

An Active Visual Estimator for Dexterous Manipulation

Alfred A. Rizzi, *Member, IEEE*, and Daniel E. Koditschek, *Member, IEEE*

Abstract—We present a working implementation of a dynamics based architecture for visual sensing. This architecture provides field rate estimates of the positions and velocities of two independent falling balls in the face of repeated visual occlusions and departures from the field of view. The practical success of this system can be attributed to the interconnection of two strongly nonlinear dynamical

systems: a novel triangulating state estimator; and an image plane window controller. We detail the architecture of this active sensor, provide data documenting its performance, and offer an analysis of its soundness in the form of a convergence proof for the estimator and a boundedness proof for the manager.

I. INTRODUCTION

WE have built a three degree of freedom robot that bats two balls into simultaneous stable periodic vertical trajectories that commonly continue for greater than an hour [26]. The juggling algorithm underlying this behavior relies on continuous estimates of ball position and velocity. This paper examines both the practical and theoretical issues involved in generating these estimates from a stereo camera system that views brightly illuminated white balls against a dark background. Despite this structured visual environment there remain a number of significant issues surrounding the efficient acquisition and use of camera data to generate the required information in a timely manner. Specifically, we have developed both a rational attention management scheme and a novel triangulating observer that ensures a stable flow of information even in the presence of unavoidable transient losses of data.

At a time when many in the robot vision community are exploring the benefits of “visual servoing” or have found the need for including attention mechanisms in their camera architectures, we offer this account as documentation of a particular system which seems to incorporate most of the essential features of an “active sensor” yet remains simple enough

Manuscript received January 22, 1996. A. A. Rizzi was supported in part by IBM through a Manufacturing Graduate Fellowship and in part by the National Science Foundation under Grants IRI-9123266 and 9396967. D. E. Koditschek was supported by the National Science Foundation in part by a Presidential Young Investigator Award, and in part under Grants IRI-9123266 and 9396967. Portions of the material presented in this paper have appeared previously in [23] and [22]. A previous draft of this paper has been circulated as a technical report [24] and presented at the 1994 International Conference on Robotics and Automation Workshop on Visual Servoing. This paper was recommended for publication by Associate Editor R. V. Dubey and Editor S. E. Salcudean upon evaluation of reviewers' comments.

The authors are with the Artificial Intelligence Laboratory, Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor MI 48109-2110 USA.

Publisher Item Identifier S 1042-296X(96)07608-2.

to permit some formal analysis. Developing an architecture amenable to analysis yields a system whose run time behavior can be understood within a simple paradigm. This in turn has greatly expedited the inevitable debugging and gain tuning cycles in our laboratory.

Developing and reasoning formally about this specific system interests us more generally in view of the apparent need to develop a theory and practice of “dexterous robots.” This term, as we understand it, denotes an autonomous machine capable of interacting with a dynamical world. The strategies of general interest to us are feedback algorithms which specify the manipulator's actions at each instant in time as a function of its current state and that of the world. For a juggling machine, the world's state reduces to the current position and velocity of one or two balls and the task of estimating this state forms the narrow focus of the present paper. It is our belief that a much larger range of dynamically dexterous tasks (of which juggling is but a simple example) will necessitate the ability to generate timely and accurate estimates for the state of a dynamical environment independent of the specific control algorithm. We are hopeful that analytically tractable sensor systems of the type proposed here will promote the development of robots that are both practically and theoretically sound as well as behaviorally complex.

A. Contributions of the Paper

Beyond the description in Section III of “yet another successful laboratory architecture,” the paper presents two separate but interrelated analytical contributions highlighting the features of this architecture that we believe account in large measure for its success.

The need for a nonlinear estimation procedure is a direct result of our decision to rely, to as great an extent as possible, on a dynamical model of ball flight rather than a geometric model of ball shape. In consequence, we choose to throw away images that are “too complicated” to interpret at any instant. Thus, we require a means of integrating the resulting intermittent measurements from the uncoordinated cameras. Section IV presents an algorithm for doing so: a new approach to Cartesian state estimation based on filtering image plane error vectors, rather than their triangulated versions. The proposed estimator makes use of a structural feature of the perspective map to achieve triangulation in much the same way visual servoing makes direct use of visual data to achieve task level goals. We are not aware of any previous work in

the field that exploits this useful structural property of the camera transformation and we call it to the reader's attention since it has already proven helpful in other contexts [17]. We demonstrate the stability of the new estimation system assuming the availability of continuous measurements. The results of this analysis provide sufficient insight into the behavior of the discretely sampled plant (relevant to a real camera system) as to afford a working implementation that we document.

The need to introduce an active component to the sensing system stems from the rapid and dramatic changes in the data available to the camera system arising from the constant appearance and disappearance of one or two balls during a successful juggling run. In Section V we study the consequences of using a feedback control to cue the data processing stage of a stereo camera system. We define the robot's state of attention as a volume of space within which the sensor will search for a ball, and prescribe a strategy for choosing the placement and size of this volume. If this volume is too large, then image processing will result either in untimely or inaccurate measurements, which, in turn, lead to increasingly inaccurate future estimates. Conversely, if the volume is chosen to be too small, then the ball may be entirely missed due to inaccuracies in the present estimates. We succeed in showing that our simple volume manager is aggressive enough to capture a lost ball but not so aggressive as to lose sight of a previously spotted ball.

B. Relation to the Existing Literature

The two themes of visual servoing and active vision have been the focus of significant attention in the recent robotics literature and it seems useful to relate our contributions to that larger body of existing work.

There are several relevant dimensions of comparison to be made. First there is the question of the task to be performed: interest may lie either in attempting to control or merely in estimating some aspect of the observable world. Second, either class of problem may be addressed with or without prior knowledge of an exact model. Finally, it may or may not be important to account for the effects of data processing errors and delays on the overall control loop. In this paper we are concerned only with the estimation problem, and assume a priori knowledge of model parameters; moreover, we find the (predominantly systematic) errors introduced by the data processing segment of our system significant enough to warrant compensation.

1) *Visual Servoing and Visual Estimation:* We conceive of our robot's perceptual apparatus as the dual of what has come to be called in the robotics literature a visual servoing system. Researchers working in this area in the last few years [8], [14], [19], [11], [12], approach the problem domain by devising visually encoded goals for controllable environments—they move the robot (or camera) until the scene (as seen by the machine) matches some desired image(s). Although the estimated state generated by our sensing system is eventually used to achieve control, in this paper we are solely concerned

with the performance of the estimator. The estimator must adjust its estimate of the world state in such a manner that its internal state visually aligns with the measured data. It is this similarity that leads us to consider the estimation system described here as a traditional dual to the visual servoing systems of our peers.

Visual estimation tasks have been pursued for some time by Allen and colleagues [1], [2] who employ statistical filtering techniques to smooth monocular camera data and predict Cartesian motion of unknown objects with sufficient accuracy to pick them up in a general purpose gripper. Even more generic visual estimation problems have been addressed by Rao, Durant-Whyte, and colleagues [21] who approach the problem as a multi-target acquisition and tracking problem reminiscent of the tracking and acquisition work of Bar-Shalom [5]. The problem at hand here is much narrower—estimate the state of a falling ball—and the success of the ultimate manipulation—bat the ball with a flat paddle at a particular position and orientation—is more sensitive to the accuracy of state estimates. In consequence, we find it more appropriate to work within a strictly deterministic framework.

As far as we are aware there is only one other group that has attempted to develop deterministic nonlinear state estimators relevant to vision applications. Ghosh and colleagues have recently announced a family of image plane state estimators for known linear Cartesian dynamical models [13]. Their observers differ from ours by taking state on the image plane (resulting in nonlinear dynamics), in contrast to the Cartesian workspace of the robot (resulting in linear dynamics with a nonlinear output map). Our proposed system is a stereo observer (its state is in Cartesian three-space) and thereby performs triangulation automatically as well, whereas the strategies of [13] only predict motion on the image plane and thus far have not been used to fuse data from multiple cameras.

2) *Model Parameters:* Our online perceptual loop is designed using a priori knowledge of the ball's dynamics and the hand-eye kinematics. Thus, as might be expected, calibration has proven terribly important in our work [24], [25]. Our modest contribution to the traditional off-line nonlinear regression paradigm has been to cast the error cost function in the image plane coordinates as opposed to the usual formulation in Cartesian coordinates and this seems to offer certain numerical benefits relative to previously reported approaches [6].

In contrast, the field has witnessed an emerging interest in parameter adaptive techniques for computing this calibration data online. Ghosh and colleagues [16] have developed an approach to monocular image plane adaptation for linear Cartesian dynamics. Were they extended to the stereo setting, these ideas might well provide us online means of calibration—to our knowledge this direction remains unexplored. Recently, Hager [14] has reconsidered the visual servoing problem without the presumption of complete a priori calibration. His methods conform to the direct adaptive control paradigm and we have begun to explore the connection between it and this work.

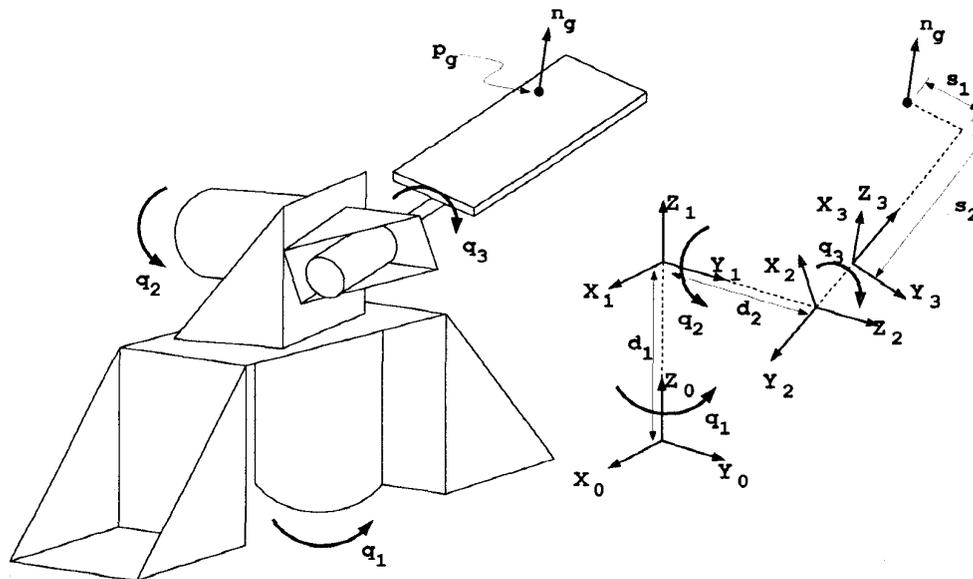


Fig. 1. The Bühgler arm (left) and its kinematics (right) [27].

3) *Active Vision*: There is a large robotics literature concerned with the introduction of active control policies into the perceptual loop. Most relevant to our work is the example provided by Dickmanns and colleagues [10] who cue their data processing system with estimates derived from various dynamical models of the world. Our attention control mechanism (an image plane windowing manager with dynamics caused by delay) may be seen as a concrete example of the sort of architectures they have proposed. Our problem domain, however, is much narrower (we require only one world model) and we focus on a specific implementation that seems amenable to formal modeling and stability analysis.

II. SETTING

The sensor systems we wish to build require an understanding both of how the world they attempt to sense evolves over time, and how they perceive that world. Thus we pause here to develop simple models for the falling and bouncing ball, the robot juggling strategy, and the robot's physical sensors (each of which will be used below).

A. Physical Models: The Robot's Environment

In general, state estimation can be only as effective as the environmental model available. For the juggling problem, the model in question will consist of two parts: ball flight, which describes the behavior of a ball under the influence of gravity; and impact, which describes how a ball will bounce when it makes contact with the robot's paddle.

