Control of a Camera for Active Vision: Foveal Vision, Smooth Tracking and Saccade

EHUD RIVLIN
Department of Computer Science, Technion, Israel Institute of Technology, Haifa 32000, Israel
ehudr@cs.technion.ac.il

HÉCTOR ROTSTEIN
Rafael—Armament Development Authority and Department of Electrical Engineering, Technion, Israel Institute of Technology, Haifa 32000, Israel
hector@ee.technion.ac.il

Abstract. Several characteristics of the human oculomotor system have been suggested to be useful also for active vision mechanisms. Among others, foveal vision and a tracking scheme based on two different modes, called *smooth pursuit* and *saccade* have often been postulated or implemented. The purpose of this paper is to formulate a setup in which the benefit of implementing these schemes can be evaluated in a systematic manner, based on control considerations but incorporating image processing constraints. First, the advantage of using foveal vision is evaluated by computing the size of the foveal window which will allow tracking of the largest possible class of signals. By using linear optimal control theory, this problem can be formulated as a one-variable maximization.

Second, foveal vision leads naturally to smooth pursuit, defined as the performance that can be achieved by the controller resulting in the optimal size of the foveal window. This controller is relatively simple (i.e., linear, time-invariant) as is to be expected for this control loop.

Finally, when smooth pursuit fails a corrective action must be performed to re-center the target on the fovea. Recent results in linear optimal control, provide the necessary tools for addressing this challenging problem in a systematic manner.

Keywords: active vision, smooth pursuit, saccade

1. Introduction

"Active Vision" refers to the ability to move an image acquisition system in a controlled manner, in order to facilitate or allow certain machine vision tasks (Bajcsy, 1998; Swain and Stricker, 1993). Active vision systems usually consist of one or more cameras mounted in such a way that their orientation and imaging parameters (focus, zoom, aperture) can be adjusted in real-time. A simple active vision mechanism includes a single camera mounted on one or more mechanical degrees of freedom. More complicated stereo vision mechanisms or "robot heads" are constructed using two moving cameras mounted on a static or moving platform; over-

all, the resulting mechanism usually attempts to match the kinemathics and dynamics of the human oculomotor system.

The first active vision systems were constructed in the late eighties (see, e.g., Krotkov and Bajcsy, 1993; Ferrier and Clark, 1993); since then, sustained advance in the hardware for active vision has given rise to high performance systems, in some respects comparable with the human oculomotor system (Fiala et al., 1994). This progress has steamed the need for both highly efficient dedicated image processing tools and for control systems capable of exploiting the potential characteristics of the mechanisms.

1.1.   Human Visual System

Following the "natural" model provided by the human visual system, gaze control in robot heads is usually organized as a number of low level control loops; these loops should interact and—hopefully—cooperate to achieve the desired performance (Brown, 1990a). Two attributes of the human visual system are of interest in this paper: non uniform resolution and eye movement. The former refers to the fact that the human eye has a relatively small region with a high concentration of visual sensors in the center called "*fovea*," and a more or less exponential decay of the concentration of receptors towards the periphery. This seems to be the result of a tradeoff between high resolution vision and the necessity to operate—and survive—in real-time as dictated by neuro computing constraints.

The fact that the fovea is relatively small implies that it should be reoriented in order to span the desired region of visual acuity. Hence the need for eye movement which, interesting enough, is not evident from everyday experience. Eye movement has been the subject of intensive studies by researchers from different fields, and a large amount of sometimes conflicting data is available in the literature. The interested reader is referred to Robinson (1968) and Carpenter (1988) for an account of the activity from a control systems perspective.

Eye movement is a highly complex task from which two main mechanisms can be recognized. The first mechanism is called *saccade*, and consists of a rapid shift in the eye position. Saccade is characterized by a fast velocity and a relatively large time delay, presumably caused by processing time of the retinal information. For some time it was thought that vision is impaired during saccade due to the so-called "saccadic suppression" (Robinson, 1968). Further studies have shown that a certain amount of visual processing occurs during the course of a saccade (Carpenter, 1988), although they affect on-going computations in an indirect manner. Specific quantitative information about saccade in human oculomotor system may be found in Fiala et al. (1994), which also contains a pointer to the literature on the subject.

The second main mechanism of the oculomotor system is called *smooth pursuit* (SP). As suggested by the name, smooth pursuit is characterized by smooth and slow eye movements, involving a relatively short time delay. The main objective of SP is to keep the target within the fovea, compensating for relatively slow target movements. It is interesting to note that, although saccades seem to be driven by positional error, there is no consensus among researchers as to whether this error, velocity of the image or some combination, drives smooth pursuit.

1.2.   Modeling Assumptions and Previous Work

High performance active vision systems require control strategies adequate for dealing with the following control problems:

– Sampled measurements. State-of-the-art commercial cameras acquire images at a rate of about 30 frames/sec, which is to say that only a sampled version of the position of the target is available for tracking. Notice, though, that performance should be achieved in "continuous" time and not at the sampling instants only.
– Large time-delays. Extracting control-relevant information, e.g., the position of the target, out of the data delivered by a camera can be an expensive and time consuming computational task. A delay then appears between the instants the measurements are done and the information on the position is available. This time-delay is made larger once other tasks, like communication and control-law generation, are taken into account. As is well known, time-delays may severely limit the performance of a control system.
– Tight specifications. Active vision systems should achieve the performance expected from current mechanical heads and computer hardware. The control strategy should get the most out of the system, given the dynamic performance of the active vision mechanism and the hardware available for computation.
– Interaction/cooperation between the single-loop controllers. Following the premise that control loops should be designed one at a time, care should be taken to guarantee that the loops "cooperate" towards the common objective.

The first issue mentioned above deserves some more discussion, since it can be confused with the second one. Basic control systems are usually classified into two groups: continuous and discrete time. Most physical systems, and a robot head is an example, evolve in continuous time. As opposed to this, a digital computer deals with events happening at discrete time intervals. The connection between these two worlds is through sampler and hold (i.e., A/D and D/A) devices.

In an active vision systems, sampling is done through the image acquisition device, while the hold is implemented by the computer when sending piece-wise constant commands to the motors of the mechanism. It is clear that the computer/controller "sees" the system as a discrete-time one: information is obtained and control actions are taken at discrete times only. On the other hand, control signals to the motors are in continuous time (although constant over the sampling intervals) and so is the movement of the target that one wants to track.

When the sampling rate is fast (as compared with the system dynamics) it is justified to model the control system as either continuous or discrete-time. Indeed, the discrete-time controller behaves almost like a continuous-time one, while the behavior of the system over the inter-sampling can be ignored. When the sampling rate is low, this modeling becomes dubious: performance should be achieved in continuous time, but control action can only be taken at discrete intervals, due to measurement and computational constraints. Consequently, if a slow-sampled control system is designed using the discrete or continuous-time paradigms, extensive simulations or experiments should be carried out to validate the design.

Let us review next the previous work on active vision control, and check how they fit the above modeling premises. Ferrier and Clark (Ferrier and Clark, 1993; Clark and Ferrier, 1993) studied the control of the Harvard Binocular Head. Their control is based on the model of the oculomotor control described by Robinson, with separate subsystems for smooth pursuit and saccadic motion. The smooth pursuit loop uses PI control plus some delay in the loop (modeled as a continuous-time one) and is inspired by the Smith predictor. Saccadic movements are controlled by a sampled-data loop. An alternative approach was pursued by Pahlavan and Eklundh (1993) for the Royal Institute of Technology Head. In their control scheme, smooth pursuit and saccade are two independent loops, designed using linear prediction for the saccade. However, since both loops have approximately the same bandwidth it appears to be difficult to distinguish between them in practical operation.

A more detailed approach was proposed by Brown, Coombs and co-workers (Brown et al., 1993; Coombs and Brown, 1991; Coombs and Brown, 1993), working on the control of the Rochester Robot Head. These researchers introduce a Smith Predictor and a Kalman Filter in an attempt to compensate for time delays in the loop, following an earlier suggestions by Brown (1990a, 1990b). Notice that the delay is modeled as a continuous-time one. The general strategy is to use PID controllers coupled with predictors for smooth pursuit, while switching to an open-loop bang-bang controller for saccadic movements. Comparing with the previous discussion, these researchers attempt to circumvent the sampled-data nature of gaze-control by including predictors that should "fill-in" the intersampling behavior.

