated, the workstation applies the control law to issue the appropriate velocity commands to the camera motors. The workstation issues the motor commands after initiating the next digitization in order to allow digitization to proceed concurrently with motor control. This causes a slight delay in issuing the motor commands, but it permits a higher overall sampling rate. The loop commonly takes 150 ms (four and one-half frame times) to complete. Thus, the system achieves a servo rate of 7.5 Hz.

7.1 The Controller

The pursuit system uses a PID controller for each of pan, vergence and tilt of the robot's gaze, as shown in Figure 10. The vergence system's error is found in the binocular disparity of the foveal images. The error signals for the pursuit controllers are the horizontal and vertical displacements of the target from the centers of the foveae.

The gaze parameters, $\vec{\theta}$, map fairly directly, though not identically onto the mechanical degrees of freedom, $\vec{\phi}$, of the robot head. There is a single tilt motor, so

$$\theta_{tilt} = \phi_{tilt}.$$

The situation for pan and verge angles is sketched in "top-view" in Figure 11. Pan angles are counter-clockwise positive viewed from below, with zero straight ahead. Tilt is CCW positive viewed from the right, with zero again straight ahead. The pan and verge angles are related to the left and right pan camera angles by

$$\begin{aligned} \theta_{pan} &= \frac{1}{2}(\phi_{right} + \phi_{left}) \\ \theta_{verge} &= \phi_{right} - \phi_{left}. \end{aligned}$$

These equations relate the static angles. However, the motor controller must convert the pan and verge velocity commands to left and right pan motor velocities. Differentiating with respect to time,

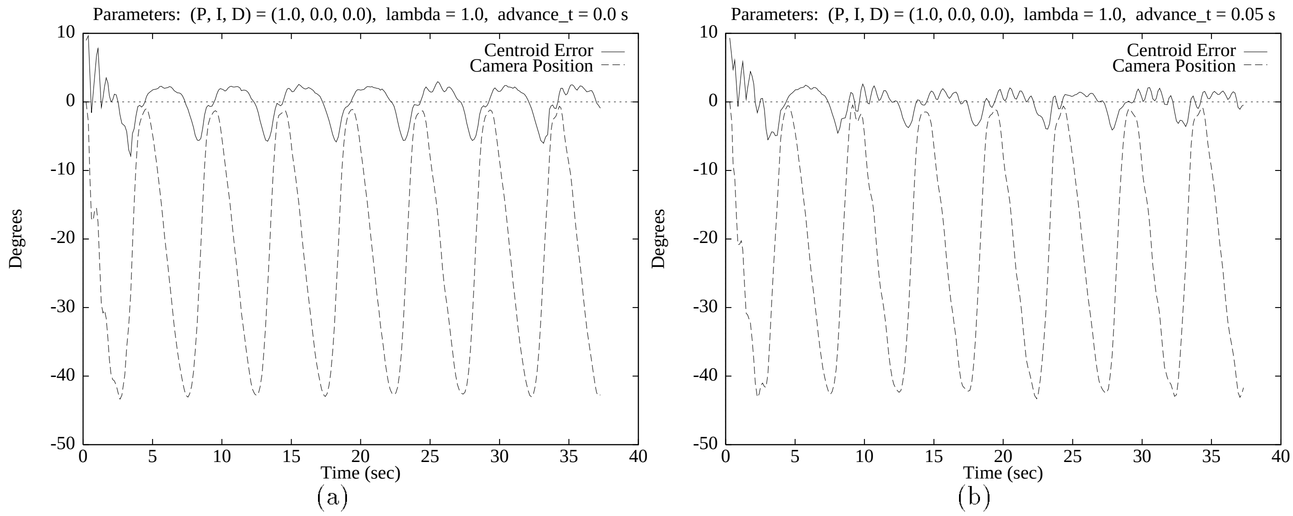


Figure 7: Camera motion and error when tracking an object in approximate harmonic motion. (a) Delay of approximately 0.08 s induces small phase lag but large tracking errors. (b) Camera motion and error using $\alpha - \beta - \gamma$ predictor to advance the signal by 0.05 s.

we obtain the motor velocities. As one might expect, the pan velocity is transmitted to both camera pans, and the vergence is split evenly between them.