1) *Flight Model*: For simplicity, we have chosen to model the ball's flight dynamics as a point mass under the influence of gravity. A position and time-sampled measurement of this dynamical system will be described by the discrete dynamics,

$$w_{j+1} = F^s(w_j) := A_s w_j + a_s;$$

$$A_s := \begin{bmatrix} I & sI \\ 0 & I \end{bmatrix}; \quad a_s := \begin{bmatrix} \frac{1}{2} s^2 \tilde{a} \\ s \tilde{a} \end{bmatrix} \\ b_j = C w_j; \quad C = [I, 0] \quad (1)$$

where s denotes the sampling period, \tilde{a} is the gravitational acceleration vector, and $w_j := (b_j, \dot{b}_j) \in \mathbb{R}^6$ embodies the entire state of the object (its position and velocity).

2) *Impact Model*: To understand the effects of impact, consider a ball with trajectory $b(t)$ colliding with the paddle in robot configuration $q \in \mathcal{Q}$ (depicted in Fig. 1) at some point, s , on the paddle which has a linear velocity v . We seek a description of how the ball's phase, $w := (b, \dot{b})$, is changed by the robot's phase, (q, \dot{q}) , by such an impact event.

As in earlier work [7], [25] we will assume that the components of the ball's velocity tangent to the paddle at the instant of contact are unchanged by impact, while the change in the normal component is governed by the simplistic (but standard [28]) coefficient of restitution law. For some $\alpha \in [0, 1]$ this impact model can be expressed as

$$(\dot{b}'_n - v'_n) = -\alpha(\dot{b}_n - v_n)$$

where \dot{b}'_n and v'_n denote the normal components of the ball and paddle velocities immediately after impact, while \dot{b}_n and v_n are the velocities prior to impact. Assuming that the paddle is much more massive than the ball, we conclude that the velocity of the paddle will remain constant throughout the impact ($v' = v$). It then follows that the coefficient of restitution law can be rewritten as $\dot{b}'_n = \dot{b}_n + (1 + \alpha)(v_n - \dot{b}_n)$, and hence

$$\dot{b}' = \dot{b} + (1 + \alpha)nn^T(v - \dot{b}) \quad (2)$$

where n denotes the unit normal vector to the paddle.

B. Behavioral Model: The Robot's Strategy

A detailed development of our juggling control strategy can be found in [22]. Briefly, the "mirror law," is a map (m)

from the phase space of a ball into the configuration space of the robot. Thus the robot's reference trajectory is specified by $q_d(t) = m(w(t))$. The function, m is defined as follows. Using (6) from [26], define the joint space position of the ball $[\phi_b, \theta_b, \psi_b, s_b]^T := g^{-1}(b)$, where g^{-1} is the inverse kinematic map (including the paddle's length s which provides an effective fourth degree of freedom) for our machine, which is shown in Fig. 1. Our robot strategy causes the paddle to: (i) track under the ball at all times; (ii) "mirror" the vertical motion of the ball through the action of θ_b on q_{d2} as expressed by the original planar mirror law [7]; (iii) raise and lower the paddle, resulting in the normal being adjusted to correct for radial deviation in the ball position; (iv) roll the paddle, again adjusting the normal so as to correct for lateral position errors.

C. Perceptual Model: Stereo Perspective Projection

Two fields of image data are simultaneously acquired from two cameras. It is then the responsibility of the sensing system to report the location and velocity of the ball (or balls) in space so the the juggling algorithm described above may be executed. As stated previously, the visual environment is structured such that an individual pixel may be identified as either part of a ball or the background simply by its intensity—we are looking for white balls against a black background. This structure allows us to use a "simplistic" geometric model of the world (pixels are either part of the ball or not) to simplify the image processing. Although we have chosen to make use of structured lighting, the environment is far from uniform. As a ball traverses the visual workspace of the machine its appearance changes shape due to lighting effects [15]. Thus a geometry based vision system could reliably report ball locations only if it were capable of taking into account these poorly modeled lighting effects. As will be seen shortly, we have chosen to make use of a dynamic model of the ball's flight to make up for this lack of geometric detail.

The simple projective stereo camera model of the form

$$c: \mathbb{R}^3 \rightarrow \mathbb{R}^4$$

(which maps positions in affine 3-space to a pair of image plane projections in the standard manner) has been sufficient for the experiments associated with this paper. More precisely, c , is formed by stacking together the perspective projections due to the two individual cameras,

$$v := c(b) = \begin{bmatrix} \pi_{f_1} \circ h_1(b) \\ \pi_{f_2} \circ h_2(b) \end{bmatrix} \quad (3)$$

where

$$\pi_f(b) := \frac{f}{b_3} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

h_i is the rigid transformation representing the base frame in the i th camera's frame, and f_i is the i th camera's focal length.

Knowledge of the cameras' relative positions and orientations together with knowledge of each camera's lens characteristics (at present we model only the focal length) permits the selection of a "pseudo-inverse" or "triangulation-

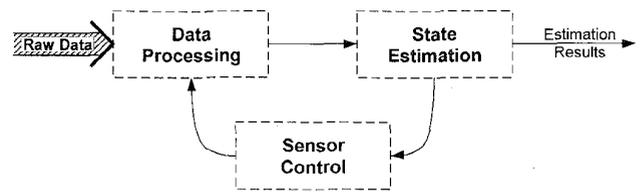


Fig. 2. A generic active visual estimator.

function,"

$$c^\dagger: \mathbb{R}^4 \rightarrow \mathbb{R}^3 \quad (4)$$

such that $c^\dagger \circ c$ is the identity on \mathbb{R}^3 . We have discussed our choice of pseudo-inverse at length in previous publications [26], and details of the calibration scheme can be found in [25].

Estimation of ball state begins anew after each impact using initial conditions predicted by (2) and the state estimates prior to impact. One drawback of using a camera system as the primary sensor is its low data rate, which makes determining the exact time of impact difficult. We have chosen to augment the juggler's sensing system with a microphone attached directly to the robot's paddle to serve as an *impact detector*.

III. AN ACTIVE VISUAL ESTIMATOR

A complete sensing system for an environment such as that just presented requires the careful integration of a number of functional submodules. We now describe our experience constructing such a system as well as the manner in which those experiences have led to revisions and complications in the architectural framework.

A. The Challenge of Constructing an Integrated System

Ideally, a purely Cartesian sensor could be purchased or designed which would be capable of providing the continuous state estimates necessary for implementation of a controller of the type described in Section II-B. Currently, however, such sensors are either prohibitively priced or lack sufficient sophistication to cope with anything but the most stringently structured environment. We thus face the task of designing our own sensing system.

It seems natural to partition a sensing system suitable for this type of dynamical task into three subsystems as shown in Fig. 2. This architecture separates the sensing system into the following modules: *data processing* encompasses the algebraic (memoryless) signal processing; *state estimation* contains the dynamic or model based processing; and finally, *sensor control* implements the feedback segment responsible for guiding the "attention" of the low level data processing. This architecture can be found in nearly any system, and we find that thinking about the overall behavior in terms of these separate modules is advantageous.

As has been described by others [9], [10], [19], an active sensing system, such as depicted in Fig. 2 can be used to reduce the load of incoming data by focusing the attention of the machine only where meaningful information is likely to be found. More fundamentally, focusing the attention of

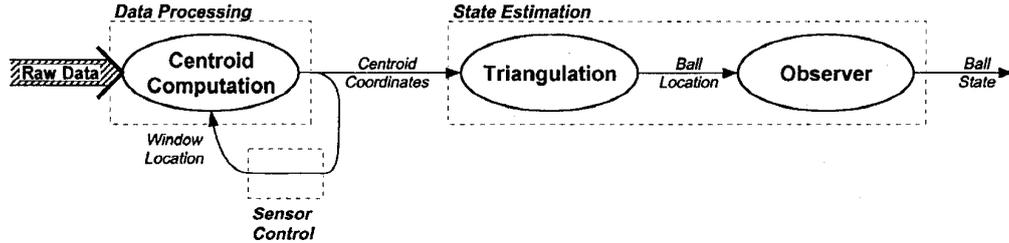


Fig. 3. Our initial implementation of the active visual estimator.

a machine can be used as a means to introduce knowledge about the environment's dynamical behavior back into the data processing task, thereby making the segmentation and feature extraction tasks more tractable.

1) *An Initial Design:* Our original design (depicted in Fig. 3) followed traditional engineering principles for reliable signal processing and state estimation modules.

Data Processing: Following Andersson's experience in real-time visual servoing [4] we chose to employ a first order moment computation applied to a small window of a threshold-sampled (thus, binary valued) image of each field. Thresholding, of course, presumes a visually structured environment, which we have provided in the form of white ping-pong balls illuminated with halogen lamps against black matte cloth. The result is a pair of pixel addresses containing the centroid of the single illuminated region seen by each camera. For the remainder of this presentation we will denote by W_k the function that takes a white ball against a black background into a pair of thresholded image plane regions and then into a pair of first order moments at the k th field

$$v_k := W_k \circ c(Cp_k).$$

We use $p_k := F^{-\tau_f}(w_k)$ as an "extra" state variable to denote the delayed image of the ball's state due to image acquisition and processing delays.

Sensor Control: Computational resources in our juggling system preclude examining more than about 2400 pixels from any given video field (our digitization system delivers individual fields at a rate of 60 Hz). Thus the system is forced to process subwindows from the images to assure completion of the image processing task before the arrival of a new field. Fig. 3 depicts the trivial sensor control strategy used in this initial design, which functions by centering the window for a new field over the location of the centroid from the previous field. This strategy implicitly presumed that objects do not move (or at least they do not move far) between images.

Triangulation: Computation of spatial locations from centroid data was initially performed via exact algebraic triangulation, c^\dagger (4), which may be written as

$$\hat{C}(p_k) := \bar{b}_k = c^\dagger \circ W_k \circ c(Cp_k) \quad (5)$$

to make explicit the role of the *data processing* module.

State Estimation: Due to digitization and processing latency, the image measurements generated by the *data processing* section are results from images that are at least one field (16 ms) old. We used an observer which operates on this delayed data,

$$\hat{p}_{k+1} = F^{\tau_f}(\hat{p}_k) - G(C\hat{p}_k - \bar{b}_k) \quad (6)$$

where the gain matrix, $G \in \mathbb{R}^{6 \times 3}$, is chosen so that $A_{\tau_f} + GC$ is asymptotically stable—that is, if the true delayed data, Cp_k , were available then it would be guaranteed that $\hat{p}_k \rightarrow p_k$.¹

Drawbacks: As detailed above, it is not the ball's position, b_k , which is input to the observer, but the result of a series of computations applied to the delayed copies of the cameras' image planes, \bar{b}_k . Prior to attempting two-juggle experiments, we ignored this "detail" and ran with the open loop sensory management procedures used to obtain data (5) [22]. It soon became clear that these procedures could not be similarly transparent in the more demanding domain of the two-juggle task. The practical limitations of our robot arm necessitated considerable enhancements to the vision subsystem, and getting these management issues right became one of the chief sources of difficulty.

For reasons detailed in [23] the considerable torque generating capabilities of our Bühgler arm did not prove sufficient to permit easily tracked ball trajectories in the two-juggle setting. We were forced to juggle higher (longer flight times between impacts) and to bring the two balls much closer together in space (shorter distance between impacts) than had been originally planned. This necessitated adding two new corresponding features to the vision system. First, we required an ability to sense and recover from out of frame events (a ball passing out of the field of view due to the height of the juggle). Second, we required that the system handle regularly occurring ball occlusions (two balls appearing at or near the same location in an image).