Other works have been reported by Milios et al. (1993), Christensen (1993) and Fiala et al. (1994) and are all based on considering pursuit and saccade as separate mechanisms, with PID's and possibly some predictors and delays on the feedback loops taking care of smooth-pursuit, and open-loop controllers based on linear predictions for the saccadic movement. The tuning of the controllers is done using classical control theory, followed by on-line adjustments based on the outcome of experiments. Switching between controllers is fired by a positional error larger than some threshold. Pursuit controllers are reset after each saccade. Some applications use Kalman filters in an attempt to predict the behavior by means of an observer (Christensen, 1995).

Murray et al. (1993) proposed a rather different scheme. First, they introduced non-uniform resolution by dividing the image into a coarse region and a foveal window on the center of the image. They also proposed an alternative scheme for the gaze control loop, based on a supervisor which should take care of deciding whether to pursue or saccade. In the additional work (Murray et al., 1995), switching from saccade to pursuit was also considered; in particular, the authors suggested that in order for the switching to succeed, both the position and the velocity of the target and the camera should be matched. This observation follows as a special case of the approach considered in the present paper.

It follows from this review that the sampled-data nature of active vision systems has somewhat been understated in the literature. The image acquisition process has been modeled as a pure time delay, which it is not: it should more accurately be modeled as a sampler followed by a discrete-time delay or *shift*. Often the overall system has been considered as discrete-time, which again is not: although the control action is taken at discrete intervals, specifications should be achieved in continuous-time. As opposed to these observations, in the present paper we will consider a model that takes the sampled-data nature of the system explicitly into account.

1.3.   Discussion and Main Problems

The discussion in the previous paragraph suggests that some active vision issues remain unsolved in spite of the activity in the area. In particular this paper attempts to clarify the following topics:

1. Smooth pursuit and saccade. In spite of its popularity and common sense appeal, to the best of our knowledge there is no proof available for the need of two separate mechanisms for tracking. A naive observer would argue for doing as "fast" tracking as possible all the time instead of slow pursuit movements possibly followed by a sudden burst of activity. The basic question is then to establish under what conditions the two mechanisms are desirable for improved performance.

2. Uniform vs. non-uniform sensing. Foveal vision has been mainly motivated by the needs of image processing in real-time. The question we would like to address is the following: Is there a need for foveal vision based on control considerations? And if so, what is the *optimal* size of the foveal window? To the best of our knowledge, no results in this area have been reported so far.

3. Interrelationship between pursuit and saccade. If two control modes are desirable (and we anticipate here that this may well be the case), then control laws must be computed for each mode, together with a mechanism for switching between them. Switching is clearly critical for a satisfactory operation. This is less so if the target of the saccade is stationary, which is one of the assumptions of many of the previous works on this subject, but becomes critical for a moving target (or platform).

The purpose of this paper is twofold. First, the modeling of an active vision system is reviewed under the lines outlined in the previous section. Second, answers are provided for the three problems mentioned above.

The paper is organized as follows. In the next section the model for active vision control is introduced, including some simplifying assumptions. In Section 3, the need for non-uniform resolution is formulated as an optimal control problem, which will eventually lead to the conclusion that the necessity of active vision depends both on the system dynamics and on the optimality criteria. A flexible framework is provided which allows the computation of the *optimal* size of the fovea whenever non-uniform resolution is found to be convenient. As a natural corollary to this problem, smooth

pursuit will be defined in rather precise terms, and the necessity of saccadic movements in order to achieve good overall performance will be established. The necessity of two control laws has been often postulated in the active vision literature and also used in actual implementation, but to the best of our knowledge the material in Section 3 is the first systematic presentation on the subject. This leads to Section 4, which deals with the problem of finding a suitable paradigm for saccadic control, and constructing a switching mechanism between the two control modes. Section 5 contains some simulation and experimental results. The good performance of the smooth pursuit control law designed following the theory in Section 3 is experimentally verified by comparing it with a controller designed using a classical control approach. The performance of saccadic control was evaluated through simulations. Finally, Section 6 presents some conclusions.

2.   Setup and Modeling Considerations

For the purpose of addressing the basic problems of foveal vision and tracking mechanism, it suffices to consider a configuration with a single camera with only one degree of freedom. For convenience, additional simplifying assumptions are introduced, which yield the simplest problem with all the essential features of active vision control built into it. Remarks as how to remove these assumptions will be outlined. Consider the setup illustrated in Fig. 1. Here the camera is mounted on a motor and has one degree of freedom, i.e., the angle θ that forms the optical axis with the horizontal. This angle can be modified by using the control signal u commanding the motor.
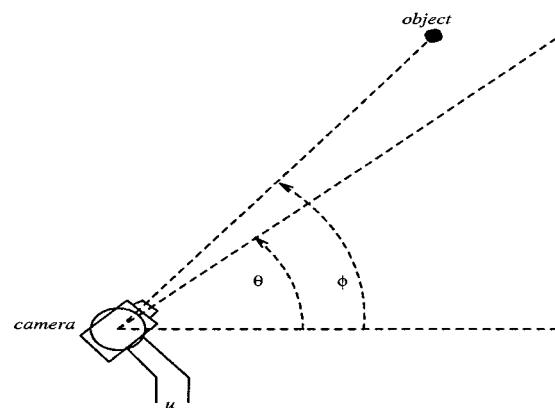

Figure 1.   A simplified setup.

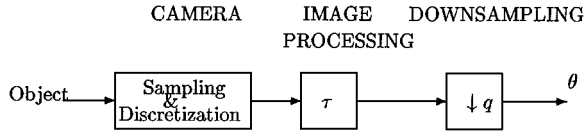CAMERA   IMAGE   DOWNSAMPLING
PROCESSING

Figure 2. Data-acquisition block diagram.

The image of the object is acquired by the camera connected to a vision card, which entails a sampling process, at a typical rate of at most 30 Hz, and also spatial discretization which will be neglected in what follows. Each image should be processed in order to extract information about the position of the object, e.g., the angle ϕ that forms the centroid of the object with the horizon, as measured from the axis of rotation. The time consumed by this processing will depend of the amount of data present, i.e., the "size" of the image, and on the sophistication of the image processing algorithm. In the model illustrated in Fig. 2, the image processing stage is lumped together with other effects like control law computation and communication delays, in a pure *time delay* τ proportional to the size of the image (plus some overhead). It is worth stressing that, as opposed to, e.g., Sharkey and Murray (1996), this time delay is of *discrete* nature.

If the delay τ is larger than the sampling period T, the sequence of images has to be down-sampled by a corresponding factor q as in Fig. 2, unless parallel processing units are employed. In the simplest possible case, q will be equal to the smallest integer larger than τ/T, but it can be made smaller subject to hardware availability. For ease of exposition, the former case is considered in the sequel. Assuming that the hardware and the image processing algorithms are given, the sub-sampling rate will only be a function of the size of the image x (see below for details); the notation q^x will be used to stress this fact.

The feedback block-diagram, including the motor and the load, is shown in Fig. 3. The block S_T is a sampler with sampling period T, which is followed by a down-sampler with down-sampling rate q; the
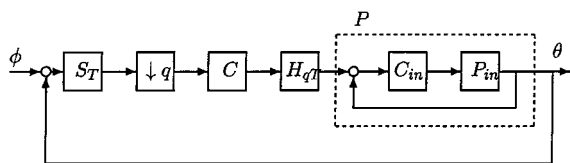
continuous-time signal is then sampled with sampling period qT. The block H_{qT} represents a hold function (typically a zero-order hold) which translates the discrete time output of the controller into a continuous-time signal. Several comments are in order:

– The system dynamics are all lumped into the plant P_{in}, and a feedback controller C_{in} is included in order to obtain good position regulation and desensitization of the electro-mechanical system from plant variations, possible neglected nonlinearities, platform motion and disturbances. Standard hardware may be used to implement this "inner" loop, which will usually work at much larger sampling rates. The transfer function of this closed-loop is called P, which is assumed to be known. Notice that if the active vision system is mounted on a static platform, then the above stabilization problem is rather straightforward. For a moving platform, however, this problem can become quite complicated; stabilization should then be achieved with respect to "inertial" space.

