Optimal Servoing for Active Foveated Vision

H. P. Rotstein E. Rivlin

Dept. of Electrical Engineering Dept. of Computer Science
Technion - Israel Institute of Technology
32000 Haifa - Israel

Abstract

Foveated vision and two-mode tracking, as inspired by the human oculomotor system, are often used in active vision system. The purpose of this paper is to provide answers to the following basic questions which arise from implementations. First, is it beneficial to have foveated vision and what is the optimal *size of the foveal window? Second, is there a need for two control mechanisms (smooth pursuit and saccade) for improved performance and how can one efficiently switch between them? In order to do so, a setup is proposed in which these strategies can be evaluated in a systematic manner. It is shown that the fovea appears as a compromise between the tightness of the tracking specifications and computational constraints. Introducing a model for the later and postulating some a priori knowledge of the target behavior, it is possible to compute the size of the fovea in an optimal way. As a by-product, "smooth-pursuit" can be defined in a natural way, and the use of a two-mode tracking scheme is justified. The second mode, i.e. "saccadic control", aims at re-centering the target on the fovea so that the smooth pursuit controller can continue to operate. It is shown that a control strategy can indeed be defined so that this objective can be met under appropriate operating conditions.*

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 [14]. The first active vision systems were constructed in the early eighties and were relatively slow and limited in scope [5, 7]. These first-generation robot heads had poor dynamic characteristics because of the presence of relatively large and heavy cameras and other mechanical and optical difficulties. As smaller and lighter cameras become available, designs were improved so that robot heads have now dynamics comparable to those of the human oculomotor visual system. This has created the need for highly efficient dedicated image processing tools and for control systems capable of exploiting the potential characteristics of the mechanisms. It is then apparent that control should play an important role in constructing fast and precise devices, since PID controllers (which are standard in current designs) may fail to achieve the limits of performance.

The gaze control of robot heads is usually modeled after the human visual system. It consists of a number of low level control loops which interact and –hopefully– cooperate to direct the attention of the system to a desired location. Gaze control can be divided into two primary categories [14]: gaze stabilization or fixation and gaze change. In this paper we will only consider the former, which is more closely related with classic control problems; the latter category usually involves high level planning tasks. Among the gaze stabilization task, one can distinguish between holding fixation on a stationary target, including focus and zoom, and tracking of moving targets.

Some of the fundamental issues in active vision control are: *1)* Large time-delays, *2)* Tight tracking specifications, *3)* Interaction/cooperation between single-loop controllers, and *4)* sampled-data measurements. With respect to the latter, note that 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. An active vision mechanism is then inherently *sampled-data*, with relatively slow sampling rates. To our knowledge, this fact has not been properly stressed in the literature, were it is vaguely modeled as a *time-delay*; this is probably because the absence of good sampled-data control design techniques called for simple, purely continuous time models.

Ferrier and Clark [5] 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 pursuit and saccadic motion. The pursuit loop uses PI control plus some delay in the loop and is inspired by the Smith predictor. Saccadic movements are controlled by a sampled-data loop.

Brown, Coombs and co-workers [2] worked on the control of the Rochester Robot Head and introduced a Smith Predictor and a Kalman Filter to compensate for time delays in the loop. They use PID controllers coupled with predictors for pursuit and switch to an open-loop bang-bang controller for saccadic movements.

Pahlavan and Ekludh [11] studied the Royal Institute of Technology head and again control pursuit and saccade independently, using linear prediction for the saccade. However, since both loops have approximately the same bandwidth it appears to be difficult

1063-6919/96 $5.00 © 1996 IEEE

to distinguish between them in practical operation.

Other works have been reported by Milios et al. [8], Christensen [3] and Fiala et al. [6] 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 (or rather, sampled-data with irregular sampling) controllers based on linear predictions for the saccadic movement. Little information is offered about the tuning of the controllers, which is presumably done on-line and based on the outcome of experiments. Switching between controllers is fired by a positional error larger than some threshold. Pursuit controllers are initialized after each saccade and Kalman filters are apparently used because they provide an easy way to design observers.

Murray et al. [10] 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 pursuit or saccade. Switching from saccade to smooth pursuit is discussed in [9].

Discussion and Main Problems

From the brief account in the previous paragraph, it is clear that several basic issues remain unsolved in spite of the activity in the area, two of which are solved bellow. First, it will be shown that the need of a foveal window can be established based on control considerations. Moreover, the calculation of the *optimal* size of this window can be formulated as a one-parameter maximization. Second, two-mode tracking is shown to be a natural consequence of foveated vision and hence can be formulated in a systematic manner. Saccades are triggered when the target slips out of the fovea, and should be such that after the completion of a saccade, the smooth pursue controller can be switched-on into the loop. It will be shown that this is indeed possible by appropriately defining a target set.