Neither the *data processing* nor the *sensor control* module described above are equipped with mechanisms suitable for

¹In principle, one might choose an optimal set of gains, G^* , resulting from an infinite horizon quadratic cost functional, or an optimal sequence of gains, $\{G_i^*\}_{i=0}^k$, resulting from a k -stage horizon quadratic cost functional (probably a better choice in the present context), according to the standard Kalman filtering methodology. Of course, this presumes rather strong assumptions and a significant amount of a priori statistical information about the nature of disturbances in both the free flight model (1) as well as in the production of \bar{b} from \hat{p} via the moment generation process. To date we have obtained sufficiently good results with a common sense choice of gains G that recourse to optimal filtering seems more artificial than helpful.

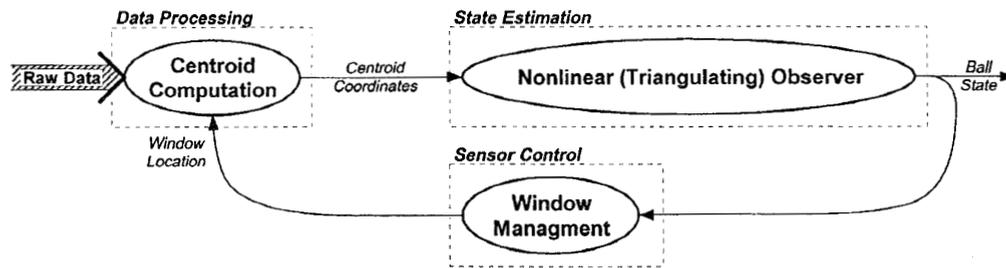


Fig. 4. Idealized implementation: The active visual estimator.

handling either of these events. In particular the *data processing* module is incapable of recognizing *occlusion* events, and produces erroneous measurements in their presence, while the naive *sensor control* strategy will never be able to reacquire a ball which leaves the field of view unless it returns near enough to the point of departure.

B. An Integrated Solution

Our solution to these problems (introduced immediately below) although a natural extension to the basic design, results in a much more complicated sensing system whose analysis forms the remainder of this paper. In order to encourage a more intuitive understanding of these modifications, we briefly digress into a historical presentation of the development of the system shown in Fig. 4.²

1) *Data Processing—Making Use of Dynamic Segmentation*: As mentioned above, the inability of the *data processing* subsystem to recognize uninterpretable images represents the fundamental weakness of our initial design. Fixing this problem is not difficult, however, the modifications result in a fundamental change in the operation of the sensing system.

The construction of a system capable of handling *occlusion* and *out of frame* events must include the capability to either detect and reject images containing such events, or to reliably extract the relevant information in spite of these events. Clearly, in the case of out of frame events, there is no choice but to predict future behavior without new information. Having already committed to measuring the first order moments of a binary image as the primary method of localization, it is natural to extend this notion and use the zeroth and second order moments as simple and robust *occlusion detectors*. Under reasonably well-structured lighting conditions, the “ballness” of an image-segment is easily determined by placing thresholds around the zeroth order moment and the the ratio of the eigenvalues of the matrix of the second order moments in conjunction with a test on the planar orientation of its eigenvectors.

2) *Sensor Control—Feedback for Active Vision*: By choosing to reject uninterpretable images in the data processing system we have implicitly assumed that the higher levels of the sensing system will be able to guide future measurement efforts. Specifically, we expect the *sensor control* module to

²Note, in our laboratory implementation we have found it expedient to incorporate two parallel observers alongside the nonlinear (triangulating) observer pictured here for reasons mentioned in Section V-D.

supply clues sufficient to guarantee that a temporarily ignored object will be reacquired.

Window Placement: Regular *occlusion events* (because the balls are purposefully juggled high and close together), in conjunction with the policy outlined above of ignoring data from occluded windows severely compromises the effectiveness of the simple window placement scheme of Section III-A1. An obvious improvement results from using the position estimates produced by the observer to place the windows. Namely, the search windows in the next image to be processed are centered at a point formed by projecting the predicted state estimate from the observer onto the camera image planes. Thus, the window location is now fed back from the output of the estimator whose inputs it provides. This connection of the observer back to the low-level data processing forms the *sensor control* module discussed above, and comprises the active vision component of this system.

Window Size Adjustment: Our inability to compute with more than a small percentage of the available pixels during the 16 ms. interval between successive camera fields forces a tradeoff between the accuracy of the centroid data input to the observer and the possibility of an unnecessary and unrecoverable out-of-window event. This tradeoff is governed by the choice of sampling resolution or, equivalently, image plane window area.

There are three principal sources of error in the sensing system. First, noise inevitably corrupts the image processing (e.g., distortions introduced by thresholding an imperfectly illuminated scene, or by insufficient spatial resolution). Second, the observer is itself compromised by parametric errors (e.g., the gravitational force, \tilde{a} in (1) is obtained through our calibration procedure) and omissions (e.g., there is no model of spin during flight). Finally, these are exacerbated by the intermittent loss of input data that attends occlusion events (e.g., out-of-frame events may easily last in excess of 0.25 s).

Section V offers a formal presentation of the system theoretic ideas which support our current implementation. Essentially, we grow the window area following any image plane measurement failure (i.e., an occlusion event), while the window area is shrunk following valid measurements. The exact size of the window needed to guarantee a successful future measurement is derived by bounding the current error in the *state estimator*, so as to ensure that the window will encompass the actual location of the ball.

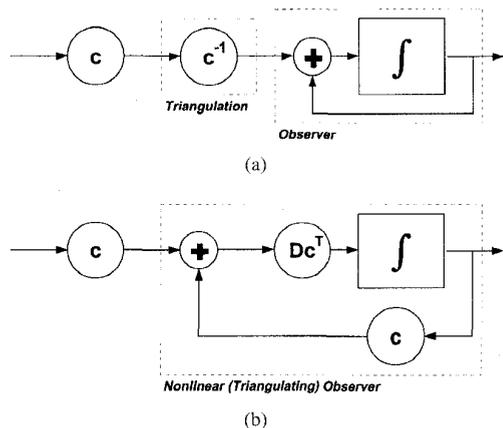


Fig. 5. A comparative illustration of the *state estimation* boxes of Figs. 3 and 4: (a) Direct or algebraic inversion, (b) Dynamical inversion through the Jacobian, Dc^T .

3) *State Estimation—A Nonlinear (Triangulating) Observer*: A central difference between the system presented in Section III-A1 and the one discussed here arises from the idea of discarding data from individual cameras whenever the image is difficult to interpret. The significant side-effect of this change is apparent when we look at the algebraic triangulator used to supply spatial ball positions to the linear observer. The system is unable to perform triangulation whenever data from *either* camera has been rejected, and thus new inputs can not be provided to the observer. Since it is unlikely for data from both cameras to be invalid simultaneously, discarding questionable data from one camera apparently forces the system needlessly to discard valuable data from the other.

IV. DYNAMIC TRIANGULATION

We now embed the triangulation process directly in an observer and thereby make use of all the available data at all times, while continuing to guarantee convergence of the state estimates. Our previously mentioned “waste” of data arose from the use of an algebraic inverse to transform image plane measurements into spatial positions. The alternative we present here performs this inversion implicitly in a dynamical filter.³ Fig. 5 depicts the structural difference between these approaches.

Underlying the new estimation technique is the simple idea of augmenting the standard (linear) Newtonian flight model, $\dot{b} = \tilde{a}$, with a nonlinear output map, $v = c(b)$, and constructing a nonlinear observer which updates its state estimates based on the image plane data rather than a spatial “measurement” (derived through triangulation). The structure for this new observer is shown in Fig. 6(b). The significant change here is to abandon the use of an algebraic triangulation function to invert $c(b)$ and instead revert to using a dynamical system

³In much the same way, an integrator in a feedback loop can be used to implicitly form the derivative of an input signal, or an analog computer can use gradient descent to find the root of a system of equations.

to simultaneously smooth, predict, and perform this inversion, all through its update law for the estimated state.⁴

A. A Useful Property of the Perspective Projection Map

Recall from (3) that the stereo camera transformation, c , is formed by stacking together the perspective projections due to the two individual cameras. In this section we note that

$$\begin{aligned} c(\hat{b}) - c(b) &= \Lambda(\hat{b}, b)C(\hat{b})(\hat{b} - b) \\ &= \Lambda(b, \hat{b})C(b)(\hat{b} - b) \end{aligned} \quad (7)$$

where $C(b)$ is the Jacobian of c evaluated at b and

$$\Lambda(\hat{b}, b) := \begin{bmatrix} \frac{\Pi_3 \circ h_1(\hat{b})}{\Pi_3 \circ h_1(b)} I_2 & 0 \\ 0 & \frac{\Pi_3 \circ h_2(\hat{b})}{\Pi_3 \circ h_2(b)} I_2 \end{bmatrix}$$

where Π_3 denotes Cartesian projection of \mathbb{R}^3 onto its 3rd coordinate.

This fact emerges directly from computation. Given b lying in the frame of reference of a camera with focal length f , we have

$$\pi_f(b) := \frac{f}{b_3} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \quad (8)$$

The Jacobian of this projection is then given by

$$D_b \pi_f(b) = \frac{f}{b_3} \begin{bmatrix} 1 & 0 & -\frac{b_1}{b_3} \\ 0 & 1 & -\frac{b_2}{b_3} \end{bmatrix}. \quad (9)$$

Expanding the right hand side of the top row of (7) in these coordinates gives

$$\frac{f}{b_3} \begin{pmatrix} \hat{b}_1 \\ b_3 \hat{b}_3 - b_1 \end{pmatrix} \quad (10)$$

and similar results follow for the remainder of the rows. This establishes the original assertion.

Throughout the remainder of this section we will need to ensure that Λ is positive definite. Geometrically, this implies that both b and \hat{b} always lie on the same side (front/back) of all the cameras at all times. In practice this is not an unrealistic assumption: it merely requires that neither the actual object cross the singularity in (3) nor that the errors in the observer system become so large as to cause the estimated object location to cross this same singularity. Formally, this assumption allows us to assert that

$$(\hat{b} - b)C^T(\hat{b})(c(\hat{b}) - c(b)) > 0 \quad (11)$$

and we will demonstrate that when (11) is initially true that the strategy of Fig. 6(b) preserves this property.

⁴The expectation here is that beyond the efficiency achieved by not wasting good data, such a system will exhibit better noise immunity since it does not directly attempt to invert c . However, we do not attempt any formal justification of this claim in the present paper.