– While the actual image and the angle θ evolve in continuous time, the acquired image and the input to the controller are discrete time signals with different sampling rates whenever $q > 1$; the resulting closed-loop is then *multi-rate*. It is worth stressing that neither the continuous time error $e(t) = \phi(t) - \theta(t)$ nor the one resulting from the "fast" sampling $e(kT) = \phi(kT) - \theta(kT)$ can be measured (the latter if $q > 1$), and only the $\hat{\epsilon}(k) = \phi(kqT) - \theta(kqT)$ are available for control. It is the belief of the authors that this observation has been somewhat neglected in the previous literature.

– Finally, the figure is an idealized setup since a more realistic configuration should take into account noisy measurements, plant/model mismatch and possible disturbances. However, most of these problems can be dealt with by using classical control techniques when designing the stabilization controller C_{in} described above.

In this paper, continuous time signals will be denoted as, e.g., $w(t), \theta(t)$ and sometimes the dependence on t will be dropped when no confusion can arise. Discrete time signals will be denoted by, e.g., $\hat{\epsilon}(k)$. When corresponding to sampling of a continuous time signal, the equality $\hat{\phi}(k) = \phi(kT)$ holds, where the sampling period T should be clear from the context.

Figure 3. Closed-loop system.

3. Is Non-Uniform Resolution Convenient?

As opposed to the human visual system, most cameras available commercially have uniform resolution, raising the question of whether it is beneficial to implement a fovea in an active vision system. Although foveated vision has been implicitly or explicitly (e.g., in Murray et al., 1993) implemented before, the objective of this section is to justify non-uniform resolution in precise terms.

A foveal window should be easy to implement in hardware or a combination of hardware and software, by keeping high resolution on a specified region of the image and reducing the resolution, e.g., by filtering and down-sampling, on the rest. The existence of a region of high resolution reduces computational times, therefore leading to faster sampling-rates and smaller time-delays and suggest the potential of a better performances. At the same time, reducing the size of the window in which a target should remain makes the tracking specification tighter, and more so the smaller the region. This describes the basic tradeoff involved in deciding the potential benefits of implementing multi-resolution sensing. The purpose of this section it to formulate this tradeoff in a systematic manner, which will allow the computation of the size of the fovea in some optimal sense.

Consider the feedback configuration illustrated in Fig. 4. A reference model M has been included which generates the position $\phi(t)$ of the object as a function of the external signal $w(t)$. Inclusion of M does not necessarily imply an a priori knowledge of the behavior of the target[1] since, for instance, M could be a single or double integrator which corresponds to assuming that ϕ is generated by the velocity or acceleration of the target which should then be characterized in some useful sense. It is worth stressing that this does not imply that $w(t)$ is available for feedback: the control system is driven by the positional error alone, since this is the only quantity available for measurement. The signal $w(t)$ is introduced as an artifice for designing the controller C.

When $w(t)$ denotes the acceleration of the target, a feasible controller should drive $e(t)$ asymptotically to
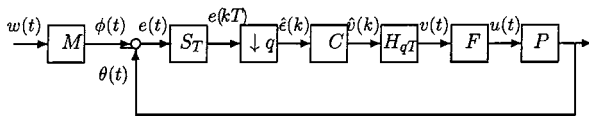
0 whenever $w(t) \equiv 0$, i.e., zero asymptotic error for constant velocity. This is a desirable characteristic also observed in the human visual system. As discussed in Yamamoto (1994), this cannot be achieved by using the discrete time controller C alone, but it is possible to connect between the output of the controller and the input of the plant a pure integrator or, in general, a filter $F(s)$ as shown in the figure. The same observation can be understood in a perhaps more intuitive manner once motion is allowed to the platform on which the active vision system is mounted. Then, the stabilization objective should be formulated with respect to inertial space, in which the velocity of the camera axis rather than its actual position is measured and controlled.

The signal w is assumed to be an integrable function belonging to a set $\mathcal{W}(\alpha)$ parameterized by a positive real number α. Examples are the sets

$$\mathcal{W}(\alpha)_\infty \doteq \{w \text{ s.t. } |w(t)| \leq \alpha \; \forall t \geq 0\} \qquad (1)$$

or

$$\mathcal{W}(\alpha)_2 \doteq \left\{w \text{ s.t. } \int_0^\infty |w(t)|^2 \leq \alpha^2\right\}. \qquad (2)$$

$\mathcal{W}(\alpha)_\infty$ and $\mathcal{W}(\alpha)_2$ correspond to signals which are uniformly bounded for each time t and signals with bounded energy respectively. The subscript in the notation reflects the fact that, in mathematical terms, 1 and 2 are the balls of radius α in the $\mathcal{L}_\infty$ and $\mathcal{L}_2$ norm respectively. In general, the set $\mathcal{W}(\alpha)$ should satisfy a monotone inclusion property as a function of α: $\mathcal{W}(\alpha_1) \subset \mathcal{W}(\alpha_2)$ if $\alpha_1 < \alpha_2$. This property is clearly satisfied by $\mathcal{W}(\alpha)_\infty$ and $\mathcal{W}(\alpha)_2$.

Together with the reference model M, the set $\mathcal{W}(\alpha)$ gives a degree of freedom available for design. In particular, the choice of $\mathcal{W}(\alpha)$ is dictated by the class of movements that the camera is expected to be able to track; $(\alpha)_\infty$ is a reasonable choice whenever little is known a priori about $w(t)$, and constitutes the main example considered in the sequel.

The other ingredient in the present approach is the *half* size of the fovea, denoted x and measured in the same units as θ and ϕ. If $e(t) = \theta(t) - \phi(t)$ denotes the difference between the position of the camera and the target at time t, the control objective is to design a discrete-time controller C such that

1. The closed-loop system is stable, and
2. $|e(t)| \leq x$ for each $t \geq 0$, whenever $w \in \mathcal{W}(\alpha)$.

Figure 4. The feedback configuration with reference model.

Notice that the specification in 2. is made in terms of the continuous time error $e(t)$ and not the sampled one $\hat{\epsilon}(k)$ which is available to the controller. The reason for this is that concentrating in $\hat{\epsilon}(k)$ may result in the target not remaining within the fovea during inter-sample time, which may be undesirable for image processing purposes; moreover, it may lead to oscillatory responses which should be avoided since the velocity of the object with respect to the camera should be relatively small to prevent image blurring. More generally, and subject to the particular application, one may want to consider a specification combining the continuous and sampled behavior as considered in Mirkin and Palmor (1997), since this could boost the achievable performance.

The existence of a controller that satisfies the above criterion will depend in general on α, since $e(t)$ cannot be guaranteed to be small for arbitrarily "large" signal. Alternatively, given a controller C, and a fovea half-size x, there exists a largest possible value of α, depending both on the controller and on the fovea size, that still satisfies the criterion. The dependence on the controller can be removed by finding the "best" possible one, which solves the optimization problem:

Problem 1 (Maximum Size of Input). Given x, find the largest α^x for which there exists a controller C^x that guarantees $|e(t)| \leq x$ for any $w(t) \in \mathcal{W}(\alpha^x)$.

A solution to this problem is presented in Appendix A, by formulating an equivalent ℓ_1 optimal control problem.

Note that the bound α^x will be small both for $x \approx 0$ and, typically, also for large values of x since image processing delays become dominant. The maximum of α^x will then be achieved for some finite value of x:

$$\alpha^* \doteq \max_x \alpha^x.$$

In principle, the maximum can be achieved for more than one value of x, so let x^* denote the largest such value less than the half size X of the camera. Then:

Non-uniform resolution is beneficial whenever $0 < x^* < X$, since then a controller may be designed such that w belongs to the largest possible set $\mathcal{W}(\alpha^*)$ such that $|e(t)| \leq x$.

The associated controller $C^s = C^{x^*}$, which for the cases considered above is linear and time-invariant, will be referred to as a smooth tracking controller. C^s guarantees that the target will remain inside the fovea for the worst case $w \in \mathcal{W}(\alpha^*)$, although w can potentially not be in $\mathcal{W}(\alpha^*)$ and still the objective $|e| < x^*$ be satisfied.