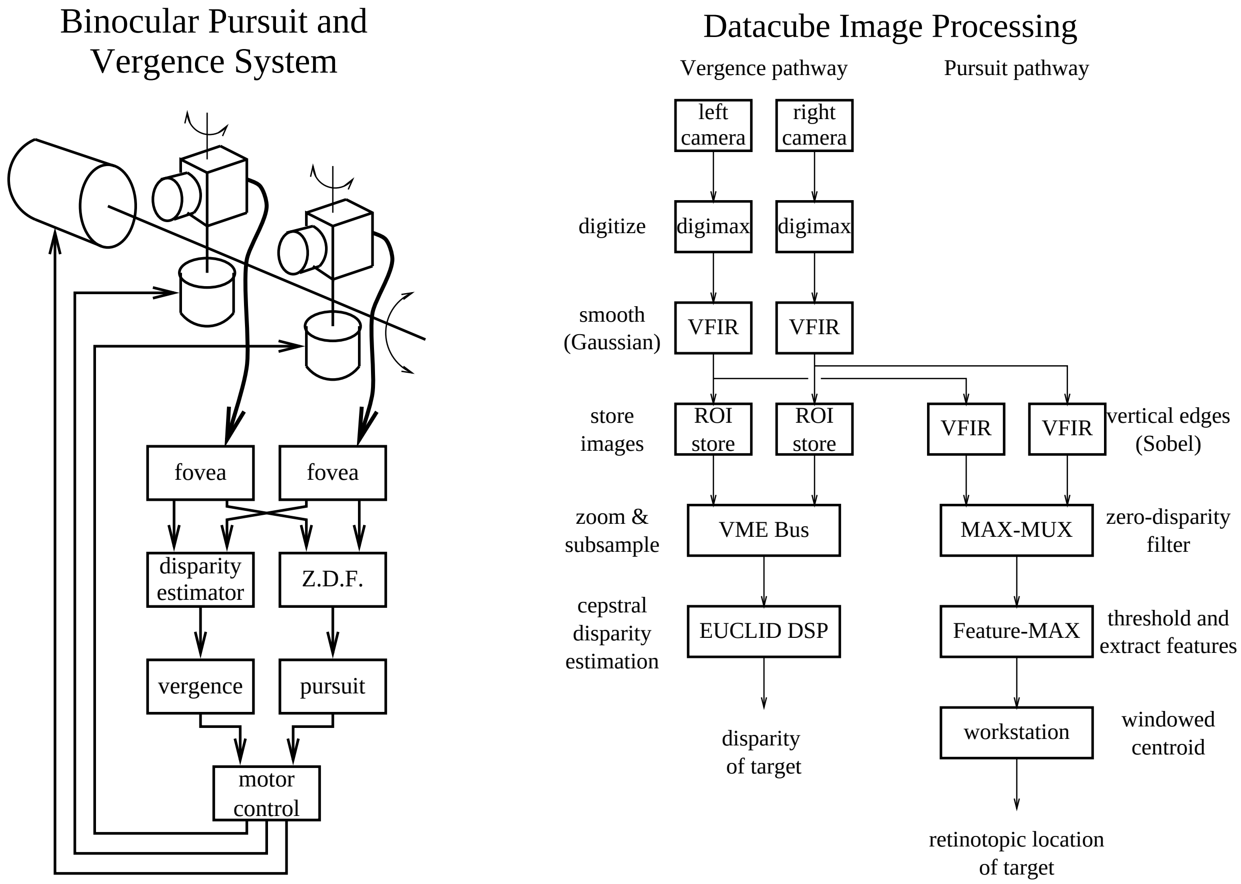
$$\begin{aligned} \dot{\phi}_r &= \dot{\theta}_p + \frac{1}{2}\dot{\theta}_v \\ \dot{\phi}_l &= \dot{\phi}_r - \dot{\theta}_v \\ &= \dot{\theta}_p - \frac{1}{2}\dot{\theta}_v . \end{aligned}$$

7.2 Gaze Holding Performance

Figure 12 shows a stereo robot's-eye view of a typical setup for a target object moving in a horizontal circle through a field of distractors, and the measured camera pan, tilt and convergence angles and visual errors. The robot views the scene from slightly above the plane of the bunny's rotation. These measurements were recorded from a run with the target rotating at 0.1 Hz, and the pan angle trace reveals rotational camera velocities as high as $0.23\,\mathrm{rad}/s$ $(13°/s)$. The cameras lagged behind the apparent target velocity, as the non-zero observed retinal error indicates.

The limits of performance of the system depend on both the observation of the target signal and the response of the control and motor system. The observation and control problems are tightly interdependent in a system of this nature, since control responses cause changes in the observed signals.

The 3-D tracking window must be large enough that the target does not maneuver out of it and yet it must be small enough to exclude clutter. If the tracking window size must be chosen to meet a firm target containment or clutter exclusion objective, the limit of performance of the other is then determined. In these experiments, the retinotopic extent of each window was about twice the area of the target's image at the object's nearest point. A tracking window of this size was sufficient to keep the target from escaping horizontally or vertically. The vergence system responds to images that are within both retinotopic tracking windows because it uses correlation to estimate the most prominent disparity in the windows. Therefore the system must not allow the visual target to escape either window. The target is most likely to escape only one window by moving in depth. As long as the object does not escape either window, vergence can continue to respond to it even if it escapes the ZDF window and the pursuit signal drops out. The zero disparity filter responded to targets with disparities within approximately $\pm 3/4$ pixel, though the ZDF was not defined in those terms, since it does not estimate disparity. The design of the ZDF and its parameters influence the depth of its responsive region. The trade-off for the size of the ZDF's sensitive region is similar to that for the retinotopic size of the windows. If the target exits the sensitive region, the pursuit system's target signal will drop out, yet if the sensitive region is very large, distracting signals can creep in and confuse pursuit. The extent of the sensitive re-