2 Setup and Modeling Considerations

For the purpose of addressing the basic problems of foveated vision and tracking mechanism, it suffices to consider a configuration with a single camera with only one degree of freedom, as well as other simplifying assumptions discussed next. It is worth stressing that one degree of freedom is considered only to simplify the exposition of the main ideas; the extension to more useful cases is essentially straightforward, but for technical details and a more elaborate notation. Hence it is assumed that 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.

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 horizontal, as measured from the axis of rotation. The time required by this processing depends on the amount of data present, i.e., the "size" of the image, and on the sophistication of the image processing algorithm. In our model, the image processing stage is lumped together with other effects like control law computation and communication times, in a pure *time delay* τ proportional to the size of the image (plus some overhead). If τ is larger than the sampling period T, the sequence of images has to be down-sampled by, say q, 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 fixed, the sub-sampling rate will only be a function of the size of the image x (see bellow 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. 1. 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. 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 to reduce the sensitivity of the electro-mechanical system to plant variations, possible neglected nonlinearities 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.

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 if $q > 1$, and only $\hat{e}(k) = \phi(kqT) - \theta(kqT)$ are available for control.

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., $\hat{e}(k)$. 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.

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. The existence of a region of high resolution reduces computational times, thus leading to faster sampling-rates and

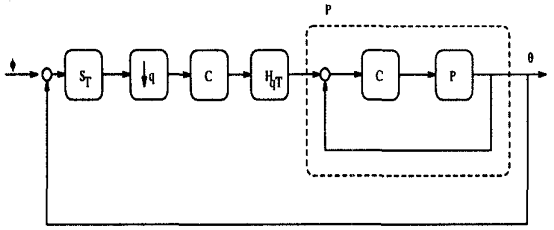

Figure 1: Closed-loop system.

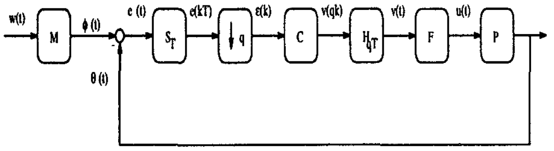

Figure 2: The feedback configuration with reference model

smaller time-delays; this suggests the potential of obtaining a better performances. At the same time, for tracking purposes the image of the target must remain inside the region of high resolution, and this becomes harder for smaller regions. 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. 2. 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 on 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 that can be measured. 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. 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 signal w is assumed to be an integrable function belonging to a set $\mathcal{W}(\alpha)$ parameterized by a positive real number α. The example considered in this paper is

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

which correspond to signals with bounded amplitude. 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 $\mathcal{W}(\alpha)_\infty$.

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; $\mathcal{W}(\alpha)_\infty$ is a reasonable choice whenever little is known a priori about $w(t)$.

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 be stable, and *2)* $|e(t)| \leq x$ for each $t \geq 0$, whenever $w \in \mathcal{W}(\alpha)$. Notice that the specification in *2.* is made in terms of the continuous time error $e(t)$ and not the sampled one $\epsilon(k)$ which is available to the controller. The reason for this is that concentrating in $\epsilon(k)$ 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.

The existence of a controller that satisfies the above criterion will in general depend on α, since $e(t)$ cannot be guaranteed to be small for arbitrarily "large" signal. It is then natural to consider 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 $w(t) \in \mathcal{W}(\alpha^x)$.*

In mathematical terms, Problem 1 may be written as:

$$\alpha^x = \sup\{\alpha \; : \; \inf_{C \in \mathcal{C}} \sup_{w \in \mathcal{W}(\alpha)} |e(t)| \leq x, \;\; t \geq 0\} \qquad (2)$$

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. Denoting by $T_{ew}(C)$ the transfer function between w and e for a given stabilizing controller C, consider the system norm:

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

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:

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

[1] Except for some smoothness assumption that any physical model should satisfy.

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

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

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 $\|e(t)\|_\infty \leq x$ 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 particular for $i = \infty$ C^x is called an $\mathcal{L}_1$-optimal controller.

The performance γ^x depends on x via the subsampling 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.

α^x will be small both for $x \approx 0$ and, assuming the typical case that γ^x grows faster than a linear function, for large values of x. The maximum of α^x will then be achieved for some finite value of x, say $\alpha^* \doteq \max_x \alpha^x$. 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. Therefore, *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 $C^s = C^{x^*}$ is 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: For the case $\mathcal{W}(\alpha)_\infty$ of interest, the controller C^s is linear time-invariant and can be computed using recently developed algorithms.

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 *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 point out that *b)* appears to be the factor which is 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. The objective is not only to center the target in the fovea at some time $t_s > t_v$ but also to guarantee that the smooth controller will be able to perform satisfactory if the assumption on $w(t)$ is satisfied for $t \geq t_s$. Since performance is very poor in the interval $[t_v, t_s]$, a natural objective is to make this interval as short as possible. This is referred in the sequel as "saccades", and is 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. 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 do that, it is necessary 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.