Remark. As explained in the appendix, the computation of C^s is made complicated by the fact that the system is sampled-data. In the $\mathcal{L}_1$ (or rather, induced $\mathcal{L}_\infty - \mathcal{L}_\infty$) optimal case, this problem was considered in Dullerud and Francis (1992) by approximating the continuous-time signals using fast sampling and then solving the resulting discrete-time problem. If the signal w is assumed to have bounded energy, then this problem was briefly considered in Bamieh et al. (1991).

The main conclusions of this section are that the optimal size of the fovea can be computed as the solution to a maximization problem, and that the benefit of implementing a foveal window depends on (compare with Murray et al., 1995) a) the limits imposed by the dynamics of the mechanical system, b) computational delays and other hardware constraints, and c) the characterization of the signal w (i.e., the definition of $\mathcal{W}(\alpha)$), which in turn reflect the set of movements of the target $\phi(t)$ one expects to track. Recent progress in robot head construction suggests that b) is the major factor now limiting the achievable performance.

4. Smooth-Pursuit and Saccade

The discussion in the previous section provides a complete answer to the first and a partial answer to the second questions raised in the introduction. Moreover, it establishes that a single linear time-invariant controller cannot generically guarantee that the target will remain within the fovea for arbitrary signals $\phi(t)$. Therefore, although smooth pursuit achieved by a single linear time-invariant controller may suffice in some cases, it may be inadequate by itself for many practical situations. For instance, almost by definition C^s cannot be used to perform fixation shifts.

The purpose of this section is to develop a control strategy for the case when the target moves out of the fovea or a fixation shift is specified by a higher level controller, which are characterized by $|e(t_v)| > x$ for some time t_v. A similar problem was discussed before in Murray et al. (1995); the main conclusions obtained in that work through simulations can be considered as a special case of the general problem considered below. It is important to stress that the objective of a saccade

is both to center the target on the fovea at some time $t_s > t_v$ and to guarantee that the smooth controller will be able to perform satisfactory for $t \geq t_s$ if the assumption on $w(t)$ is satisfied. Since performance can be poor in the interval $[t_v, t_s]$, a natural objective is to make this interval as short as possible.

It is instructive to think of the example of a television broadcast of a football game[2]. If the ball remains in a relatively small sector of the field moving at low velocities, the camera tracks its trajectory with smooth movements, but a ball kicked strongly will usually require rapid camera movement in an attempt to maintain it or bring it back to the field of view. These last movements, referred to as "saccades" in the sequel, are of a different nature than the ones required for smooth pursuit. First, they appear to be more *reflective* in the sense that they involve higher level of processing on the part of the operator. As another example which highlight this fact, it is interesting to note that the visual system of a newborn is able to perform some smooth pursuit, but a child is able to follow the flight of a ball only after he or she is several years old! Second, they involve larger control actions as compared to the ones generated by smooth pursuit. Third, the error $|e(t)|$ is reduced only at some time t_s in the future as opposed to the uniformity achieved by the smooth controller. In order to that, it is necessary to be able to predict $\phi(t_s)$, which in turn requires finding a suitable model for $\phi(t)$. As will become clear, this model is critical for the success of the saccadic correction. Fourth, and related with the previous one, the control system appears to become refractory to new input, which is consistent with our previous treatment and closed-loop stability. Going back to the example, if the ball hits against an obstacle and bounces back after the saccade has been triggered, then this is not immediately taken into account but rather a saccade is completed to the *wrong* place before a second saccade is executed to correct for the error. These examples capture some of the main characteristics of the saccadic movements of the human oculomotor system (Robinson, 1968).

4.1.   *Switching Between Controllers*

As described above, a saccade is a fast movement triggered when the smooth controller cannot keep the target within the fovea, with the objective of bringing the target back to the smooth-pursuit regime. It is worth stressing that the saccadic control should not only achieve fast target re-engagement, but it should also guarantee

that the target could be tracked by the smooth pursuit controller right after the saccade. The driving idea behind our approach to saccadic control is to generate the control law in such a way that from the instant the saccade has been completed, the smooth pursuit control loop can be closed without introducing transients. This is achieved by making the internal states of the physical system right after the saccade identical to the ones that could have been reached by the smooth-pursuit closed-loop.

A systematic derivation of the control law is rather technical, especially because of the difficulties introduced by the sampled-data nature of the controller. A derivation of the equations governing the saccade, as well as surrounding theoretical results, are given in Appendix B.

4.2.   *Saccadic Control*

The discussion in the previous section provides the framework for the systematic treatment of saccadic control. Following the approach in Rivlin et al. (1997), four different stages are considered.

Switch On.   Suppose that the constraint on $|e(t)|$ is violated at time t_v so that the smooth controller can no longer guarantee good performance or even continue its normal operation. A saccadic action is then triggered, which requires relatively lengthy computations. Meanwhile, the camera should somehow be operated in a way that will possibly facilitate the future correction. In the absence of additional information about the variations of the position of the target, then one could select a fictitious signal w^{t_v} in such a way that the error criterion remains constant from t_v and until the saccadic control is employed. Taking then $e_v \doteq e(t_v)$,

$$x_C(j) = A_C^{j-k_v} x_C(k_v) + \sum_{i=k_v}^{j-1} A_C^{j-i} b_C e_v$$

where $x_C(k_v)$[3] is the internal state of the smooth pursuit controller C_1 at sampling instant k_v and $j > k_v$. The output signal of C_1 is

$$
\begin{aligned}
u_C^1(j) &= c_C x_C(j) + d_C e_v v \\
&= c_C A_C^{j-k_v} x_C(k_v) + \big[c_C A_C \big(A_C^{j-k_v} - I \big) \\
&\quad \times (A_C - I)^{-1} b_C + d_C \big] e_v
\end{aligned}
$$

and

$$v(j) = v(j-1) + u_C^1(j-1). \qquad (3)$$

Take now $jh < t \le (j+1)h$. The control input to the plant is

$$u(t) = u(jh) + v(j)(t - jh)$$

due to the continuous time integrator, giving

$$\theta(t) = ce^{A(t-t_v)}x_P(t_v) + c\int_{t_v}^t e^{A(t_v-s)}bu(s)\,ds$$
$$= ce^{A(t-t_v)}x_P(t_v) + c\int_{t_v}^t e^{A(t_v-s)}b(u(\lfloor t/h\rfloor h)$$
$$+ v(\lfloor t/h\rfloor)(t-jh))$$

and

$$\phi(t) = e_v + ce^{A(t-t_v)}x_P(t_v) + c\int_{t_v}^t e^{A(t-s)}b(u(\lfloor s/h\rfloor h)$$
$$+ v(\lfloor s/h\rfloor)(s-jh))\,ds.$$

Taking derivatives:

$$\dot{\phi}(t) = cAe^{A(t-t_v)}x_P(t_v) + cb(u(jh) + v(j)(t-jh))$$
$$w^v(t) = \ddot{\phi}(t) = cA^2e^{A(t-t_v)}x_P(t_v) + cbv(j).$$

Replacing $v(j)$ with (3) gives an expression for the virtual reference signal, which can be computed on-line and hence used for driving the camera while performing the two computational stages discussed next.

Modeling. In order to reduce the error signal below the fovea half-size at some future instant τ, it is necessary to predict the values of the signal $\phi(t)$ for $t \ge \tau$, based on measurements which are usually costly to obtain and potentially contaminated by noise. The success of the saccadic control action may depend on the accuracy of these predictions. As an example, suppose that the target is located at some stationary point lying outside the foveal window (this is a standard experiment when evaluating human saccades (Robinson, 1968)); then the modeling problem reduces to determining the new position, which can presumably be done accurately. On the other hand, suppose that in the football example discussed in the beginning of this section, the ball bounces back after being kicked; the prediction

will then most probably be poor and in the real life example requires additional saccadic corrections before returning to smooth pursuit regime.