(a) (b)

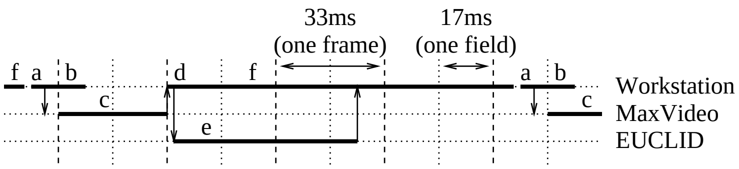
Figure 8: (a) Binocular Vergence and Pursuit system. (b) Datacube Image Processing. Nearly all of the visual processing is carried out on a Datacube MaxVideo image processing system.

gion of the ZDF is distance dependent, since it is defined by image disparity. The region is narrow for near fixation distances, and it gets wider as the fixation distance increases. The retinotopically defined "cones" also increase in cross-sectional area with increasing distance. Thus the 3-D volume of the tracking window increases as the fixation distance increases.

The scene clutter that can be tolerated is primarily a function of the volume of the tracking window, but there are dynamic affects that permit the system to withstand a small amount of clutter in the tracking window. If a distractor passes through the tracking window, the momentum of the estimation, motor and control systems limits the system's response to transitory signals of this type. The ZDF signal will be disturbed temporarily, but the disturbance may be insufficient to pull gaze off the target object. Similarly, the target can be occluded briefly and the prediction and momentum of the system may be able to sustain tracking temporarily, tolerating the brief signal dropout.

The spatiotemporal characteristics of the estimations are also important. The vergence system's disparity estimation will suffer for a target affected by motion blur since high frequency components of the signal are lost. The object's signal in the disparity landscape will be weaker, and precision will be diminished. The pursuit system's ZDF response to the target will be weakened as well, and the signal may well drop out. Loss of medium and high frequencies will reduce the quality of the match. Neither vergence's disparity estimation nor pursuit's ZDF will perform well on a blurred image of the target. If the object's retinal slip is large, the image will be blurred by the

Pursuit and Vergence Timing Chart



a --- Set up frame buffers to capture images
Set up feature extractor to collect ZDF pixels

b --- Read motor angles
(Update gaze-vector display overlay)
Estimate retinotopic target location
Issue motor commands
(Update crosshair display overlays)
(Issue stimulus motor commands)

c --- Frame buffers capture images
Feature Extrator collects ZDF pixels

d --- Fork cepstral disparity estimator on EUCLID

e --- EUCLID grabs subsampled images and estimates disparity

f --- Collect ZDF pixels and calculate centroid

NB: times are approximate, for illustrative purposes.

Figure 9: Binocular Pursuit Loop Timing Diagram.

motion, and signal dropout will result. Eye movements as well as target motions affect retinal slip, so aggressive eye motions were avoided.

The tracking speeds achieved by the system are a function of control bandwidth (which depends on delays and servo control rates) and target signal quality (*e.g.*, steadiness). The system can observe the target signal only when the object remains within the 3-D tracking window (retinotopically, in disparity, and spatiotemporally). Assuming the target signal does not drop out, tracking performance depends on the ability of the control system to respond to target motion. The control systems were tuned to be slightly underdamped, to respond as briskly as possible with minimal overshoot.

Even a small phase lag contributes significantly to positional error. The observed errors in pan angle of up to 0.05 rad (3°) were near the limit that could be tolerated given the window size and control bandwidth. Control bandwidth is limited by the delay (one time interval) and the servo control rate of 7.5 Hz. The control gains, already slightly underdamped, could not be increased much without destabilizing the system. For instance, overly brisk response of the vergence system leads to intermittent dropout of the zero-disparity silhouette if the amplitude of the ringing gets too large. Conversely, overly damped vergence response results in loss of track of the target if the object is able to slip away in range too fast for the vergence system to keep the horopter on it. If the system responds too briskly to an error, the resulting motion blur and overshooting (which take the tracking window off the target) cause the signal to drop out. When the signal reappears (assuming it remains within the tracking window spatially), the error is large. This introduces large step inputs in the target signal, resulting in unstable or marginally stable following as the system jumps toward the observed target location. Even if the system response itself is stable, the motion blur induced by the vigorous following can degrade the image of the object enough to impair observation of the target, as previously noted.