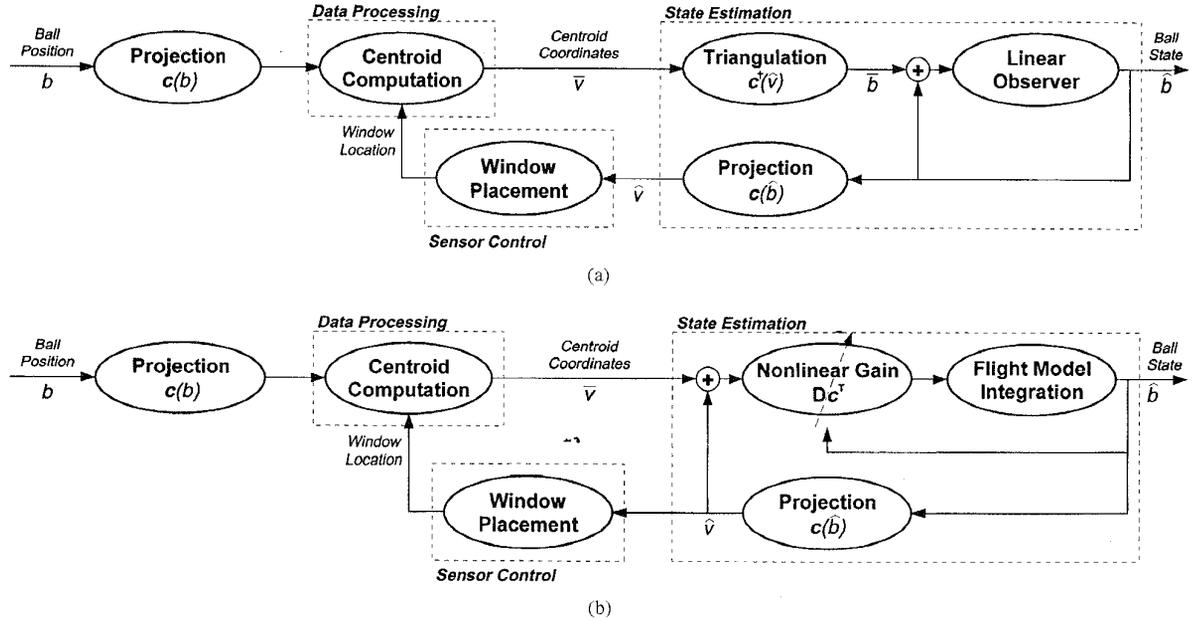


Fig. 6. Structure of the two observer schemes: (a) Original design; (b) Improved design.

B. An Image Based Observer for Second Order Linear Dynamics

Let us now reconsider the system, $\ddot{b} = \ddot{a}$, written more generally as

$$\begin{aligned} \dot{w}_1 &= w_2 \\ \dot{w}_2 &= A_1 w_1 + A_2 w_2 + u \\ v &= c(w_1) \end{aligned} \quad (12)$$

where $w_1 := b$ and $w_2 := \dot{b}$ represent the position and velocity of the object respectively. This system is of particular interest since it includes our model for the ball falling under the influence of gravity. The associated observer now takes the form

$$\begin{aligned} \dot{\hat{w}}_1 &= \hat{w}_2 - \Gamma_1 \mathbf{C}^T(\hat{w}_1)(\hat{v} - v) \\ \dot{\hat{w}}_2 &= A_1 \hat{w}_1 + A_2 \hat{w}_2 + u - \Gamma_2 \mathbf{C}^T(\hat{w}_1)(\hat{v} - v) \\ \hat{v} &= c(\hat{w}_1) \end{aligned} \quad (13)$$

with gain matrices Γ_1 and Γ_2 free to be chosen. Taking differences yields the error dynamics

$$\begin{aligned} \dot{\tilde{w}}_1 &= \tilde{w}_2 - \Gamma_1 \mathbf{C}^T(\hat{w}_1)(\hat{v} - v) \\ \dot{\tilde{w}}_2 &= A_1 \tilde{w}_1 + A_2 \tilde{w}_2 - \Gamma_2 \mathbf{C}^T(\hat{w}_1)(\hat{v} - v) \end{aligned} \quad (14)$$

which simplifies to

$$\begin{aligned} \dot{\tilde{w}} &:= \begin{bmatrix} \dot{\tilde{w}}_1 \\ \dot{\tilde{w}}_2 \end{bmatrix} \\ &= \begin{bmatrix} -\Gamma_1 \mathbf{C}^T(\hat{w}_1) \mathbf{\Lambda}(\hat{w}_1, w_1) \mathbf{C}(\hat{w}_1) & I \\ A_1 - \Gamma_2 \mathbf{C}^T(\hat{w}_1) \mathbf{\Lambda}(\hat{w}_1, w_1) \mathbf{C}(\hat{w}_1) & A_2 \end{bmatrix} \begin{bmatrix} \tilde{w}_1 \\ \tilde{w}_2 \end{bmatrix}. \end{aligned} \quad (15)$$

1) *Stability Proof:* We now demonstrate that this nonlinear dynamical system can be stabilized by the appropriate choice of gains under a set of reasonable assumptions about the operating conditions. Specifically, we will show that if b and \dot{b} both start beyond some minimum distance from the two cameras and if w and \hat{w} start within a nominal distance of one another then not only will \hat{w} converge to w but also b and \dot{b} are guaranteed to stay in front of both cameras. Note that this is an “almost” global result, in that the gains may be adjusted to stabilize an arbitrarily large set of initial errors, but that a bound on $\|\tilde{w}(0)\|$ is required.

Assumptions: We will assume that the motion to be tracked, $b(t)$, remains a bounded distance away from the two cameras, that is

$$\max_{t \geq 0} [\|b(t) - o_1\|, \|b(t) - o_2\|] < \beta_0$$

denoting by o_1, o_2 the origin (in the world coordinates) of the two camera frames, respectively. Using the notation

$$\Delta(b) := \begin{bmatrix} \Pi_3 \circ h_1(b) I_2 & 0 \\ 0 & \Pi_3 \circ h_2(b) I_2 \end{bmatrix}$$

developed above, this implies the matrix inequality

$$\text{for all } t: \Delta \circ b(t) < \beta_0 I_4.$$

Similarly, we will also assume that the motion to be tracked remains in front of both cameras—that is, $\Delta \circ b$ is bounded below

$$\text{for all } t: \Delta \circ b(t) > 2\delta_0 I_4.$$

This last assumption implies as well that

$$\min_{t \geq 0} [\|b(t) - o_1\|, \|b(t) - o_2\|] > 2\delta_0 > 0.$$

Furthermore, for simplicity, we presume that the structure of the gain matrices is of the form⁵

$$\Gamma_1 := \frac{\gamma_2}{\gamma_1} I, \quad \Gamma_2 := \gamma_2 I.$$

Bounded Initial State Estimation Error: In order to choose feedback gains, γ_1, γ_2 , and a quadratic Lyapunov function, ϑ (19), that depends upon them, we require some initial knowledge of the state estimation error, \tilde{w} . As usual, the more conservative the initial estimation error bounds, the larger will be the required gain magnitudes.

We now further assume that the initial state estimation error, \tilde{w}_0 , is bounded by some known amount,

$$\vartheta(\tilde{w}_0) \leq \vartheta_0 \quad (16)$$

and we choose γ_1 large enough so that

$$\gamma_1^2 > \vartheta_0 / 2\delta_0^2. \quad (17)$$

Note that the Lyapunov candidate below, ϑ (19), satisfies the inequality

$$\|\tilde{b}\|^2 \leq \frac{1}{2\gamma_1^2} \vartheta(\tilde{w}).$$

If $\vartheta \circ \tilde{w}(t) \leq \vartheta_0$ for all $t > 0$, then

$$\|\tilde{b}\|^2 < \delta_0^2,$$

from which we conclude that $\hat{b}(t)$ also remains in front of both cameras.

This last condition may again be rendered as

$$\text{for all } t: \Delta \circ \hat{b}(t) > \delta_0 I_4.$$

Consequence—A Lower Bound for $C^T \Lambda C$: We have already noted that $\beta_0 I_4 > \Delta(b) > \delta_0 I_4 > 0$ and $\frac{d\vartheta}{dt} < 0$ implies $\Delta(\hat{b}) > \delta_0 I_4$. This, in turn, yields

$$\Delta(\hat{b}) \cdot \Delta^{-1}(b) > \frac{\delta_0 I_4}{\beta_0} > 0.$$

Moreover, since \hat{b} remains in front of the cameras, and within a δ_0^2 radius of b and the latter remains within a compact set, it follows that the Jacobian matrix is bounded below for all time as well,

$$\text{for all } t: \|C(\hat{b})\| \geq \xi > 0.$$

Thus, under the assumptions above, we have

$$\|C^T(\hat{b})\Delta(\hat{b})\Delta^{-1}(b)C(\hat{b})\| \geq \mu := \xi^2 \frac{\delta_0}{\beta_0} > 0. \quad (18)$$

⁵In principle there is significantly more freedom available in the choice of gains than we have chosen to exploit here. This simple gain structure affords simple analysis and has performed well in our experimental and simulation environments.

The Lyapunov Function and Its Derivative: Consider a Lyapunov function of the form

$$\begin{aligned} \vartheta &:= \tilde{w}^T P \tilde{w} \\ P &:= \begin{bmatrix} 2\gamma_1^2 I & -\gamma_1 I \\ -\gamma_1 I & I \end{bmatrix} = \begin{bmatrix} \gamma_1 I & \gamma_1 I \\ 0 & -I \end{bmatrix} \begin{bmatrix} \gamma_1 I & \gamma_1 I \\ 0 & -I \end{bmatrix}^T \end{aligned} \quad (19)$$

which is positive definite since P is the symmetric square of a nonsingular matrix.

The derivative of ϑ along the motion of (15) is given by

$$\begin{aligned} \dot{\vartheta} &= -\tilde{w}^T Q \tilde{w} \\ Q &:= \begin{bmatrix} \gamma_1(\gamma_2 C^T \Lambda C - A_1) & -\gamma_1(\gamma_1 I + \frac{1}{2}A_2) + \frac{1}{2}A_1^T \\ -\gamma_1(\gamma_1 I + \frac{1}{2}A_2) + \frac{1}{2}A_1^T & \gamma_1 I - A_2 \end{bmatrix}. \end{aligned} \quad (20)$$

Choice of Feedback Gains: Choosing

$$\gamma_2 > \gamma'_2 + \|A_1\|/\mu \quad (21)$$

and

$$\gamma_1 > \gamma'_1 + \|A_2\| \quad (22)$$

for some $\gamma'_1, \gamma'_2 > 0$, insures that the diagonal block matrices of Q (20) are positive definite. We may now write

$$\tilde{w}^T Q \tilde{w} > [\|\tilde{b}\|, \|\dot{\tilde{b}}\|] Q' \begin{bmatrix} \|\tilde{b}\| \\ \|\dot{\tilde{b}}\| \end{bmatrix}$$

where

$$Q' := \begin{bmatrix} \gamma'_1 \gamma'_2 & \gamma'_0 \\ \gamma'_0 & \gamma'_1 \end{bmatrix}$$

and

$$\gamma'_0 = -\frac{1}{2} \left[\gamma_1 \left(\gamma_1 + \frac{1}{2} \|A_2\| \right) + \frac{1}{2} \|A_1\| \right].$$

Thus if we require

$$\gamma'_2 > \left(\frac{\gamma'_0}{\gamma'_1} \right)^2 \quad (23)$$

it follows that Q' is positive definite so $\dot{\vartheta}$ remains negative semidefinite.

C. Implementation

As usual, the real world departs from the assumptions underlying these models in certain important regards. What follows is a brief discussion of the differences between the previous section and the actual system, along with both experimental and simulation results demonstrating the utility and pitfalls for this type of observer.

1) *Time Sampled Implementation for Second-Order Systems:* Real cameras do not deliver continuous measurements in time—present day affordable devices take snapshots of the world at a fixed sampling rate, in our case, 60 Hz. Since the observed system's motion is significant relative to this rate (near impact, the ball often travels in excess of 10 cm between successive images), sampling considerations cannot be ignored. For the observer of Section III-A1 (with algebraic triangulation), implementation in the presence of sampling

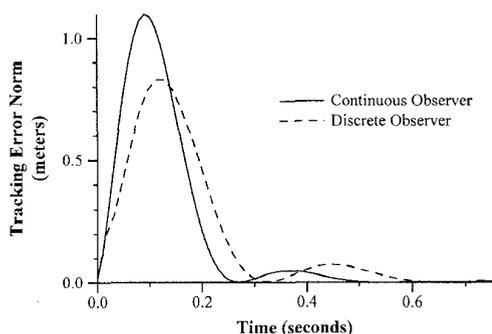


Fig. 7. Simulation: Convergence of the constant gain continuous and discrete time observers for small initial error.