The reference Bar-Shalom and Fortmann (1988) contains an array of different algorithm for the computation of models for predicting signals under various sets of assumptions. The algorithm of choice should be selected depending on the standing assumptions for $\phi(t)$ and the noise level corrupting the measurements. This selection is important since it sets the time lag required to have a prediction of future position and determines the a priori accuracy of the predictions. A popular choice in the active vision field is to select $\alpha - \beta$ or $\alpha - \beta - \gamma$ filters for prediction. This filters have the advantage of their simplicity, and the coefficients of these filters are usually selected by using the steady-state solution to a corresponding Kalman filtering problem (Bar-Shalom and Fortmann, 1988). However, much better predictions can be made if a priori knowledge of the variations of $\phi(t)$ are available and exploited.

Saccade. Once the model is available at time, say, t_p, it is possible to compute $\hat{x}_M(t_s)$ for some future time instant and hence the time-varying target set $\mathcal{O}(\hat{x}_M(k_s))$. The problem is now to generate the control signal $u^{sac}(t)$ that drives the plant from $\hat{x}_P(t_p)$ to $\mathcal{O}(\hat{x}_M(t_s))$. A natural objective is to do this in the shortest possible time, not only because of the tracking objective but also since the future prediction of $\hat{x}_M(t_s)$ potentially deteriorates with time. It is implicitly assumed that the internal state of the plant is measurable for feedback; this can be achieved at least approximately if the internal control loop discussed before is designed so that P can be accurately approximated by a second order system, for which both position and velocity are measured.

The computation of the saccadic control appears to be challenging; it can be approximated by using fast-sampling, i.e., replacing the continuous-time virtual input $w(t)$ by a piece-wise constant function:

$$w(t) = \acute{w}(k) \quad k\hat{h} \le t < (k+1)\hat{h}$$

where $\hat{h} \ll T$. This reduces the problem to a discrete-time multi rate one. The advantage is that in that case linear programming based algorithms exist (de Vlieger et al., 1982) for solving these problems, and they allow the inclusion of additional constraints, like bounds on the tolerable control actions. Notice that the constraint

on the target set is a linear one, and so can be incorporated with minor modification into the formulation.

Switch Off. Linear optimal controllers such as an $\mathcal{L}_1$-optimal, assume that the initial state for the plant to be controlled is zero. If the initial state is non-zero and unknown, then the controller can no longer guarantee the desired performance and should be replaced by a usually more complicated one (e.g., non linear, time-varying). As claimed above, if the initial state is known, then the same controller can be used if it is properly initialized, since it amounts to finding a fictitious but legal disturbance that would drive the state of the plant to the actual non-zero initial state, when the plant is interconnected with the optimal controller. Then, it is possible to "read-out" the state of the controller and initialize the actual configuration so that the optimal performance can be guaranteed.

5.   Simulation and Experimental Results

The control strategy presented above was tested on the Technion Robot Head. The head consists of two cameras mounted on a mechanism providing four mechanical degrees of freedom (see Fig. 5). For the purposes of the present research, the pan axis of one camera was used for control, while a laser pointer was attached to the other camera. This laser pointer provides a target for the control system which can be controlled by the computer.

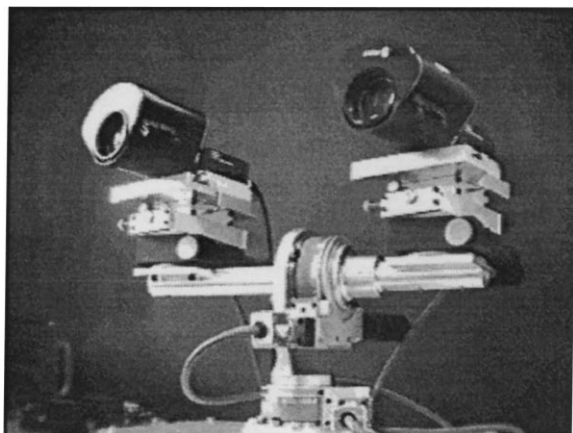Through experiments, the following open-loop transfer function was obtained for the pan axis of interest, relating normalized voltage input to the motor with angular displacement of the axis:

$$P_{in}(s) = \frac{-3.3012s + 3.9197 \times 10^4}{s^2 + 64.7983s}$$

The parameters of the controller available with the controller board of the Technion Head was subsequently designed using classical control considerations. It was found that proportional control suffices to provide very good open loop performance, in terms of stability, bandwidth and noise rejection properties. After taking all the gains in the loop into account, the controller was seen to be

$$C_{in} = -.5.$$

Next, the theory presented above for the design of a smooth pursuit controller was applied. In order to do this, an integrator was added in cascade with the internal closed loop of P_{in} and C_{in}:

$$\frac{1}{s}\frac{P_{in}C_{in}}{1 + P_{in}C_{in}} = .5\frac{-3.3012s + 3.9197 \times 10^4}{s^3 + 63.1477s^2 + 1.96 \times 10^4 s}.$$

The optimal performance of the system when considering discrete-time PID controllers with sampling time $h = q(1/30)$ sec is illustrated in Fig. 6. As expected from the interconnection, the performance degrades monotonically with decreasing sampling rate. These PID controllers were optimally tuned following the theory in Section 3. The reason for this is that actual optimal controllers could be of high complexity (indeed, optimality is achieved only for infinite dimensional controllers); also, PID controllers are easier to implement and require minimum software changes.
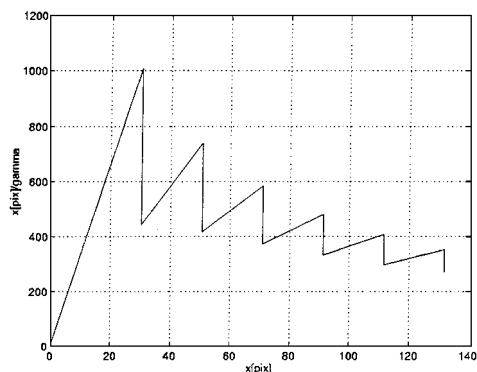

Figure 5.   The Technion Robot Head.

Figure 6.   The optimal performance of the system when considering discrete-time PID controllers. The vertical axis x(pix)/gamma measures the worst-case error (see Appendix A).

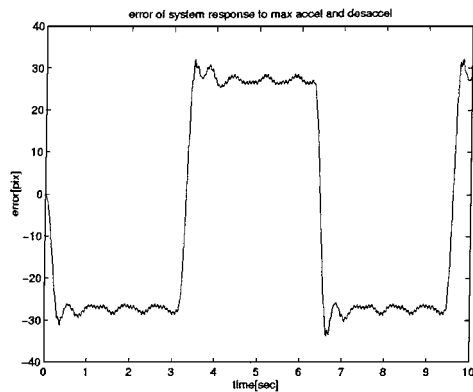*Figure 7.* The performance of the smooth pursuit controller.

Figure 8. The performance of the smooth pursuit controller (the solid line) vs. a standard PID controller (dashed line).

Figure 9. The performance of the smooth pursuit controller (the solid line) vs. a standard PID controller (dashed line).

A model for the image processing block is now required to trade-off against the performance degradation induced by the sampling-rate. Due to the character of the target, a simple correlation algorithm was employed to localize the target from each essentially one-dimensional image. As a consequence, the computational time can be model through the equation

$$\tau = a_{\mathrm{o}} + a_1 x$$

where x is half the size of the fovea, normalized between zero and one. Comparing this cost with the performance in Fig. 6, it was found that the optimal trade-off is achieved at $q = 2$, meaning that the optimal fovea has a half-size of 30 pixels, and that only one out of two images should be used as input to the control system. The resulting controller is capable of tracking within the fovea, target accelerations of $250°/\mathrm{sec}^2$. The performance of the smooth pursuit controller is illustrated in Fig. 7. There, the target is initially at rest at the center of the fovea, then it accelerates with the maximal admissible acceleration in one direction, and finally decelerates and accelerates in the opposite direction. The figure shows that at overall the trajectory, the target remains inside the fovea.