In order to illustrate the function of each component of the gaze holding system, selected control components were removed from the system, and

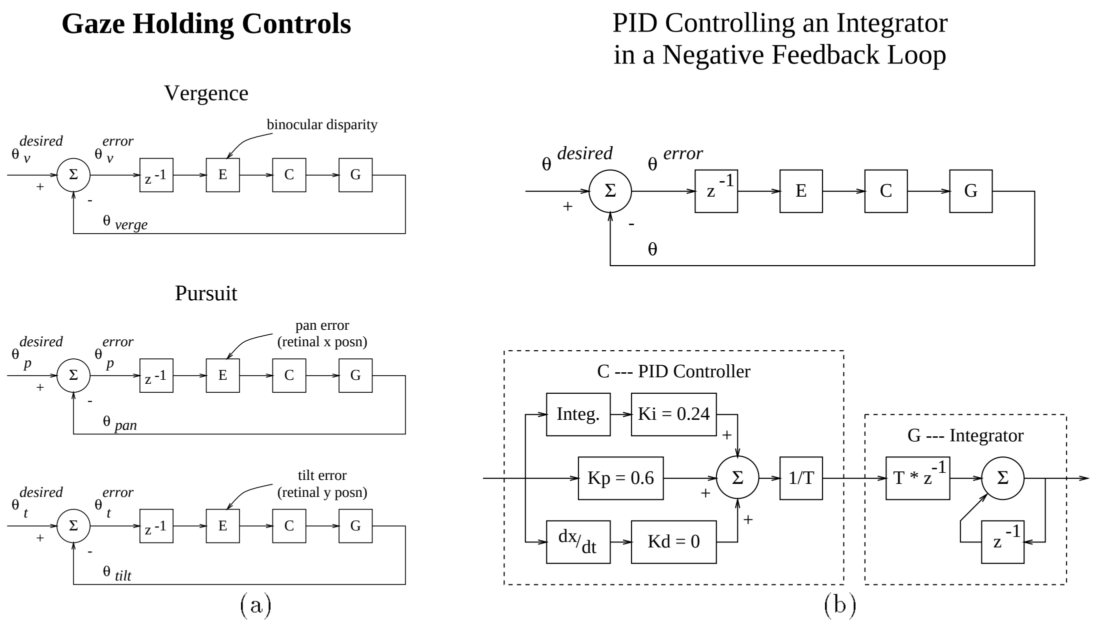


Figure 10: Gaze Holding Control Systems: The vergence and pursuit controls are driven by three independent feedback controllers, one each for pan, tilt, and vergence (a). The vergence system's error is found in the binocular disparity of the foveal images. The error signals for the pursuit controllers are the horizontal and vertical displacements of the target from the center of the fovea. Each of the controllers is a PID controller, and each degree of freedom of the gaze control system can be modeled approximately by the system shown in (b).

the resulting behaviors are compared with the behavior of the complete system (Figure 13). The first trace shows the behavior of the unimpaired system for comparison. This pan trace is similar to the pan trace in Figure 12(c). The subsequent traces show performance without peripheral suppression (*i.e.*, without narrowing attention retinotopically), with fixed vergence angle, and without disparity filtering (*i.e.*, without narrowing attention in depth).

The second trace illustrates the loss of track on the target object that results without foveal attention windowing (or peripheral suppression). The system's gaze bounces around the scene from object to object as its attention is drawn by various objects. Recall that the pursuit system's target object is selected by the vergence system and the disparity filter. The vergence target is determined by the strongest signal in the disparity space (*i.e.*, the single disparity that best describes the difference in the stereo images). Attention is no longer narrowed to the retinotopic neighborhood of the target, so the disparity competition is opened up to the entire scene. (A background with rich texture would likely command the robot's attention eternally.) The disparity competition is influenced by the bandwidth and the spatial extent of the of the image patches involved. As the scene changes due to the bunny's motion and the robot's camera movements, different objects win the robot's attention in turns.

The third trace demonstrates the system's inability to track the target if the vergence angle is held constant. By disabling the vergence system, the horopter is held at approximately a fixed range. (The shape of the horopter deforms a bit as the eccentricity of gaze changes when the cameras pan from side to side. However, it remains essentially at a constant range.) Since the pursuit system only "sees" objects in the horopter, the system drifts until it either locks onto a target that lies in the horopter or drifts to a stop. This trace is taken after enough time has passed for the target to locate a stationary object in the horopter.

The final trace shows how the system is distracted by objects at all distances when zero-disparity filtering is eliminated. Thus, the system follows the centroid of all the edges visible in the scene. (The vergence system has no influence in target selection in this situation. For stability reasons, peripheral suppression was also eliminated