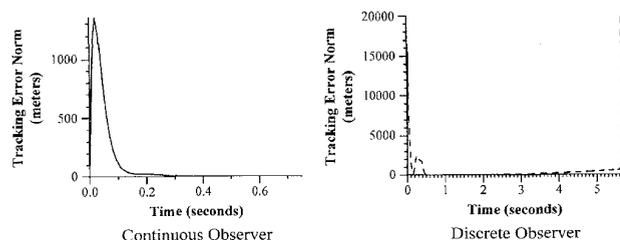


Fig. 8. Simulation: Convergence of the constant gain continuous and discrete time observers for larger initial error.

presents no problem since the dynamical system we are observing is linear, and traditional discrete time systems theory affords a reliable observer (6).

In contrast, since there is no general nonlinear sampling theory, the new triangulating estimator requires a separate stability argument for its discrete time version, which we present in Appendix A. Not surprisingly, one can no longer tolerate arbitrarily large initial errors, and furthermore the gain structure introduced in Appendix A becomes state dependent. In our laboratory implementation we have chosen, for reasons of computational simplicity, to implement a sampled version of the constant gain continuous time algorithm (13) as opposed to the conservative time varying structure of Appendix A. To understand the trade-offs in doing so, we have studied numerically both the continuous and sampled (constant gain discrete time) versions of (13). Figs. 7 and 8 demonstrate how an increase in the initial errors can result in instability for the time sampled system, while the continuous version remains well behaved, as predicted by theory. Fig. 7 depicts a case where the discrete (sampled at 60 Hz) and continuous constant gain system demonstrate comparable behavior for identical gains and small initial errors, Fig. 8 demonstrates that these systems can display markedly different behavior for different initial conditions. In this particular example the continuous system converges reasonably quickly, while the discrete version initially behaves reasonably well, then slowly begins to fail until 5.5 s, when it “explodes.”

2) *Integration in the Juggling System:* In our laboratory setting, we are fortunate to have good control over the initial conditions of the estimator—balls are presented at predetermined locations in the workspace with zero velocity

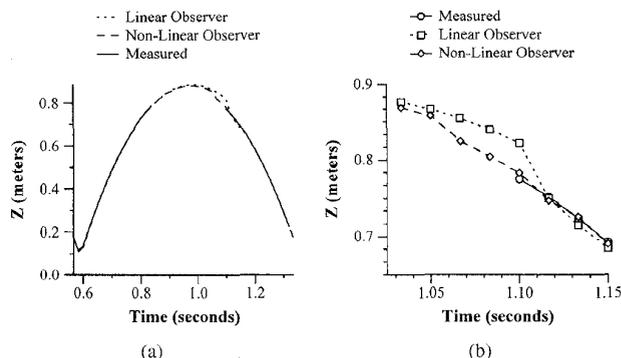


Fig. 9. Experimental data: Triangulated ball height and estimated ball height from both observers during recovery from a typical out-of-frame event.

and the sensing system performs algebraic triangulation on the first valid measurement it receives to generate an initial location estimate for the ball, while initializing the velocity estimate to zero. The resulting initial errors have been sufficiently small that we have neither observed destabilization of the observer due to the sampling effects noted above, nor have we had difficulty choosing observer gains based on the simpler design of (13), as opposed to the provably stable design of Appendix A.

Fig. 9 demonstrates the difference between this new observation scheme and the traditional triangulator/linear-observer scheme. Fig. 9(a) shows the overall flight of the ball as estimated by both observers, and measured by the triangulator (absence of the solid line indicates that the ball was outside the field of view of one or both cameras). In this example the ball travels out of frame for approximately 0.2 s. As can be seen in Fig. 9(b) (a blowup of the ball returning into the field of view) the dynamical triangulator is capable of updating its estimate while the triangulator/observer pair are forced to simply predict the trajectory (note the differing behavior from 1.05 to 1.10 s). Significant reduction in tracking error then results as the ball reappears in both camera’s fields of view at 1.10 s. This anecdotal picture is confirmed by experimental statistics. Fig. 10 shows the mean and standard deviation of the norm squared tracking errors (position only) for the first four frames after recovery from an out-of-frame event for 102 typical events. Although the improvement may seem slight, our machine generally has fewer than one dozen frames of data between reacquisition and impact, thus the improved transient greatly increases the accuracy of our batting.

V. ACTIVE VISION: CONTROLLING THE STATE OF ATTENTION

Whether choosing what segment of an image to process, where to look with a camera, or what camera to look with, many modern vision based systems incorporate an implicit control system—the control of the state of attention of the machine. In the case of our robot juggler this problem appears quite explicitly as a result of the limited real-time vision hardware. The machine is limited to only processing a small fraction of the total available data and must thus choose what data to process. This problem of *active vision* introduces a

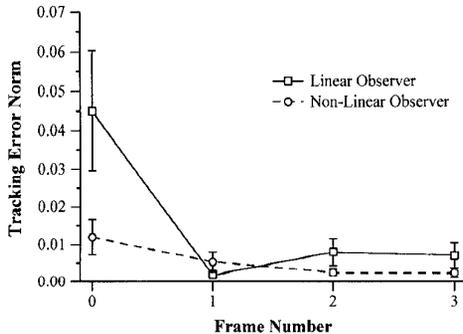


Fig. 10. Experimental data: Mean and standard deviation for the spatial observer errors immediately after recovery from out of frame, averaged over 102 events.

novel aspect of control in the design of the sensing system: the system responsible for control of attention must balance the benefit of examining only a small amount of sensor data against the risk of failing to generate useful measurements.

This problem might be posed within an optimal framework—for example by introducing the notion of “dual control” [5], [18], [21]. However, the experimental setup in our laboratory has never motivated the kind of careful statistical models required to yield practical utility from such an approach. What follows is a detailed examination of this problem under the presumption that there are three noise sources (sensor noise, modeling inaccuracy, and measurement inaccuracy due to the “area of attention”) to be balanced against the need for the state estimate to converge. Our presentation is focused on ensuring the convergence of the underlying state estimator in conjunction with guaranteeing future acquisition of measurements.

While in general the ideas presented here can be applied to an arbitrary estimation scheme, in particular the nonlinear scheme of Section IV, for reasons of clarity we will limit our formal analysis to the case of a linear estimation scheme, and point out how to make the generalization in the textual remarks.

A. The Sensor Control Variables as a “State of Attention”

The *sensor control* module of Fig. 2 is responsible for controlling the locus and extent of the image plane windows used for information extraction by the *data processing* subsystem. Thus, we offer as a definition for the system’s *state of attention* at some field interval, k , as the pair

$$a_k = (\hat{b}_k, \rho_k) \in \mathbb{R}^3 \times \mathbb{R}^+ \quad (24)$$

where \hat{b}_k denotes an estimate of where the falling ball is expected to appear, and the positive scalar ρ_k is a measure of certainty of this estimate. With respect to a norm, $\|\cdot\|_P$, which will be defined below, a_k induces two windows on the two camera image planes including all stereo image pixel pairs, x , in the set

$$\mathcal{N}(a_k) := \{x \in \mathfrak{c}(\mathbb{R}^3) : \|\hat{b}_k - \mathfrak{c}^\dagger(x)\|_P \leq \rho_k\}.$$

Note that this *state of attention* represents a dynamic notion of certainty and is thus structurally different from the geometric

notions of certainty offered in [3], [20]. The *data processing* subsystem will process these windows, and if the “ballness” tests (zeroth and second order moment tests) are passed, the first order moments will be passed to the *state estimation* to be interpreted as a spatial position. Otherwise, an empty window will be reported. For the sake of notational simplicity, we will denote the situation that first order moments are successfully formed inside the windows of the k th camera field as

$$\mathfrak{c}(b_k) \in \mathcal{N}(a_{k-1}).$$

The dependence of the k th measurement on a_{k-1} immediately suggests the *dynamics* intrinsic to the general sensor management problem—appearing here as mere delay. Regardless of how it is computed, the state of attention, a_k must be assembled from information derived from existing sensory observations. Thus, the acquisition of new data is necessarily mediated by old knowledge and a feedback loop is formed.

For a suitable norm, we look back to the stabilized observer equations (6). Because these systems are asymptotically stable there exists a positive definite symmetric matrix, P , such that

$$[A_{\tau_f} + GC]^T P [A_{\tau_f} + GC] < P$$

and we will denote the Euclidean norms induced by this matrix as

$$\|x\|_P := (x^T P x)^{1/2}; \quad \|A\|_P := \sup_{\|x\|_P=1} \|Ax\|_P.$$

For ease of exposition we introduce the notational conventions,

$$\alpha := \|A_{\tau_f}\|_P; \quad \bar{\alpha} := \|A_{\tau_f} + GC\|_P \quad (25)$$

and assume, purely for further notational convenience, that the poles of the closed loop observer equation (6) have been placed on the real line with multiplicity two with the consequence that

$$\|[A_{\tau_f} + GC]^{-1}\|_P = 1/\bar{\alpha}.$$

Similar arguments follow for the nonlinear and discrete time versions of this system, the only significant difference being that the norm of (25) are no longer constant, and must be replaced with conservative bounds on their state dependent values.

B. Observer Errors from a Noisy Model

The task at hand is to develop a control scheme for updating the state of attention, a_k as a function of its previous value and presently available data. To do so we must append to our previous state estimation procedure some notion of its changing degree of certainty. Thus, reconsider the Newtonian flight model (1), with the addition of both a process and a sensor noise model. We wish to model the inaccuracies in the Newtonian flight law as well as the salient features of the inaccuracies in ball position measurement introduced through the use of the camera. The latter includes two central phenomena: the absence of data when the ball lies outside of its assigned window; and the imprecision of spatial localization as the size of the window grows (and either delay grows or resolution shrinks correspondingly). For present exploratory purposes, we

will be content with a crude deterministic representation of the imprecision inherent in these process and sensor models.

We substitute for (1) and (5) the system

$$\begin{aligned} w[(j+1)\tau_r] &= F^{\tau_r}(w(j\tau_r)) + n_N(j\tau_r) \\ p_{k+1} &= w[k\tau_f] \\ \bar{b}_k &= \hat{C}_k[p_k + n_S(\rho_{k-1})]. \end{aligned} \quad (26)$$

As a first crude model for the failings of the putative Newtonian free-flight model (1) we take n_N to be a bounded deterministic sequence of uncontrolled inputs (perhaps generated via a map on the state space), and n_S to be the sensor noise introduced by thresholding a finite resolution image before computation of the moments. Because the window resolution must decrease as the window size increases (as a consequence of subsampling), n_S is nondecreasing in its argument. Since no subsampling is required for sufficiently small windows, n_S is a positive constant for small values of its argument. These considerations suggest an affine model of sensor noise as a function of window radius

$$\|n_S(\rho_k)\|_P \leq \nu_0 + \nu_1 \rho_k. \quad (27)$$

We choose to ignore the details of how $\mathbf{c}(\cdot)$ and $\mathbf{c}^\dagger(\cdot)$ influence the creation of errors in the measurement of \bar{b}_k , since this would require a careful assessment of the reflectance properties of the balls—a distant second order effect given the current structured lighting. In contrast, we are greatly concerned with developing correct window management logic, and we will explicitly embed the influence of $W(\cdot)$ in \hat{C} as follows.