4.1 Switching Between Controllers

A scheme is proposed in this section for switching between smooth-pursuit and saccades, which is later used for defining saccadic control. Due to space constraints, only the major ideas will be outlined. Let T_{ew} denote the closed-loop operator from w to e. Although T_{ew} is time-varying, it is possible to come up with a state-space realization with internal state $x_s(k)$, formed by stacking together the state vectors of the reference model, the plant and the controller, x_M, x_P and x_C respectively. Given an initial state $\hat{x}_0$ at time t_0 and some (integrable) function $w(t)$, let $\mathcal{F}_S(k, \hat{x}_0, t_0, w)$ denote the linear function mapping x_0 into the state trajectory $\hat{x}_S(k)$:

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

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

$$\mathcal{R}_S \doteq \{\hat{x}_S = \mathcal{F}_S(k_f, 0, 0, w) \text{ for some } k_f, w \in \mathcal{W}(\alpha^*)\}.$$

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

$$\begin{aligned} [\ I \quad 0 \quad 0\ ]\,\mathcal{F}_S(k_s,0,0,w) &= \hat{x}_M^o \\ [\ 0 \quad I \quad 0\ ]\,\mathcal{F}_S(k_s,0,0,w) &= \hat{x}_P. \end{aligned}$$

The set $\mathcal{O}(\tau, x_M^o)$ 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

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

It follows from the reasoning in [13], that $|e(t)| < x$ for $t \geq \tau$ if the future disturbances $|w(t)| \leq \alpha^*$. This is because 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).

4.2 Saccadic Control

The discussion in the previous section provides the framework for the systematic treatment of saccadic control. 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 trigger, 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^{tv} in such a way that the error criterion remains constant from $t_v = k_v h$ and up to the instant where the saccadic control is employed.

Modeling

In order to reduce the error signal bellow x at some future instant τ, it is necessary to predict the values of the signal $\phi(t)$ for $t \geq \tau$, based on measurements which are usually costly to obtain and potentially contaminated by noise. The success of the saccadic control action is usually related with the accuracy of these predictions. As an example, suppose that the target changes its position to some stationary point lying outside the foveal window (this is a standard experiment when evaluating human saccades [12]); then the modeling problem reduces to determine the new position, which can presumably be done accurately. Otherwise, predictions on moving objects are much harder, so that additional "corrective" saccades may be required.

The computation of models for prediction under different sets of assumptions is considered in detail in [1], which contains an array of different algorithms. The specific algorithm should be selected depending on the standing assumptions for $\phi(t)$ and the noise which is possibly corrupting the measurements. This selection is important since it will determine the time lag required to have a prediction of future position and how accurate that prediction will be. A popular choice in the active vision field is to select an $\alpha-\beta$ or $\alpha-\beta-\gamma$ filters, which have the advantage of their simplicity. Coefficients of this filters are usually selected by using the steady-state solution of a Kalman filtering problem [1]. However, much better predictions can be made if a priori knowledge of the variations of $\phi(t)$ are available and exploited, for instance, if the objective is not to track a moving target but to do a gaze shift.

Finally, since all the internal state of the reference model is required for computations, this information should be readily obtainable.

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 assume 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:

$$w(t) = \hat{w}(k) \qquad k\hat{h} \leq t < (k+1)\hat{h}$$

where $\hat{h} << T$. This reduces the problem to a discrete-time multi-rate one. The advantage is that in that case linear programming based algorithms exists [4] 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 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.

An idealized setup was considered in this paper to simplify the presentation of the ideas, and several adjustments might be required in actual implementations. 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.

As a final remark, the present paper differs from most of the works on the area of active vision by stressing and proposing solutions to the control aspects of the problem. In doing so, a framework has been provided on which issues that have been discussed before in the field can be accommodated, including image processing topics and/or the modeling of the position of the target.

References

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

[2] C. Brown, D. Coombs, and J. Soong. Real-time smooth pursuit tracking. In A. Blake and A. Yuille, editors, *Artificial Intelligence*, pages 123–136. MIT Press, 1993.

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

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

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

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

[7] E. Krotkow. *Active Computer Vision by Cooperative Focus and Stereo.* Heidelberg: Springer Verlag, 1989.

[8] E. Milios, M. Jenkin, and J. Tsotsos. 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, 1993.

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

[10] D. W. Murray, F. Du, P. F. McLauchlan, I. D. Reid, P. M. Sharkey, and M. Brady. Design of stereo heads. In A. Blake and A. Yuille, editors, *Artificial Intelligence*, pages 155–172. MIT Press, 1993.

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

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

[13] H. Rotstein and L. Mirkin. On static feedback for the $\mathcal{L}_1$ and other optimal control problems. In preparation, 1995.

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