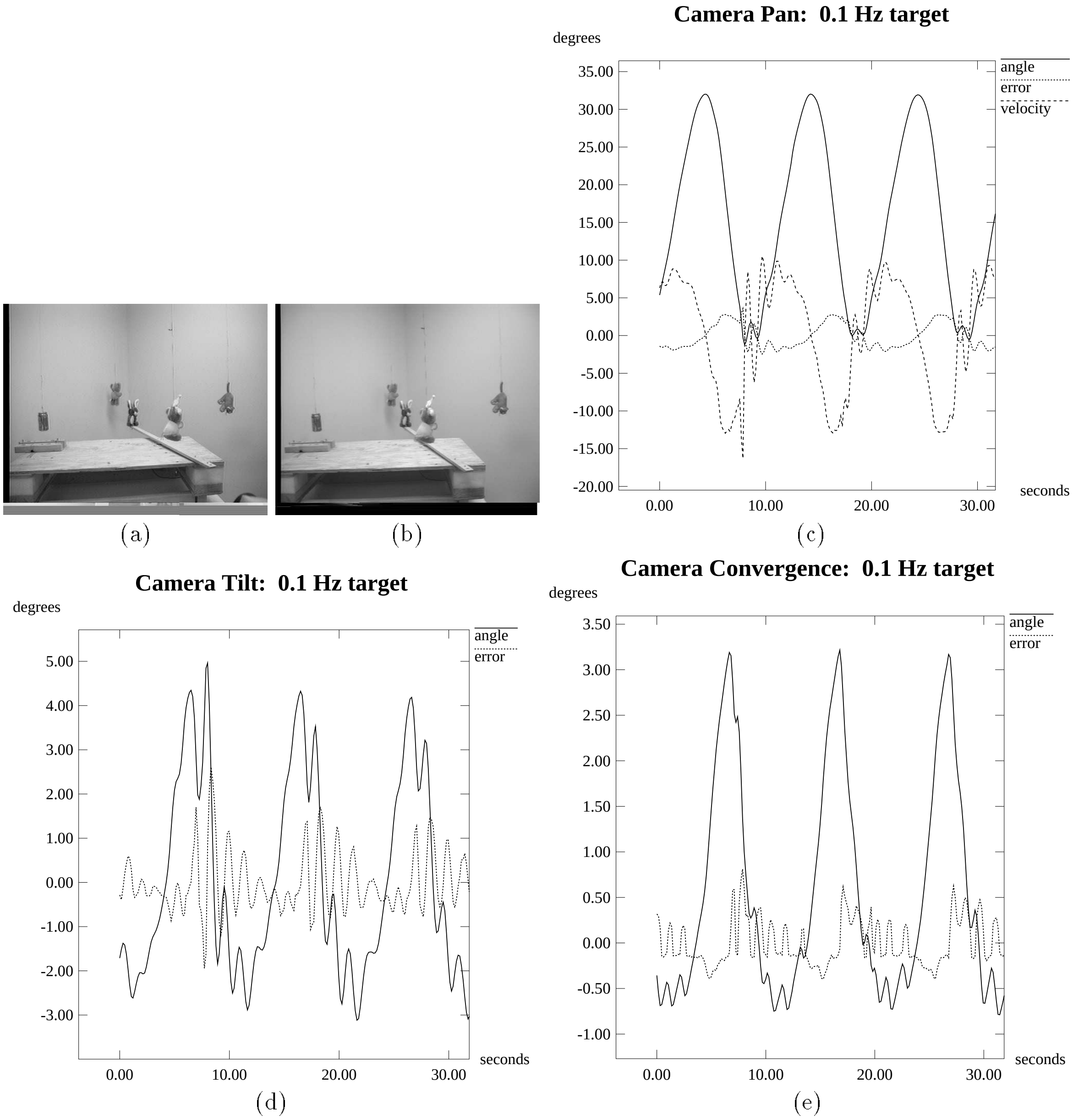


Figure 12: Gaze holding camera traces. Measured traces of the pan (c), tilt (d) and vergence changes (e) show the performance of the gaze holding system in following a target moving in 3-D through a field of distractors. A robot's-eye stereo view of a typical stimulus setup is shown in (a,b).

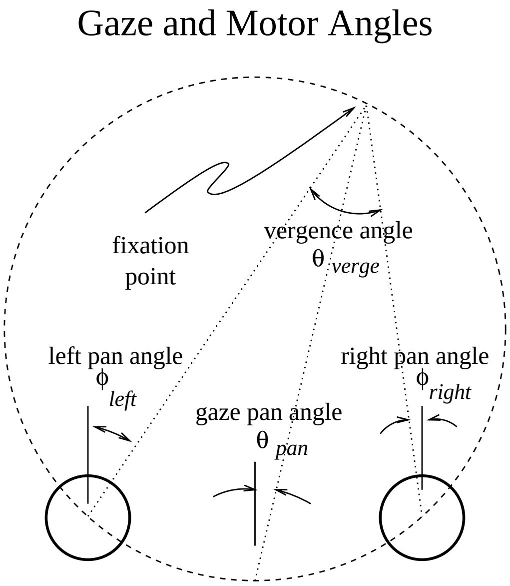


Figure 11: The relation between gaze angles $\vec{\theta}$ and motor angles $\vec{\phi}$ is sketched in this view of the "gaze plane" defined by the camera and the fixation point. Pan angles are counter-clockwise positive viewed from below, with zero straight ahead. Tilt is CCW positive viewed from the right, with zero again straight ahead.