The deterministic output map, \hat{C}_k returns the value $C = [I, 0]$ as in (1) when the body's image is in the examined area of the image plane, and vanishes otherwise

$$\hat{C}_k := \begin{cases} C: \mathbf{c}(b_k) \in \mathcal{N}(a_{k-1}) \\ 0: \mathbf{c}(b_k) \notin \mathcal{N}(a_{k-1}) \end{cases}. \quad (28)$$

This models the salient behavioral features of the *data processing* subsystem introduced in Section III-A1, as it returns no data (zero) when an “out of frame” event occurs. This results in the observer simply extrapolating the present state estimate in such situations. The resulting observer takes the same form as (6) only with \hat{C}_k from (28) incorporated,

$$\begin{aligned} \hat{p}_{k+1} &= F^{\tau_f}(\hat{p}_k) + G(\bar{b}_k - \hat{C}_k \hat{p}_k) \\ \hat{w}(k\tau_f + j\tau_r) &= F^{\tau_f + \iota_k + j\tau_r}(\hat{p}_k), \\ j &= 0, 1, \dots, \tau_f + \iota_{k+1} - \iota_k \\ \hat{b}_k &= C F^{\tau_f}(\hat{p}_k). \end{aligned} \quad (29)$$

Here, we distinguish between the state estimate, $\hat{w}(\cdot)$, that is sent forward to the juggling algorithm, and the attention variable, \hat{b} , that will be sent back to the *sensor control* module. The robot gets $\hat{w}(k\tau_f)$ as soon as it is formed, with future predictions being made at the faster physical rate, τ_r . The *sensor control* module will make use of \hat{p}_k in the form of \hat{b}_k to handle the $(k+1)$ st image.

The result is a system with two distinct kinds of error,⁶ each with its own causes and effects. The first is the standard error due to the observer,

$$\tilde{p}_k := p_k - \hat{p}_k,$$

and is governed by the dynamics

$$\begin{aligned} \tilde{p}_{k+1} &= (A_{\tau_r} + G\hat{C}_k)\tilde{p}_k + n_k \\ n_k &:= G n_S(\rho_{k-1}) + n_N[(k-1)\tau_f]. \end{aligned} \quad (30)$$

Denoting the present error magnitude by $\vartheta_k := \|\tilde{p}_k\|_P$, we can conclude that

$$\begin{aligned} \vartheta_{k+1} &\leq \lambda_k \vartheta_k + \|n_k\|_P \\ \lambda_k &:= \begin{cases} \bar{\alpha} < 1: \mathbf{c}(b_k) \in \mathcal{N}(a_{k-1}) \\ \alpha > 1: \mathbf{c}(b_k) \notin \mathcal{N}(a_{k-1}) \end{cases}. \end{aligned} \quad (31)$$

(α and $\bar{\alpha}$ are defined in (25)) and it follows that the necessary and sufficient condition on \hat{C}_k and λ_k for a measurement to be successfully taken may now be expressed as

$$\begin{aligned} \mathbf{c}(b_k) \in \mathcal{N}(a_{k-1}) &\iff \\ \|C^T(Cw[(k-1)\tau_f] - \hat{b}_{k-1})\|_P &< \rho_{k-1}. \end{aligned} \quad (32)$$

Thus, there is a second sort of error associated with this event. It is due to the conjunction of process noise with time delay in the formation of the extrapolated state estimate. For, assuming $\|n_N\|_P$ is bounded above by the scalar ν_N , we have

$$\begin{aligned} \|C^T(Cw[(k-1)\tau_f] - \hat{b}_{k-1})\|_P &\leq \|w[(k-1)\tau_f] - F^{\tau_f}(\hat{p}_{k-1})\|_P \\ &\leq \alpha(\vartheta_{k-1} + \tau_f \nu_N). \end{aligned} \quad (33)$$

It follows that if ρ_{k-1} is at least as large as the last expression, we are guaranteed (within the limits of our noise model) that the k th window will not be empty—that condition (32) will hold.

C. The Window Radius Control Policy and Its Consequences

The construction of a functional observer of the form presented in (26) necessitates the implementation of a *sensor controller*. Specifically, this amounts to choosing window sizes, ρ_k , and locations, \hat{b}_k , in such a fashion that the acquisition of new measurements can be guaranteed in conjunction with the estimated state converging to the actual state.

1) *Certainty Estimates from a Parallel Observer*: The result of (33) implies that ρ_k should be set in relation to ϑ_k in order to insure data to the observer. But, unfortunately, we are not in possession of the error magnitude, ϑ , for the very reason that we were led to build an observer in the first place (our inability to measure ball velocities). Since \hat{p} represents our only knowledge of p , the best estimate of ϑ is 0 as matters stand presently. To address this deficit, we will build a second

⁶Note that there is actually a third sort of error, which concerns the quality of the estimate passed forward to the robot. If $\tilde{w}_k := w(k\tau_f + \iota_k) - \hat{w}(k\tau_f)$ we have, $\|\tilde{w}_k\|_P \leq \alpha^{\iota_k}(\vartheta_k + (\tau_f + \iota_k)\nu_N)$, where $\tau_r \leq \iota_k \leq \tau_f$. Thus, $\|\tilde{w}_k\|_P$ is a nondecreasing function of both ϑ and ρ . But this error is never seen in the sensory loop.

state estimator and attempt to construct and estimate of ϑ by comparing the two.

In the case of linear estimation scheme, we can make use of the invertibility of the observability matrix,

$$\Theta := \begin{bmatrix} C \\ CA_{\tau_f} \end{bmatrix}$$

to define a very different estimate of p taking the form

$$d_k = F^{\tau_f} \left(\Theta^{-1} \left(\begin{bmatrix} \bar{b}_{k-1} \\ \bar{b}_k \end{bmatrix} - \begin{bmatrix} 0 \\ CA_{\tau_f} \end{bmatrix} \right) \right). \quad (34)$$

This is a dead-beat observer in the sense that $\tilde{d}_k := p_k - d_k$ converges to zero in two steps from all initial estimates, d_0 in the absence of noise, $n_S = n_N = 0$.⁷

Through careful comparison of the estimates provided by these two observers (as detailed in Appendix C.1) we are led to define a worst case estimate for ϑ as

$$\hat{\vartheta}_{k-1} := [\|\hat{p}_k - d_k\|_P + \nu_{\Delta}(\rho_{k-1}, \rho_{k-2})] / \bar{\alpha} \quad (35)$$

where

$$\begin{aligned} \nu_{\Delta}(\rho_{k-1}, \rho_{k-2}) &:= \|n_{k-1}\|_P + \alpha \tau_f \nu_N \\ &+ \frac{\alpha}{\|\Theta\|_P} (\nu_N + \|n_S(\rho_{k-1})\|_P + \|n_S(\rho_{k-2})\|_P). \end{aligned}$$

This guarantees that $\hat{\vartheta}_{k-1} \geq \vartheta_{k-1}$ when \bar{b}_k and \bar{b}_{k-1} are both available. If either of these is unavailable (making d_k uncomputable), a conservative estimate for ϑ is given by

$$\hat{\vartheta}_{k-1} := \alpha \hat{\vartheta}_{k-2}. \quad (36)$$

2) *Control of Window Radius*: Equipped with a worst case estimate for ϑ , we are now in a position to adjust ρ . According to the previous calculations (33), a window radius management strategy that achieves the relation

$$\rho_k \geq \alpha(\vartheta_k + \tau_f \nu_N)$$

guarantees data to the observer at step $k+1$. Noting that ϑ_k is causally determined by ρ_k , and thus cannot be estimated directly by the procedure (35) at stage k , we appeal to (31) and note that the desired relation is implied by

$$\rho_k \geq \alpha(\lambda_{k-1} \vartheta_{k-1} + \|n_{k-1}\|_P + \tau_f \nu_N).$$

This demonstrates that the radius adjustment procedure

$$\rho_k = \alpha(\lambda_{k-1} \hat{\vartheta}_{k-1} + \|n_{k-1}\|_P + \tau_f \nu_N) \quad (37)$$

will always yield a window large enough to capture the next centroid, up to the limits of the error models employed.

⁷More generally any six independent measurements of the underlying dynamical system could be used to produce a similar dead-beat estimate of the balls state (e.g., any three perspective images of a ball's flight could be used).

3) *Boundedness of Estimator Errors*: This then leaves the question of observer convergence. Recall that as ρ increases, the quality of the estimates deteriorates. Eventually, the recourse to subsampling might begin to have a net destabilizing effect through the injection of noise represented by n_k in (31). We must show that the coupled dynamical system defined by (31) and (37) remains bounded. The following analysis verifies that if the observer gains (G in (29)) are chosen to result in sufficiently fast convergence, and the noise due to sub-sampling (ν_1 in (27)) grows slowly enough in relation to window size, then our radius management scheme will grow windows quickly enough to reacquire lost balls, but not so quickly as to drown the estimation process in resolution noise.

As derived in Appendix B the coupled dynamics for ϑ_k and ρ_k may be bounded by

$$\begin{aligned} \vartheta_{k+1} &\leq \lambda_k \vartheta_k + \nu_1 \rho_{k-1} + \gamma \nu_0 + \nu_N \\ \rho_{k+1} &\leq \alpha \left(\tau_f \nu_N + \gamma(\nu_0 + \nu_1 \rho_{k-1}) + \nu_N \right. \\ &\quad \left. + \frac{\lambda_k^2}{\bar{\alpha}} [\vartheta_k + 2\nu_{\Delta}(\rho_k, \rho_{k-1})] \right). \end{aligned}$$

Moving to the coordinate system, $x := [\chi_1, \chi_2, \chi_3]^T$, where $\chi_1(k) \geq \vartheta_k$ bounds the actual Lyapunov magnitude of (29) and $\chi_2(k) \geq \rho_k$, $\chi_3(k) \geq \rho_{k-1}$ represent bounds on the most recent window radius values, we obtain the dynamics

$$\begin{aligned} x(k+1) &= Q_k x(k) + r \\ Q_k &:= \begin{bmatrix} \lambda_k & 0 & \nu_1 \\ \frac{\alpha \lambda_k}{\bar{\alpha}} & \alpha \nu_1 g_1 & \alpha \nu_1 g_2 \\ 0 & 1 & 0 \end{bmatrix} \\ r &:= \begin{bmatrix} r_1 \\ r_2 \\ 0 \end{bmatrix} \end{aligned} \quad (38)$$

where the symbols $g_i, r_i, i = 1, 2$ denote constants derived from the computations developed above.

By construction of the radius adjustment procedure (37), the state of this system must enter a region where $\lambda_k = \bar{\alpha} < 1$ after an initial transient. Now, elementary root locus analysis of the characteristic polynomial of this system,

$$s^2(-\bar{\alpha} + s) + \alpha \nu_1 [(g_2 - 1)\bar{\alpha} + (\bar{\alpha} g_1 - g_2)s + g_1 s^2]$$

shows that the matrix Q has roots in the unit circle of the complex plane for small enough values of ν_1 (they originate at $\{\bar{\alpha}, 0, 0\}$). This implies that if the noise coefficient, ν_1 is sufficiently small relative to the other parameters then the window management system succeeds in keeping the windows large enough to retain the required image, but not so large as to destabilize the estimation procedure.