Figure 8 presents the performance of the smooth pursuit controller (the solid line) vs. a standard PID controller (dashed line) to an $\phi(t) = t - t * \cos(2\pi/7)$ as performed running on the Technion Robot Head. The size of the fovea is 4 degrees (about 60 pixels). One can see that the standard PID controller takes the target outside of the fovea at about the tenth second. We also checked the difference in reaction to acceleration of the two controllers. The results ar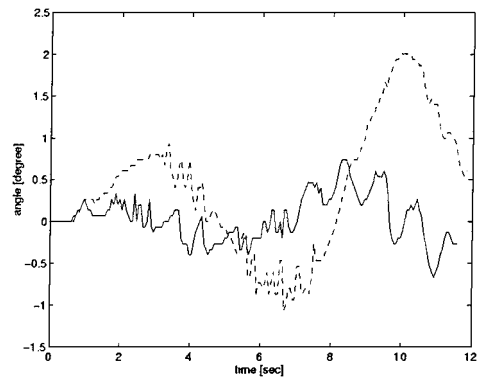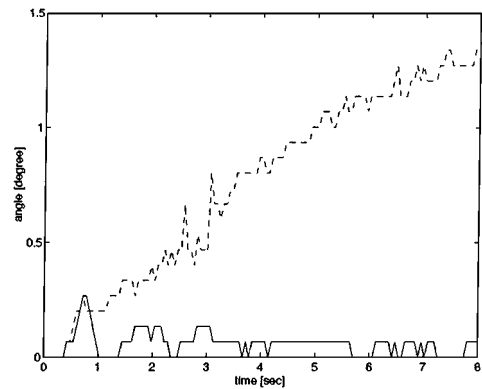e presented in Fig. 9. One can see that the standard PID controller converges to a steady-state after about 8 seconds, with a relatively large error (1.5 degree, about 22 pixels).

Saccadic Control

The implementation and evaluation of saccadic control is much more complex and, consequently, the performance will be illustrated by means of a simulation example which captures the behavior of the two loops in a quite nice manner. Figure 10 presents the performance of the system to sinusoidal input when running in two mode control. The input signal is such that smooth pursuit is broken twice during each period. However, saccadic control is able to bring the target again within the fovea and switch back to smooth pursuit. It is interesting to observe that since the target is not static, most of the previous approaches to saccadic control would probably fail to switch successfully back to smooth
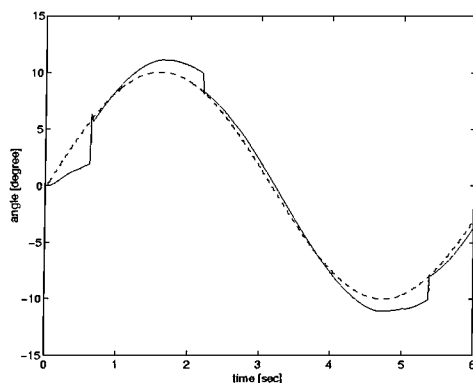
Figure 10. The performance of the system in a two-mode control.

pursuit. A possible exception to this is the control scheme proposed in Murray et al. (1995).

6. Conclusions

In this paper some of the fundamental problems regarding the control of an active vision system have been addressed. It has been shown that the benefit of implementing foveal vision can be formulated as an optimization problem, since a trade-off appears between having a small window which would yield small computational delays but tighter control objectives or relaxing the control objectives but obtaining more challenging dynamics. Following the current approach, the size of the fovea is chosen as the one giving best tracking capabilities, as measured by the size of the signals which the system is guaranteed to track. It was also shown that foveal vision is tightly related with smooth pursuit, since the solution to the former provides a controller which makes the latter meaningful.

If the performance provided by the smooth controller does not suffice, which will be the case if the camera is expected to perform in a realistic environment, then one is necessarily led to consider a two-mode controller, in which the smooth controller is replaced whenever the error fails to meet specifications by a saccadic controller. This latter controller is substantially more sophisticated since it has to have capabilities of modeling the evolution of the target, generating a signal which will drive the system momentarily and then another one that should position the camera in such a way that the smooth pursuit controller can be switched-back into the loop and perform according to specifications. It was established that, in the light of recent developments in

optimal control theory, all this requirements can be formulated in a systematic manner.

The setup considered in this paper is clearly oversimplified, and several adjustments might be required in order to implement the different stages. Computation of the optimal fovea size and smooth controller appear to be straightforward, although several modifications should be introduced to make the design more realistic; for instance, model uncertainty and noise corrupting the measurements should be introduced into the picture. Although this makes the calculations slightly more involved, since for instance the calculation of α^x requires itself some iterations, they can still be performed off-line in a reasonable time. As for saccadic control, it is clear that since it involves intensive on-line computations it should be implemented carefully to obtain satisfactory results, and a priori knowledge on the type of target the system is expected to track should be used to speed up computations. In this respect, notice that the modeling stage has been de-emphasized although it is critical for achieving good performance, and hence constitutes the degree of freedom that the designer has at hand for tailoring saccades to specific applications.

Although the present paper differs from most of the works on the area of active vision by stressing the control aspects of the problem, it is the feeling of the authors that it is also tightly connected with them since issues that have been discussed before can be easily accommodated into the current approach, including image processing techniques or modeling of the position of the target.

As stated in the introduction, the most simple setup was considered in order to highlight the most important problems involved in tracking by an active vision system. The next steps are implementing these ideas in an actual system and increasing the complexity by including one more camera and other degrees of freedom. The latter appears to require rewriting the theory in this paper in terms of *multivariable* systems (e.g., systems with more than one input and/or output) while the former seems to be more challenging since specifications are harder to write down in a systematic manner. It will be interesting to find out in what sense the organization of the human oculomotor system provides clues as to how this should be done. As an incidental remark, several apparently desirable characteristics have been hypothesized about the behavior of this system, including the $\mathcal{L}_1$ optimal characteristic. Experiments are planned to contrast these hypotheses with real data.

Appendix

A.   A Solution to Problem 1

In mathematical terms, Problem 1 may be written as:

$$\alpha^x = \sup\{\alpha \; : \; \inf_{C \in \mathcal{C}} \sup_{w \in \mathcal{W}(\alpha)} |e(t)| \le x, \; t \ge 0\} \quad \text{(A1)}$$

where $C \in \mathcal{C}$ is used to denote that the controller is stabilizing. This problem is closely related to optimal control problems with an induced-norm criterion. To see this, let $T_{ew}(C)$ denote the transfer function between w and e for a given controller C. For C stabilizing, $T_{ew}(C)$ is stable and it is possible to define the system norm:

$$\begin{aligned}\|T_{ew}(C)\|_{\infty,i} &\doteq \sup_{w \in \mathcal{L}_i} \frac{\|e(t)\|_\infty}{\|w(t)\|_i} \\ &= \sup_{w \in \mathcal{W}(\alpha)} \frac{\|e(t)\|_\infty}{\alpha}\end{aligned}$$

where the last equality follows from linearity and the definition of $\mathcal{W}(\alpha)$. The relevance of this norm is that, given an input w, it is possible to bound the norm of the output as:

$$\|e(t)\|_\infty \le \|T_{ew}(C)\|_{\infty,i} \|w\|_i$$

and the bound is tight in the sense that there always exists an input w such that it holds as an equality. The controller C can be chosen optimally as a solution to the problem

$$\begin{aligned}\gamma_i^x &= \inf_{C \in \mathcal{C}} \|T_{ew}(C)\|_{\infty,i} \\ &= \inf_{C \in \mathcal{C}} \sup_{w \in \mathcal{L}_i} \frac{\|e\|_\infty}{\|w\|_i}.\end{aligned}$$

A solution C^x to this problem is *min-max* optimal in the sense that it guarantees that the norm of the output will remain smaller than $\gamma_i^x \|w\|_i$ for a given input $w \in \mathcal{W}(\alpha)_i$. It follows that $\|e(t)\|_\infty \le x$ if $\alpha = x/\gamma_i^x$ and for $\alpha > x/\gamma_i^x$ there always exists $w \in \mathcal{W}(\alpha)_i$ such that the constraint on the norm of $e(t)$ is violated. From a computational point of view, C^x as above may be found by using control theory techniques: in the case $i = 2$ C^x is known as generalized $\mathcal{H}_2$ controller while for $i = \infty$ it is called an $\mathcal{L}_1$ controller. The performance γ^x depends on x via the sub-sampling rate q^x, and hence will be piece-wise constant: only variations of x large enough to change the integer q will affect

it. Given that x is positive and bounded above by the physical size of the camera, the computation of only a finite number of values for γ^x is required. Notice that γ^x will typically be an increasing function of x since continuous time performance deteriorates as the sampling period becomes larger.