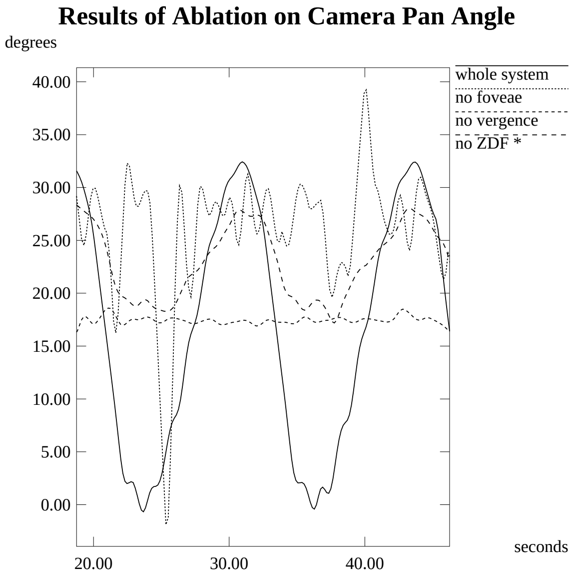


Figure 13: These camera traces result from ablation of various system components. The first trace shows the behavior of the unimpaired system for comparison. This pan trace is similar to the pan trace in Figure 12(c). The second trace illustrates the loss of track on the target object that results from the removal of foveal narrowing of attention (or peripheral suppression). The third trace demonstrates the system's inability to track the target if the vergence angle is held constant. The final trace shows how the system is distracted by objects at all distances when zero-disparity filtering is eliminated.

for this experiment. When foveal windowing was used, the abrupt entrance and exit of features in the window resulted in large impulses in the target location estimate. The robot responded so vigorously to these signals that its behavior was unstable. Including extra-foveal edges in the "target" centroid calculation diluted the destabilizing effect of features entering and leaving the "foveal" area.) The effect of the bunny's motion on the centroid can be clearly seen as the trace is essentially a damped and attenuated version of the unimpaired performance of the system. That is, the centroid of the scene edges follows the bunny around, but its motion is smaller than the motion of the bunny's edges alone. In a sufficiently cluttered scene, a sufficiently small target would have negligible influence on the scene centroid.

These ablation experiments show that each piece of the system contributes to the performance, and it is the combination of the simple components that allows each part to be simple. In additional experiments the Puma arm moved the robot head during the pursuit task. Since gaze-holding is purely driven by visual feedback we would expect the system to be robust to head motion, and indeed it proved to be.

8 Conclusion

A system is presented that holds a robot's gaze on an object while both are moving. The system uses only precategorical visual processing, *i.e.*, it does not require the ability to recognize the target. The approach is based on the premise that visual processing and visual behavior should be considered as tightly interacting components of visual perception. By exploiting constraints that can be maintained by active control of camera movement, simplified visual processing is sufficient to hold the robot's gaze on the visual target.

A system implemented on the Rochester robot demonstrates the idea, holding the binocular gaze of a moving robot on an object that moves through a cluttered scene. The vergence and pursuit components of the system cooperate to simplify the visual processing required, as illustrated by Figure 2. The vergence system controls the vergence angle between the cameras to minimize the stereo disparity of the foveated target. A fast correlation-based technique estimates the most prominent disparity in foveal stereo images and the vergence system converges the cameras to the distance of the visual target. The pursuit system controls the pan and tilt angles of the cameras to center them on the object on which the cameras are verged. A simple zero-disparity filter provides the target location to the pursuit system by detecting features that have no stereo disparity. Each of the pursuit and vergence systems maintains the quasi-invariant that is required by the other. The vergence system provides a zero-disparity target for pursuit, and pursuit keeps the target foveated for vergence. The system is able to maintain these quasi-invariants in the retinal images by its active control of the camera angles.

The chief contributions of the present work are:

- That localizing visual attention in 3-D space enables simple precategorical processing to suffice for holding gaze on a visual target. It is demonstrated (in Section 5) that simple binocular cues are able to distinguish a relatively near target from scene clutter. Other precategorical cues with physical 3-D interpretations should prove useful in situations that do not provide binocular cues.
- A demonstration of harnessing the interaction between sensing and control to support sensory-motor behavior, which is also exemplified by the ability of binocular fixation segmentation to pick out the target during gaze holding.
- A demonstration of the symbiotic cooperation of simple subsystems, *e.g.*, pursuit and vergence (in Section 7). The contribution of selected components is illustrated by disabling each of them in turn.

The limitations of the system described here should be clear. First, neither is any attempt is made to identify an interesting target for initiation of tracking, nor is the system able to acquire fixation of a non-foveated moving target. Both of these activities are crucially important for the completeness of a gaze control and tracking system. Second, binocular cues provide reliable information only for near targets (with, say, two or three "arm's lengths", *i.e.*, about 20 times the interocular separation) because there must be sufficient range discontinuities to permit disparity differences to distinguish the target. Biological pursuit systems have capabilities of these kinds that the demonstration system lacks. Further, the normal mode of pursuit operation for animal systems is not likely to be precategorical; rather, the creature probably is able to recognize any object that it is tracking, and it is not clear how much and what sort of influence this has on the process of tracking an object. Certainly any advantages in robustness and performance that can be gained by the use of such knowledge should be sought by robot visual systems as well.

Nevertheless, the central claim is that localizing attention in 3-D space makes simple precategorical visual processing sufficient to hold gaze. The precategorical nature of the visual sensing means the algorithms are simple and do not require delicate tuning. Vergence control is not even required since the disparity filter can be designed to respond to any finite disparity. This basic idea can be used to advantage when holding gaze in dynamic situations.