D. Implementation and Results

We have implemented and performed experiments on a window management system identical to that described here. Specifically, while our juggling system has made use of the output from the nonlinear estimation system of Section IV, we have continued to use a triangulator and linear estimation

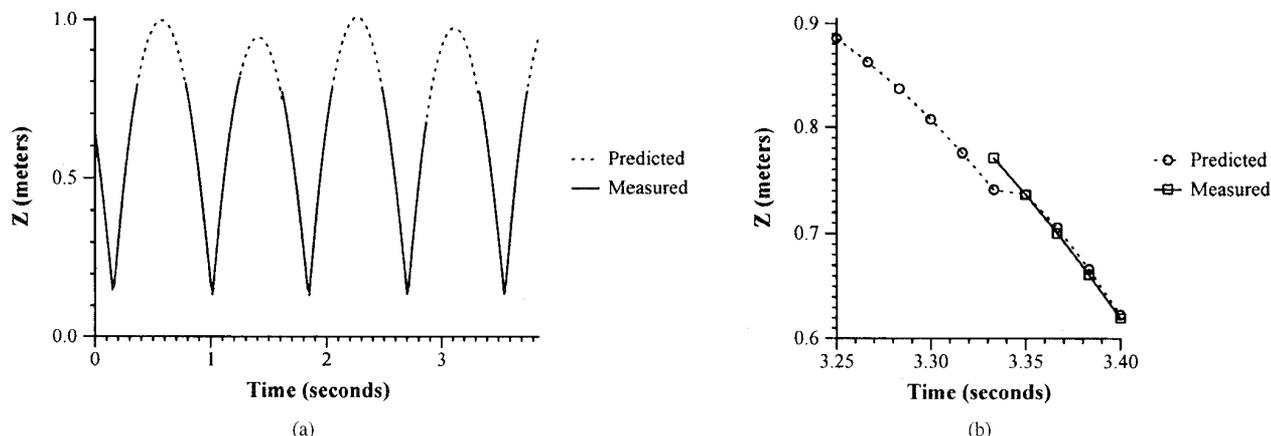


Fig. 11. (a) Measured and predicted (by the observer) ball heights for an out of frame juggling sequence, and (b) an expanded view of a single recovery event.

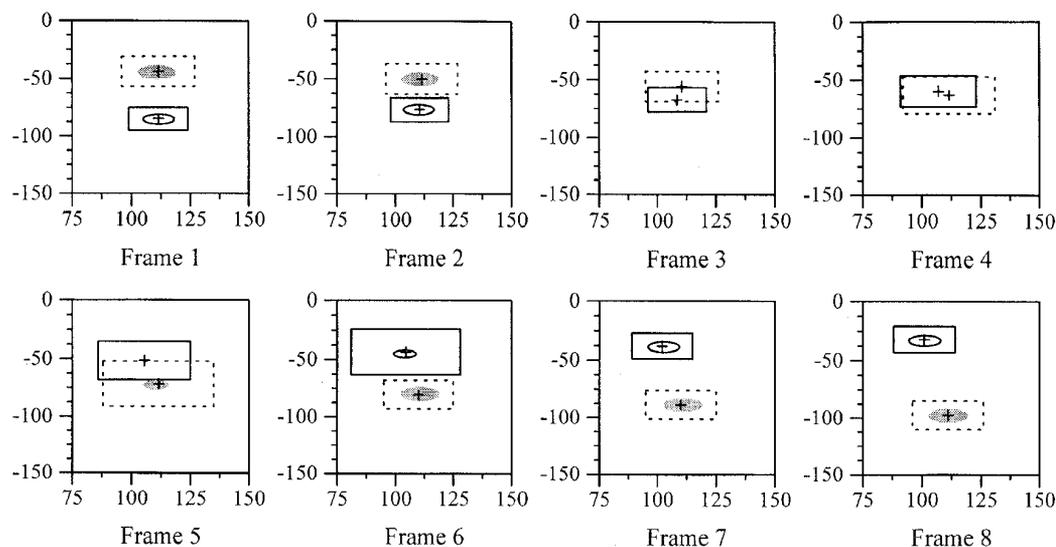


Fig. 12. Image plane tracks and search windows for two balls during an occlusion event.

scheme to drive the sensor management system. Thus our implementation incorporates three parallel observers, rather than the minimally required two.⁸

The two significant areas of improvement, for which we present anecdotal evidence, are the ability to recover a ball which leaves the system's field of view and to recover from two balls occluding one another.

Recovery from Out-of-Frame: As mentioned above, the use of the output from the *state estimator* to place the search windows, has allowed the juggling height to be increased to the point that every juggle passes out of the field of view of our vision system. Fig. 11(a) and (b) depict exactly such a sequence. The top 0.25 to 0.4 seconds of each flight are outside the field of view, as is evident by the lack of position measurements during this period. Nevertheless the observer continues to predict the ball's location, and it is recovered as it passes back into the system's field of view. Fig. 11(b) shows

⁸Note this is the minor deviation from the idealized system of Fig. 4 mentioned earlier.

a detail of a single recovery. Evidently, there is a slight build up of prediction error (approximately 5 cm vertical error) over the near 0.5 second that this ball was outside of view. However since the measurement window has grown, this magnitude of error is readily accommodated.

Recovery from Ball-Ball Occlusions: Similarly we have been able to observe the occlusion events discussed earlier. Fig. 12 depicts eight sequential image plane measurements spanning an occlusion event (for clarity we have only presented a small region from one of the two image planes). The grey ellipses represent centroid information assigned to ball 0, while the white ellipses are those associated with ball 1. The solid and dotted boxes denote the windows over which the binary moments were calculated for ball 0 and 1 respectively, and the small crosses indicate the window centers which are also the expected ball locations.

In this particular sequence ball 1 (the open ellipse) is rising toward its apex as ball 0 falls behind it causing an occlusion in the 3rd frame. The balls remain occluded (lying within

the overlap region between the two large windows), and both windows continue to grow until the 5th frame at which point ball 0 reappears from behind the search window for ball 1, and finally in frame 6 ball 0 becomes visible due to the search window for ball 1 shrinking and exposing it.

VI. CONCLUSION

This paper discusses a working active visual sensor—a computational dynamical system internal to our juggling robot’s perceptual apparatus that combines state estimation and sensor control in a nonlinear feedback loop. The architecture of this system has been described in detail, its functional behavior in laboratory experiments has been documented, and formal proofs of soundness have been presented. Throughout, we have tried to take an integrated view, stressing the implications of our design choices for the coordinated system as a whole. In particular, the analytical framework is motivated by the hope of characterizing and managing the behavior of dynamically dexterous robotic systems that incorporate such active perceptual subsystems.

We have defined the state of attention of our sensor to be a representation of a prescribed volume of space whose features the system will attempt to relate to a particular world model—a falling ball. We have introduced a control of attention policy by means of which this volume is adjusted with the dual effect of bringing high level clues to bear on the low level image plane segmentation task and providing a recovery strategy when segmentation fails. We have demonstrated mathematically that this recovery strategy is sound: that the volume of space being examined will grow as fast as required to encompass the missing ball (up to the limits of our model of ball motion) without growing so fast as to misjudge its state when found again (up to the limits of our model of sensor noise).

We have also proposed a new nonlinear state estimation scheme—the triangulating observer—that employs stereo image plane measurement errors to update estimates of the falling ball’s spatial position and velocity and both mathematically and experimentally demonstrated its convergence. This construction follows from a structural feature of the perspective camera transformation relating an image plane difference vector to the Jacobian projection of the corresponding Cartesian difference vector. Direct use of image plane measurements avoids passing the data through a sensitive algebraic inverse and seems to result in a system better conditioned with respect to noise sources. It is particularly important in our juggling application that the new estimator promotes an efficient use of incomplete stereo image data: updates can be formed from only one camera’s report when the other’s is missing or corrupted.

Our motivation for pursuing this line of research remains quite practical. During the construction of our juggling robot, we found that the original specification of its perceptual capabilities was inadequate to the task at hand. As more and more enhancement modules were added, anticipating and appropriately redesigning both their mutual interaction as well

as the net effect on the targeted juggling behavior became increasingly challenging. This motivated the analytical study presented here. We are convinced that this interactive process of pragmatic building followed by theoretical reflection leading to further refined building, and so on, is the best way to advance the emerging field of robotics.

APPENDIX A

DISCRETE TIME TRIANGULATING OBSERVER

A. The Nonlinear Observer “Inherits” the Gains of the Linear Observer

When considering the stability of the discrete time variant of (14) it is essential that we chose gains which are large enough to stabilize the system yet small enough to avoid instability due to the time sampled nature of the system. Our approach is to begin with the sampled linear model of the system, construct a stabilizing set of gains (using pole placement, or whatever design method seems appropriate), use these gains to induce a Lyapunov function over the nonlinear error system, and finally use the structure of this Lyapunov function to derive stabilizing gains for the actual system. Note, these derived gains stabilize the system for initial errors, $\|\tilde{w}(0)\|$, sufficiently small as to guarantee $\|A - I\|$ starts and remains sufficiently small.

B. The Lyapunov Difference as a Quadratic Form in Gain

Given that $A \in \mathbb{R}^{6 \times 6}$ is the discrete time system matrix of the Cartesian dynamics, $\tilde{b} = \tilde{a}$, with the (possibly state dependent) output matrix $E_1 = [M^T, 0]^T$. Furthermore presume that (E_1^T, A) is an observable pair, and that a matrix of gains, G has been chosen such that $A - GE_1^T$ is an asymptotically stable matrix. It then follows that for any $Q > 0$ there can be found a $P > 0$ such that

$$Q = P - [A^T - E_1 G^T] P [A - GE_1^T] > 0 \quad (39)$$

and it then follows that for all $w := (w_1, w_2) \in \mathbb{R}^6$

$$0 < w^T L w - 2w_1^T M^T G^T K w + w_1^T M^T G^T P G M w_1 \quad (40)$$

where

$$L := \begin{bmatrix} L_1 & L_0^T \\ L_0 & L_2 \end{bmatrix} = P - A^T P A$$

$$K := [K_1, K_2] = P A.$$

From this we conclude that L_2 must be positive definite so that this inequality is preserved when $w_1 = 0$. Rearranging terms and completing the square for L_2 we may write this as

$$0 < w^T S_2 L_2^{-1} S_2^T w + w_1^T S_1^T R S_1 w_1 \quad (41)$$

where

$$S_1 := \begin{bmatrix} G M \\ I \end{bmatrix}; \quad S_2 := \begin{bmatrix} L_0^T - M^T G^T K_2 \\ L_2 \end{bmatrix}$$

$$R := \begin{bmatrix} R_1 & R_0^T \\ R_0 & R_2 \end{bmatrix} \\ = \begin{bmatrix} P - K_2 L_2^{-1} K_2^T & K_1 + K_2 L_2^{-1} L_0 \\ (K_1 + K_2 L_2^{-1} L_0)^T & L_1 - L_0 L_2^{-1} L_0^T \end{bmatrix}$$

and note that $\text{Ker } S_2^T$ includes the image of

$$\begin{bmatrix} I \\ -L_2^{-1}(L_0 - M^T G^T K_2)^T \end{bmatrix}.$$

Thus it also must be true that $S_1^T R S_1 > 0$, or else the inequality of (39) would fail to hold for $w \in \text{Ker } S_2^T$.

C. Choice of Gains

For the error system associated with the nonlinear observer of Section IV the relevant problem is to find a (possibly state dependent) choice of gains which stabilize the discrete form of (14). In this case $M = C^T \Lambda C$. To begin we note that it must be the case that R_1 from (41) is negative definite. Were this not true an arbitrarily large gain, G , could be chosen such that it aligned with a positive eigenvector of R_1 forcing (40) to hold for all w in an arbitrarily large set, however this is a discrete system and it must be the case that the system is destabilized by arbitrarily large gains.