B.   Saccade Computation

Let the plant P be described by the state space equations:

$$\begin{aligned}\dot{x}_P(t) &= A x_P(t) + b u(t) \\ \theta(t) &= c x_P(t)\end{aligned}$$

which can be denoted in compact notation

$$P = \left(\begin{array}{c|c} A & b \\ \hline c & 0 \end{array}\right).$$

In the sequel, this compact notation will be used to denote transfer matrices whenever convenient. The discrete-time controller has a state-space representation

$$\begin{aligned}\hat{x}_C(k+1) &= A_C \hat{x}_C(k) + b_C \hat{\epsilon}(k) \\ \hat{v}(k) &= c_C \hat{x}_C(k)\end{aligned}$$

and the same compact notation may be used; the correct interpretation can be obtained from the context. Since a sampling device is linear but not time-invariant, the closed-loop of a sampled-data system is *time-varying*. Fortunately, it is periodically time-varying and hence *lifting* techniques may be used in order to write down the equations in a compact and elegant form; the reader is referred to Bamieh and Pearson (1992) for a comprehensive introduction to lifting for sampled-data systems. In a nutshell, continuous-time signals can be *lifted* into discrete-time ones by "cutting" them into pieces, and difference-equations can be found for the correspondingly lifted systems (assuming they are finite-dimensional). As a consequence, the closed-loop behavior is defined by a set of discrete-time state-space equations, albeit with operators replacing the state-space realization matrices.

In order to make our treatment specific, the model M will be assumed to be a *double-integrator*:

$$M = \left(\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \hline 1 & 0 & 0 \end{array}\right).$$

In order to have a) bounded $\mathcal{L}_1$ norm between w and e and b) zero steady-state error whenever w is constant, we take F to be the integrator:

$$F = \left(\begin{array}{c|c} 0 & 1 \\ \hline 1 & 0 \end{array} \right)$$

and set $C(z) = \frac{1}{z-1} C_1(z)$, which also makes the discrete-time controller strictly causal, a condition that every feasible controller should satisfy since $\hat{\epsilon}(k)$ cannot affect the current control signal. It follows from the internal model principle for sample-data systems (Yamamoto, 1994) that these two conditions are necessary and sufficient for satisfying a) and b)). It is implicitly assumed here that the plant contains no pole at zero, since otherwise either the continuous or discrete time integrator (or both) may not be required. S_T is an ideal sampler while the interconnection of S_T and the down-sampler is an ideal sampler with sampling rate $h = qT$. The hold H_h is assumed to be a zero-order hold while the controller C_1 has a state-space realization

$$C_1 = \left(\begin{array}{c|c} A_C & b_C \\ \hline c_C & d_C \end{array} \right)$$

and then

$$C = \left(\begin{array}{c|c} 1 & h \\ \hline 1 & 0 \end{array} \right) \left(\begin{array}{c|c} A_C & b_C \\ \hline c_C & d_C \end{array} \right) = \left(\begin{array}{cc|c} 1 & hc_C & hd_C \\ 0 & A_C & b_C \\ \hline 1 & 0 & 0 \end{array} \right)$$

Using lifting techniques, it is possible to get a representation for the closed-loop when $kh \leq t < (k+1)h$ in terms of state-space equations of the form:

$$x_S(k+1) = A_S x_S(k) + \int_0^h B_S(h-s) w(kh+s) \, ds$$

$$e(t) = C_S(t) x_S(k) + \int_0^h D_S(h-s) w(kh+s) \, ds$$

where

$$A_S = \begin{bmatrix} 1 & h & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & c\int_0^h e^{Ar} dr & -\psi_a(h)b & 0 \\ 0 & 0 & 0 & e^{Ah} & -\psi_b(h)b & 0 \\ hd_C & 0 & hd_C & 0 & 1 & hc_C \\ b_C & 0 & b_C & 0 & 0 & A_C \end{bmatrix}$$

$$C_S(t) = [\, 1 \quad t \quad 1 \quad c\int_0^t e^{Ar} dr \; - \psi_a(t)bc_C \quad 0 \,]$$

$$B_S(t) = \begin{bmatrix} t \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$D_S(t) = t$$

and

$$\begin{bmatrix} \psi_a(t) \\ \psi_b(t) \end{bmatrix} = \begin{bmatrix} c\int_0^t \int_0^{t-s} e^{Ar} \, dr \, ds \\ \int_0^t e^{A(t-s)} \, ds \end{bmatrix}.$$

The corresponding state vector is formed by stacking the ones of the reference model, plant and controller:

$$\hat{x}_S = \begin{bmatrix} \hat{x}_M \\ \hat{x}_P \\ \hat{x}_C \end{bmatrix}.$$

It follows by the internal model principle that the *unstable* states of M should be unobservable. To exhibit this fact, note that A is invertible (since otherwise the plant P would have some poles at zero), assume that $c \cdot b \neq 0$ and consider the change of variables:

$$\begin{bmatrix} \hat{x}_M \\ \hat{x}_a \\ \hat{x}_b \end{bmatrix} = \begin{bmatrix} \hat{x}_M \\ \hat{x}_P + \begin{bmatrix} 1 & 0 \\ 0 & A^{-1}b/cb \end{bmatrix} \hat{x}_M \\ \hat{x}_C + \begin{bmatrix} 0 & -1/cb \\ 0 & 0 \end{bmatrix} \hat{x}_M \end{bmatrix}.$$

Then, it is straightforward to verify that the closed loop system may be written as $S = S_{\text{uo}} + S_{\text{o}}$, where

$$S_{\text{uo}} = \left(\begin{array}{cc|c} 0 & h & t \\ 0 & 1 & 1 \\ \hline 0 & 0 & 0 \end{array} \right)$$

$$S_{\text{o}} = \left(\begin{array}{cccc|c} 1 & c\int_0^h e^{Ar} dr & -\psi_a(h)b & 0 & t \\ 0 & e^{Ah} & -\psi_b(h)b & 0 & A^{-1}b/cb \\ hd_C & 0 & 1 & hc_C & -1/cb \\ b_C & 0 & 0 & A_C & 0 \\ \hline 1 & c\int_0^t e^{Ar} dr & -\psi_a(t)bc_C & 0 & t \end{array} \right);$$

here S_{uo} is unobservable from the error signal. These systems generate the error signal $e(t)$ as a function of $w(t)$.

Given $\phi(t_s)$ at some future time $t_s = qk_s$, the objective of the saccadic control is to synthesize a control action that would allow to switch the smooth controller back into the loop at time t_s. A moment of reflection shows that it is not enough to guarantee that $|e(t)|$ will remain smaller than x for $t > t_s$ since, for instance, there may be a large velocity mismatch between the camera and the target at t_s. The problem is then to find an appropriate "target set" for the saccade; a similar problem was addressed in Rivlin et al. (1997) for the discrete-time case.

Before proceeding, some notation is required. Given a periodically time-varying system G in lifted form, an initial state $\hat{x}_0$ at time t_0 and some (integrable) function $w(t)$, let $\mathcal{F}_G(k, \hat{x}_0, t_0, w)$ denote the linear function mapping x_0 into the state trajectory $\hat{x}_G(k)$:

$$\hat{x}_G(k) = \mathcal{F}_G(k, \hat{x}_0, t_0, w).$$

Consider the *Reachable Set* $\mathcal{R}_G$ of G, defined as the set of all states that can be reached from 0 in a finite number of samples by using inputs $w \in \mathcal{W}(\alpha^*)$:

$$\mathcal{R}_G \doteq \{\hat{x}_G \ : \ \exists k_f, \ w \in \mathcal{W}(\alpha^*)$$
$$\text{giving } \hat{x}_G = \mathcal{F}_G(k_f, 0, 0, w)\}.$$

Let $\mathcal{R}_{S_o}^P$ denote the projection of the reachable set $\mathcal{R}_S^P$ into the states of the plant:

$$\mathcal{R}_{S_o}^P = \begin{bmatrix} I & 0 \end{bmatrix} \mathcal{R}_{S_o}.$$

It was shown in Blanchini and Sznaier (1995) that if the system is discrete or continuous-time and $\mathcal{W}(\alpha^*) = \mathcal{W}(\alpha^*)_\infty$, then if $x_P \in \mathcal{R}_S^P$ for some time instant, it is possible to construct a non-linear, static state-feedback controller that will force the future states to remain within $\mathcal{R}_S^P$. This fact was shown also to be true for the sampled-data case in Mirkin et al. (1998)[4] were, perhaps more interesting, it was also shown that it is possible to select a state $x_C(t_s)$ for the smooth controller C^s in such a way that the reduced closed-loop state $\begin{bmatrix} \hat{x}_a \\ \hat{x}_b \end{bmatrix}$ will remain inside $\mathcal{R}_{S_o}$ for future time samples if $w \in \mathcal{W}(\alpha^*)$ and hence $|e(t)|$ will remain bounded by x. This motivates the following definition of target set for saccadic control.