Acknowledgments

The work and its presentation have benefited from discussions with Dana Ballard, Randal Nelson, Tom Olson, Ray Rimey, Mike Swain, and Lambert Wixson, and from the thoughtful suggestions of Martin Herman, Ernest Kent, Russell Kirsch, Ron Lumia and other reviewers. Mike King and Larry Snyder shared their expertise in biological eye movement systems and control systems. Tim Becker deserves many thanks for his help with monocular pursuit experiments using prediction. Dave Tilley and Ray Rimey enhanced the MaxVideo programming environment with Zebra and Zed [Tilley, 1990], and Tim Becker made the Puma control software easier to use.

This material is based upon work supported in part by the National Science Foundation under Grants numbered IRI-8903582, CDA-8822724, and IRI-89220771, and by ONR/DARPA research contract number N000114-82-K-0193. Final preparation of the manuscript was supported in part by NIST.

References

[Agre and Chapman, 1987] Philip Agre and David Chapman, "Pengi: an implementation of a theory of activity," In *AAAI*, pages 268–272, 1987.

[Allen, 1989] Peter Allen, "Real-Time Motion Tracking Using Spatio-Temporal Filters," In *Proc. of the DARPA Image Understanding Workshop*, 1989.

[Bar-Shalom and Fortmann, 1988] Yaakov Bar-Shalom and Thomas Fortmann, *Tracking and data association*, Academic Press, 1988.

[Bogert *et al.*, 1963] B. Bogert, M. Healy, and J. Tukey, "The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo-autocovariance, Cross-Cepstrum, and Saphe Cracking," In M. Rosenblatt, editor, *Proc. Symp. Time Series*

Analysis, pages 209–243, New York, 1963. John Wiley and Sons.

[Brown, 1990] Christopher Brown, "Prediction and cooperation in gaze control," *Biological Cybernetics*, 63:61–70, May 1990.

[Brown and Coombs, 1991] Christopher Brown and David Coombs, "Notes on Control with Delay," Technical Report 387, University of Rochester, Computer Science Department, Rochester, New York 14627 USA, August 1991.

[Carpenter, 1988] Roger Carpenter, *Movements of the eyes*, Pion, 2nd ed., revised and enlarged edition, 1988.

[Clark and Ferrier, 1988] James Clark and Nicola Ferrier, "Modal Control of an Attentive Vision System," In *Proc. of ICCV'88, the Second International Conference on Computer Vision,*, Tampa, Florida, December 5–8, 1988.

[Coombs and Brown, 1991] David Coombs and Christopher Brown, "Cooperative Gaze Holding in Binocular Vision," *IEEE Control Systems*, June 1991.

[Coombs *et al.*, 1992] David Coombs, Ian Horswill, and Peter vonKaenel, "Disparity Filtering: Proximity Detection and Segmentation," In *Proc. of the SPIE Conf. on Intelligent Robots and Computer Vision XI: Algorithms, Techniques, and Active Vision,* Boston, Massachusetts, November 15–20, 1992.

[Corke and Paul, 1989] Peter Corke and Richard Paul, "Video-Rate Visual Servoing for Robots," Technical Report MS-CIS-89-18, Department of Computer and Information Science, University of Pennsylvania, GRASP Lab, Philadelpha, Pennsylvania 19104, February 1989.

[Dorf, 1980] Richard Dorf, *Modern control systems*, Addison-Wesley, 3rd edition, 1980.

[Howard, 1982] Ian Howard, *Human Visual Orientation*, Wiley & Sons, 1982.

[Jenkin, 1991] Michael Jenkin, "Using Stereomotion to Track Binocular Targets," In *Proc. of CVPR'91, the IEEE Conference on Computer Vision and Pattern Recognition,* Maui, Hawaii, June 3–6, 1991.

[Krauzlis and Lisberger, 1989] R. Krauzlis and S. Lisberger, "A Control Systems Model of Smooth Pursuit Eye Movements with Realistic Emergent Properties," *Neural Computation*, 1989.

[Land, 1975] Michael Land, "Similarities in the Visual Behavior of Arthropods and Men," In Michael Gazzaniga and Colin Blakemore, editors, *Handbook of Psychobiology*, pages 49–72. Academic Press, 1975.

[Lee and Wohn, 1988] Sang Wook Lee and K. Wohn, "Tracking Moving Objects by a Mobile Camera," Technical Report MS-CIS-8-97, University of Pennsylvania, Computer and Information Science Department, November 1988.

[Lisberger *et al.*, 1987] S. Lisberger, E. Morris, and L. Tychsen, "Visual Motion Processing and Sensory-Motor Integration for Smooth Pursuit Eye Movements," *Annual Review of Neuroscience*, 10:97–129, 1987.

[Matthies *et al.*, 1989] L. Matthies, T. Kanade, and R. Szeliski, "Kalman Filter-Based Algorithms for Estimating Depth From Image Sequences," *International Journal of Computer Vision*, 3:209–236, 1989.