Since $R_1 < 0$ it follows that we may rewrite R as

$$\begin{bmatrix} R_1 \\ R_0 \end{bmatrix} R_1^{-1} \begin{bmatrix} R_1 & R_0^T \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & R_2 - R_0 R_1^{-1} R_0^T \end{bmatrix}.$$

We conclude that $S_1^T R S_1 > 0$ only if

$$R_2 - R_0 R_1^{-1} R_0^T > 0.$$

This suggests $-\gamma R_1^{-1} R_0 M^{-1}$ as the natural choice of gains. Unfortunately M^{-1} , because of its dependence on Λ , is not available to us, and we are forced to consider

$$G = -\gamma R_1^{-1} R_0 (C^T C)^{-1}.$$

Expanding the first term of $S_1^T R S_1$ in this case yields

$$(I - C^T \Lambda C (C^T C)^{-1}) R_0^T R_1^{-1} R_0 (I - (C^T C)^{-1} C^T \Lambda C)$$

and we note that this has the desired property of approaching 0 as $\Lambda \rightarrow I$. As detailed in Section IV-B, Λ is bounded above and below. Similar algebraic manipulation results in initial bounds on $\|\hat{w}(0)\|$ that guarantee $\|\Lambda - I\|$ starts and remains sufficiently small for all time.

APPENDIX B

MATHEMATICAL DETAILS OF WINDOW MANAGEMENT

A. Upper Bound for $\hat{\vartheta}_{k-1}$

In Section V-C1 a worst case estimate for $\hat{\vartheta}_{k-1}$ was developed, its derivation follows from the fact that

$$\tilde{d}_k = \sum_{j=1}^{\tau_f} A_{\tau_r}^{k-j} n_N(\tau_r j) - A_{\tau_f} \Theta^{-1} \\ \times \begin{bmatrix} \hat{C}_{k-1} n_S(k-1) \\ \hat{C}_k (n_S(k) + n_N(k-1)) \end{bmatrix}.$$

Noticing that

$$\|\hat{p}_k - d_k\|_M = \|\tilde{d}_k - \tilde{p}_k\|_M \\ = \left\| (A_{\tau_r} + G \hat{C}_k) \tilde{p}_{k-1} + n_{k-1} \right. \\ \left. + \sum_{j=1}^{\tau_f} A_{\tau_r}^{k-j} n_N(\tau_r j) - A_{\tau_f} \Theta^{-1} \right. \\ \left. \times \begin{bmatrix} \hat{C}_{k-1} n_S(k-1) \\ \hat{C}_k (n_S(k) + n_N(k-1)) \end{bmatrix} \right\|_M \\ \geq \frac{1}{\|(A_{\tau_r} + G C)^{-1}\|_M} \vartheta_{k-1} - \nu_\Delta(\rho_{k-1}, \rho_{k-2})$$

where

$$\nu_\Delta(\rho_{k-1}, \rho_{k-2}) := \|n_{k-1}\|_M + \alpha \tau_f \nu_N \\ + \frac{\alpha}{\|\Theta\|_M} (\nu_N + \|n_S(\rho_{k-1})\|_M \\ + \|n_S(\rho_{k-2})\|_M)$$

we are led to define a worst case estimate for ϑ as

$$\hat{\vartheta}_{k-1} := [\|\hat{p}_k - d_k\|_M + \nu_\Delta(\rho_{k-1}, \rho_{k-2})] / \bar{\alpha}. \quad (42)$$

B. Bounded Coupled Dynamics for ρ_k and ϑ_k

The bounded coupled dynamics for ρ_k and ϑ_k used in Section V-C3 is constructed by first approximating the appearance of ρ in n_k and ν_Δ to first order (27). This results in

$$\|n_k\|_P \leq \gamma(\nu_0 + \nu_1 \rho_{k-1}) + \nu_N \\ \nu_\Delta(\rho_k, \rho_{k-1}) \leq (1 + \alpha \tau_f) \nu_N + \gamma(\nu_0 + \nu_1 \rho_k) \\ + \frac{\alpha}{\|\Theta\|_P} (\nu_N + 2\nu_0 + \nu_1 \rho_k + \nu_1 \rho_{k-1}) \\ = (1 + \alpha(\tau_f + 1/\|\Theta\|_P)) \nu_N + \left(\gamma + 2 \frac{\alpha}{\|\Theta\|_P} \right) \\ \times \nu_0 \nu_1 \left(\gamma + \frac{\alpha}{\|\Theta\|_P} \right) \rho_k + \nu_1 \frac{\alpha}{\|\Theta\|_P} \rho_{k-1}.$$

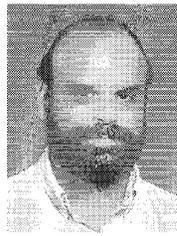
The coupled dynamical inequalities in question now may be written

$$\vartheta_{k+1} \leq \lambda_k \vartheta_k + \nu_1 \rho_{k-1} + \gamma \nu_0 + \nu_N \\ \rho_{k+1} \leq \alpha \left(\tau_f \nu_N + \gamma(\nu_0 + \nu_1 \rho_{k-1}) + \nu_N \right. \\ \left. + \frac{\lambda_k^2}{\alpha} [\vartheta_k + 2\nu_\Delta(\rho_k, \rho_{k-1})] \right).$$

REFERENCES

- [1] P. K. Allen, A. Timcenko, B. Yoshimi, and P. Michelman, "Trajectory filtering and prediction for automated tracking and grasping of a moving object," in *IEEE Int. Conf. on Robotics and Automation*, 1992, pp. 1850-1856.
- [2] ———, "Automated tracking and grasping of a moving object with a robotic hand-eye system," in *IEEE Int. Conf. on Robotics and Automation Workshop on Visual Servoing*, 1994.

- [3] P. Anandan, "Measuring visual motion from image sequences," Tech. Rep. COINS-TR-87-21, University of Massachusetts, Amherst, 1987.
- [4] R. L. Andersson, *A Robot Ping-Pong Player: Experiment in Real-Time Intelligent Control*. Cambridge, MA: MIT Press, 1988.
- [5] Y. Bar-Shalom, Ed., *Multitarget-Multisensor Tracking: Advanced Applications*. Norwood, MA: Artech House, 1990.
- [6] D. J. Bennett, J. M. Hollerbach, and D. Geiger, "Autonomous robot calibration for hand-eye coordination," in *Int. Symp. of Robotics Res.*, 1989.
- [7] M. Bühler, D. E. Koditschek, and P. J. Kindlmann, "A family of robot control strategies for intermittent dynamical environments," *IEEE Control Syst. Mag.*, vol. 10, pp. 16–22, Feb. 1990.
- [8] P. I. Corke, "Video-rate robot visual servoing," in *Visual Servoing—Real-Time Control of Robot Manipulators Based on Visual Sensory Feedback*, K. Hashimoto, Ed. Singapore: World Scientific, 1993.
- [9] E. D. Dickmanns and V. Graefe, "Applications of dynamic monocular machine vision," *Mach. Vis. Applicat.*, pp. 241–261, 1988.
- [10] ———, "Dynamic monocular machine vision," *Mach. Vis. Applicat.*, pp. 223–240, 1988.
- [11] A. Fox and S. Hutchinson, "Exploiting visual constraints in the synthesis of uncertainty-tolerant motion plans i: The directional backprojection," in *IEEE Int. Conf. on Robotics and Automation*, May 1993, vol. 1, pp. 305–310.
- [12] ———, "Exploiting visual constraints in the synthesis of uncertainty-tolerant motion plans ii: The non directional backprojection," in *IEEE Int. Conf. on Robotics and Automation*, May 1993, vol. 1, pp. 311–316.
- [13] B. K. Ghosh, M. Jankovic, and Y. T. Wu, "Perspective problems in systems theory and its application to machine vision," *J. Mathemat. Syst. Estimation Contr.*, vol. 4, no. 1, pp. 3–38, 1994.
- [14] G. D. Hager, W.-C. Chang, and A. S. Morse, "Robot feedback control based on stereo vision," *IEEE Control Systems*, vol. 15, pp. 30–39, Feb. 1995.
- [15] B. K. P. Horn, *Robot Vision*. Cambridge MA: MIT Press, 1986.
- [16] M. Jankovic and B. K. Ghosh, "Visually guided ranging from observations of points, lines and curves via an identifier based nonlinear observer," *Syst. Contr. Lett.*, to appear.
- [17] D. Kim, A. A. Rizzi, G. D. Hager, and D. E. Koditschek, "A robust convergent visual servoing system," in *RSJ/IEEE Int. Conf. on Intelligent Robotics and Systems*, 1995.
- [18] P. R. Kumar and P. Varaiya. *Stochastic Systems: Estimation, Identification, and Adaptive Control*. Englewood Cliffs, NJ: Prentice-Hall, 1986.
- [19] B. Nelson, N. P. Papanikolopoulos, and P. K. Khosla, "Visual servoing for robotic assembly," in *Visual Servoing—Real-Time Control of Robot Manipulators Based on Visual Sensory Feedback*, K. Hashimoto, Ed. Singapore: World Scientific, 1993.
- [20] N. P. Papanikolopoulos and P. K. Khosla, "Adaptive robotic visual tracking: Theory and experiments," *IEEE Trans. Automat. Contr.*, vol. 38, no. 3, pp. 429–445, 1993.
- [21] B. S. Y. Rao, H. F. Durrant-Whyte, and J. A. Sheen, "A fully decentralized multi-sensor system for tracking and surveillance," *Int. J. Robot. Res.*, vol. 12, no. 1, Feb. 1993.
- [22] A. A. Rizzi and D. E. Koditschek, "Further progress in robot juggling: The spatial two-juggle," in *IEEE Int. Conf. on Robotics and Automation*, vol. 3, pp. 919–924, May 1993.
- [23] ———, "Toward the control of attention in a dynamically dexterous robot," in *Visual Servoing—Real-Time Control of Robot Manipulators Based on Visual Sensory Feedback*, K. Hashimoto, Ed. Singapore: World Scientific, 1993.
- [24] ———, "An active visual estimator for dexterous manipulation," Rep. CSE-TR 213-94, Univ. of Michigan, Ann Arbor, May 1994.
- [25] A. Rizzi and D. E. Koditschek, "Preliminary experiments in robot juggling," in *Proc. Int. Symp. on Experimental Robotics*, Toulouse, France, June 1991.
- [26] ———, "Progress in spatial robot juggling," in *IEEE Int. Conf. Robot. Aut.*, Nice, France, pp. 775–780, May 1992.
- [27] A. A. Rizzi, L. L. Whitcomb, and D. E. Koditschek, "Distributed real-time control of a spatial robot juggler," *IEEE Computer*, vol. 25, no. 5, May 1992.
- [28] J. L. Synge and B. A. Griffith, *Principles of Mechanics*. London: McGraw Hill, 1959.



Alfred A. Rizzi (M'89) received the Sc.B. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1986, and the M.S. and Ph.D. degrees in electrical engineering from Yale University, New Haven, CT, in 1990 and 1994, respectively. Prior to graduate school, he worked in the aerospace industry as a designer of electro-optical test equipment.

In 1994 and 1995, he held a post-doctoral position at the University of Michigan Center for Display Technology and Manufacturing, and since January 1996, has been a post-doctoral Fellow, Microdynamic Systems Lab, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA. His research interests include robotics, real-time distributed control of intelligent machines, and the application of dynamical systems theory to dexterous manipulation problems.



Daniel E. Koditschek (S'80–M'83) received the Ph.D. degree in electrical engineering from Yale University, New Haven, CT, in 1983.

He was appointed to the faculty at Yale University in 1984, and moved to the University of Michigan, Ann Arbor, in 1993, where he is presently a Professor, Department of Electrical Engineering and Computer Science with a joint appointment in the Artificial Intelligence Lab and Control Systems Lab. His research interests include robotics, the application of dynamical systems theory to intelligent mechanisms, nonlinear control, and applications of real time distributed control technology.

Dr. Koditschek is a member of the AAAS, ACM, AMS, MAA, SIAM, and Sigma Xi.