Definition 1 (Target Set).   Given an internal state of the reference model $\hat{x}_M^o$, the state $\hat{x}_P$ belongs to the *target set* $\mathcal{O}(\hat{x}_M^o)$ if there exists k_s and $w \in \mathcal{W}(\alpha^*)$ such that

$$\begin{bmatrix} I & 0 & 0 \end{bmatrix} \mathcal{F}_S(k_s, 0, 0, w) = \hat{x}_M^\circ$$
$$\begin{bmatrix} 0 & I & 0 \end{bmatrix} \mathcal{F}_S(k_s, 0, 0, w) = \hat{x}_P.$$

The set $\mathcal{O}(\tau, x_M^\circ)$ contains the states of the plant which can be reached by signals $w \in \mathcal{W}(\alpha^*)$ in a finite number of sample intervals if the internal state of the reference model is constrained to be equal to the one at τ, $\hat{x}_M(\tau)$. The important observation is that if u^{sac} is now computed so that $\hat{x}_P(\tau) \in \mathcal{O}(\hat{x}_C(\tau))$, then the smooth controller can be switched back into the loop at time $q\tau$ by initializing its internal mode to $\hat{x}_C(\tau) = [0\,0\,I]\mathcal{F}_S(k_s, 0, 0, w^v)$ where $w^v \in \mathcal{W}(\alpha)$ is such that

$$\begin{bmatrix} \hat{x}_M(\tau) \\ \hat{x}_P(\tau) \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \end{bmatrix} \mathcal{F}_S(k_s, 0, 0, w^v).$$

It follows from the reasoning in Rivlin et al. (1997) (see also Mirkin et al., 1998, for a more detailed treatment) that $|e(t)| < x$ for $t \geq \tau$ if the future disturbances $|w(t)| \leq \alpha^*$. The reason is that the closed-loop system will behave for $t > \tau$ as if the past input to the system would have been w^v (a similar interpretation can be made for the case of normed bounded signals).

Acknowledgments

Notes

1. Except for some smoothness assumption that any physical model should satisfy.
2. Here and in what follows, by "football" we refer to the *real* football as opposed to American football. No previous knowledge on the game is assumed on the part of the reader.
3. State vectors are denoted below with x and a sub-index. Not to be confused with the fovea half-size x without indeces.
4. As a matter of fact, the results in Mirkin et al. (1998) evolve from Rivlin et al. (1997) which in turn is inspired by the present work.

References

Bajcsy, R. 1988. Active perception, *Proceedings of the IEEE*, 76(8). Special issue on Computer Vision.

Bamieh, B., Pearson, J.B., Francis, B., and Tannenbaum, A. 1991. A lifting technique for linear periodic systems with applications to sampled-data control. *Systems & Control Letters*, 17:79–88.

Bamieh, B. and Pearson, J.B. 1992. A general framework for linear periodic systems with applications to $\mathcal{H}^\infty$ sampled-data control. *IEEE Transactions on Automatic Control*, 37(4):418–435.

Bar-Shalom, Y. and Fortmann, T.E. 1988. *Tracking and Data Association*, Mathematics in Science and Engineering. Academic Press.

Blanchini, F. and Sznaier, M. 1995. Persistent disturbance rejection via static state feedback. *IEEE Transactions on Automatic Control*, 40(6):1127–1131.

Brown, C. 1990a. Gaze controls with interactions and delays. *IEEE Transactions on Systems, Man and Cybernetics*, 20(1):518–527.

Brown, C. 1990b. Prediction and cooperation in gaze control. *Biolobical Cybernetics*, 63:61–70.

Brown, C., Coombs, D., and Soong, J. 1993. Real-time smooth pursuit tracking. In *Active Vision*, A. Blake and A. Yuille (Eds.). MIT Press, pp. 123–136.

Carpenter, R.H.S. 1988. *Movements of the Eyes*. Pion Limited.

Christensen, H.I. 1993. A low-cost robot camera head. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(1):69–84.

Christensen, H.I. 1995. Private communication.

Clark, J.J. and Ferrier, N.J. 1993. Attentive visual servoing. In *Active Vision*, A. Blake and A. Yuille (Eds.). MIT Press, pp. 137–154.

Coombs, D.J. and Brown, C.M. 1991. Cooperative gaze holding in binocular vision. *IEEE Control Systems Magazine*, 11(3):24–33.

Coombs, D.J. and Brown, C.M. 1993. Real-time binocular smooth pursuit. *International Journal of Computer Vision*, 11(2):147–164.

de Vlieger, J.H., Verbruggen, H.B., and Bruijn, P.M. 1982. A time-optimal control algorithm for digital computer control. *Automatica*, 18(2):239–244.

Dullerud, G. and Francis, B. 1992. $\mathcal{L}_1$ design and analysis in sampled-data systems. *IEEE Transactions on Automatic Control*, 436–446.

Ferrier, N.J. and Clark, J.J. 1993. The harvard binocular head. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(1):9–31.

Fiala, J., Lumia, R., Roberts, K., and Wavering, A. 1994. TRICLOPS: A tool for studying active vision. *International Journal of Computer Vision*, 12(2/3):231–250.

Krotov, E. and Bajcsy, R., Active vision for reliable ravsing: co-operating, focus, stereo, and vergence. *International Journal of Computer Vision*, 11(2):187–203.

Milios, E., Jenkin, M., and Tsotsos, J. 1993. Design and performance of TRISH, a binocular robot head with torsional eye movements. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(1):51–68.

Mirkin, L. and Palmor, Z. 1997. Sampled-data $\mathcal{H}_\infty$-optimal control with mixed discrete/continuous specifications. *Automatica*, 33(1):1997–2014.

Mirkin, L., Rivlin, E., and Rotstein, H. 1998. On static feedback for the $\mathcal{L}_1$ and other optimal control problems, *International Journal of Control*.

Murray, D., Bradshaw, K., McLauchlan, P., Reid, I., and Sharkey, P. 1995. Driving saccade to pursuit using image motion. *International Journal of Computer Vision*, 16:205–228.

Murray, D.W., Du, F., McLauchlan, P.F., Reid, I.D., Sharkey, P.M., and Brady, M. 1993. Design of stereo heads. In *Active Vision*, A. Blake and A. Yuille (Eds.). MIT Press, pp. 155–172.

Pahlavan, K. and Eklundh, J.-O. 1993. Heads, eyes and head-eye systems. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(1):33–49.

Rivlin, E., Rotstein, H., and Zeevi, Y. 1997. Two-mode control: An oculomotor-based approach to tracking systems. *IEEE Transactions on Automatic Control*, 43(6):833–843.

Robinson, D.A. 1968. The oculomotor control system: A review. *Proceedings of the IEEE*, 56(6):1032–1049.

Sharkey, P.M. and Murray, D.W. 1996. Delays versus performance of visually guided systems. *IEE Proceedings—Control Theory and Applications*, 143(5):436–447.

Swain, M.J. and Stricker, M.A. 1993. Promising directions in active vision. *International Journal of Computer Vision*, 11(2):109–126.

Yamamoto, Y. 1994. A function space approach to sampled-data control systems and tracking problems. *IEEE Transactions on Automatic Control*, 39(4):703–713.