[McDonald *et al.*, 1983] J. McDonald, A. Bahill, and M Friedman, "An Adaptive Control Model for Human Head and Eye Movements While Walking," *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(3):167–174, April 1983.

[Nishihara, 1984] H. Nishihara, "Practical real-time imaging stereo matcher," *Optical Engineering*, 23(5):536–545, September/October 1984.

[Olson, 1991] Thomas Olson, "Stereopsis for Fixating Systems," In *Proc. IEEE International Conference on Systems, Man and Cybernetics*, Charlottesville, Virginia, October 1991.

[Olson and Coombs, 1991] Thomas Olson and David Coombs, "Real-Time Vergence Control for Binocular Robots," *International Journal of Computer Vision*, 7(1):67–89, November 1991.

[Olson and Lockwood, 1992] Thomas Olson and Robert Lockwood, "Fixation-Based Filtering," In *Proc. of the SPIE Conf. on Intelligent Robots and Computer Vision XI: Algorithms, Techniques, and Active Vision,* Boston, Massachusetts, November 15–20, 1992.

[Pahlavan *et al.*, 1992] Kourosh Pahlavan, Tomas Uhlin, and Jan-Olof Eklundh, "Integrating Primary Oculumotor Processes," In *Proc. of ECCV'92, the Second European Conference on Computer Vision,* Santa Margherita Ligure, Italy, May 18-23, 1992.

[Papanikolopoulos *et al.*, 1991] N. Papanikolopoulos, P. Khosla, and T. Kanade, "Vision and Control Techniques for Robotic Visual Tracking," In *Proc. International Conference on Robotics and Automation*. IEEE, 1991.

[Reading, 1983] R. Reading, *Binocular Vision: Foundations and Applications*, Butterworth, Boston, 1983.

[Rojer and Schwartz, 1990] Alan Rojer and Eric Schwartz, "Design Considerations for a Space-Variant Visual Sensor with Complex-Logartihmic Geometry," In *Proc. of the International Conference on Pattern Recognition*, Philadelphia, Pennsylvania, June 1990.

[Shinkman *et al.*, 1985] P. G. Shinkman, M. R. Isley, and D. C. Rogers, "Development of Interocular Relationships in Visual Cortex," In R. N. Aslin, editor, *Advances in Neural and Behavioral Development*, pages 187–268. Ablex Publishing, Norwood, New Jersey, 1985.

[Thorpe, 1983] Charles Thorpe, "An Analysis of Interest Operators for FIDO," Technical Report CMU-RI-TR-83-19, Carnegie-Mellon University, December 1983.

[Tilley, 1990] David Tilley, "Zebra for MaxVideo: an application of object oriented microprogramming to register level devices," Technical Report 315, University of Rochester, Computer Science Department, 1990.

[Tölg, 1992] Sebastian Tölg, "Gaze Control for an Active Camera System by Modeling Human Pursuit Eye Movements," In *Proc. of the SPIE Conf. on Intelligent Robots and Computer Vision XI: Algorithms, Techniques, and Active Vision,* Boston, Massachusetts, November 15–20, pages 585–598, 1992.

[Tsotsos, 1988] John Tsotsos, "A 'Complexity Level' Analysis of Vision," *International Journal of Computer Vision*, 1(4), January 1988.

[Tyler and Scott, 1979] C. Tyler and A. Scott, "Binocular Vision," In R. Records, editor, *Physiology of the Human Eye and Visual System*, pages 643–74. Harper and Row, New York, 1979.

[Van der Spiegel *et al.*, 1989] J. Van der Spiegel, G. Kreider, C. Claeys, I. Debusschere, G. Sandini, P. Dario, F. Fantini, P. Bellutti, and G. Soncini, "A Foveated Retina-Like Sensor Using CCD Technology," In C. Mead and M. Ismail, editors, *Analog VLSI Implementation of Neural Systems*. Kluwer, 1989.

[Wavering *et al.*, 1993] A. Wavering, J. Fiala, K. Roberts, and R. Lumia, "TRICLOPS: A High Performance Trinocular Active Vision System," In *Proc. of the IEEE International Conference on Robotics and Automation,* Atlanta, GA, May 2–7, 1993.

[Waxman *et al.*, 1988] A. Waxman, W. Wong, R. Goldenberg, S. Bayle, and A. Baloch, "Robotic Eye-Head-Neck Motions and Visual Navigation Reflex Learning Using Adaptive Linear Neurons," *Neural Networks Supplement: Abstracts of 1st INNS Meeting*, 1:365, 1988.

[Whitehead and Ballard, 1990] Steven Whitehead and Dana Ballard, "Active Perception and Reinforcement Learning," *Neural Computation*, 2(4), 1990, (Also In the Proceedings of the Seventh International Conference on Machine Learning, Morgan Kaufmann, June 1990).

[Young, 1971] L. Young, "Pursuit Eye Tracking Movements," In P. Bach y Rita, C. Collins, and J. Hyde, editors, *Control of Eye Movements*. Academic Press, 1971.