Visual Control of Robot Reaching

N. J. Ferrier*
Department of Mechanical Engineering
University of Wisconsin-Madison, WI, 53706
Email: ferrier@engr.wisc.edu

August 14, 1998

Abstract

In visual control of motion the performance of the system is typically tied to the speed at which the visual information can be processed. Visual processing is unusual in that the data is not the *information* and the *quality* of the information can often be improved through extended processing. A sensor tracking an object will only provide *sporadic observations* due to the processing delay, and, computational delay can pose difficulty within the control structure. We explore the trade off between the computational delay, the control stability, and the accuracy both of the visual processing and resulting visual-guided motion. The task of visually tracking a target in order to localize it accurately enough to "touch" it provides a demonstration of the trade off between motion accuracy and the motion execution time for visual guided robot reaching tasks.

1 Motivation

When visual information is used to control motor systems, such as robots, control inputs for actuators are derived from the processing of image data. Typically the visual processing involves the extraction of contours, features, corners, or other visual primitives. Then, motion commands are determined by considering the image location of visual primitives, calibration information, and the desired robot task (see e.g. [1, 5, 6, 13, 15, 19]). Tracking stability can only be achieved if the sensing delays are sufficiently small, and/or, if the dynamical model is sufficiently accurate. Various approaches have been taken to compensate for the delay of the visual system in visual servo control (see e.g. [6, 8, 18, 13] and other works on visual tracking [2, 14, 16]).

We consider a simple robotic reaching task: the robot must touch a moving target using target information derived from the processing of visual data. This task can be decomposed into the analysis of three separate parts: (1) the target - the dynamics of the target including the system noise; (2) the motor system - speed and accuracy of the motor system; (3) the vision system - resolution and accuracy of processing; and the relationships between these parts: (a) the measurement noise (vision system tracking the target); (b) the distance from the manipulator to the target, (c) visual-motor calibration. Of interest here is the relationship between the visual processing, its compuational delay, and both the accuracy and resolution of the resulting target localization. For a particular manipulator, adequate tuning of the controller should enable motion to the commanded position with sufficient accuracy in time that depends only on the distance. Thus for a fixed distance to the target, any variation in motion time must depend on the accuracy of the commanded position (hence the visual system).

This paper revisits the integration of vision in robot motion control where vision processing may take place in parallel with the motion control *but* the motion *execution* is independent of the visual system. This form of motor control is thought to exist in humans where the decision of where to move and the movement itself are often thought to be independent [10, 11, 12, 17]. Studies of visual attention and human motor control suggest that vision *is not* utilized during motion execution (motions are ballistic). Some evidence suggests that vision is used at the onset of the task (to determine initial conditions), at the end of the task (to evaluate the motion and determine success, failure, etc.), and for periodic correction or modulation of the pattern of movements. In contrast, visual guided robot motions typically perform in a closed loop fashion and thus motion time is limited by the speed with which the visual sensing can be performed. In hunam systems with visual processing in parallel with motion control the motion is not 'servo-ed' on vision. While entire tasks (e.g. driving) cannot be performed without visual feedback, many tasks or *portions* of a task, are readily performed without tight integration of the vision system within the control loop. In this paper we explore the interaction of vision and motion control through analysis

*This research supported in part by NSF IRI-9703352

of robot reaching tasks for a simple manipulator. The ability to touch a target depends on accuracy, resolution, and target dynamics. The size of the target effects the required resolution of the visual processing, which in turn effects the computation time. The uncertainty in the target dynamics influences the required resolution. The interplay of these factors for touching a static target was addressed in [7]. Here we further explore the interaction of the target dynamics, the resolution of visual data and the accuracy of reaching.

2 Visual Tracking

There has been considerable effort to produce real-time algorithms for tracking moving objects. The ability to track a target depends upon the dynamics of the target (how fast it is moving), the computation time (amount of data to process), and the desired tracking accuracy. Often the visual tracking problem is formulated using a recursive filter framework [2, 14]. A stochastic model of the dynamics of the target with Gaussian noise is used (with the notation based on [14]):

$$dx(t) = Ax(t)dt + Bdw(t) \tag{1}$$

where $w(t) \sim N[0,Q]$ and with measurement process $y(t) = Hx(t)+v$, $v \sim N[0,R]$, where $R(t)$ is the resolution of the measurement process. Under Gaussian noise assumptions, an estimate of the target, $\hat{x}$, and the associate covariance $P = E[(x-\hat{x})\cdot(x-\hat{x})^T]$ are obtained using a Kalman-Bucy filter (see e.g. [4, 9]). Accuracy of tracking is expressed via the covariance $P(t)$. If $P(t)$ diverges then the process is not observable – tracking is not possible.

Olivier[14] gives a precise discussion on real-time observability of a one dimensional linear stochastic process with variable sample times. In visual tracking the sample time is typically the time used for processing visual data. Often this time is taken to be a constant (e.g. [13]). Mathematical formulations and analysis of tracking also assume is that the search window is sufficiently large such that the measurement process "works" , i.e. the target is found. (typically a 2σ search scale is used – see [2, 16]). If the resolution within that window is too large then tracking will also fail. Figure 1 graphically represents these ideas. If the resolution, R is proportional to the covariance, P, i.e. $R = \alpha P$ $(0 < \alpha \ll 1)$, and the sample time is proportional to the search window area (which is a function of the covariance P) then tracking is real-time observable (the covariance does not diverge) for **sufficiently slow dynamics** [14]. For example, the one dimensional linear system:

$$\begin{aligned} dx &= axdt + dw \ ; \ E[dw^2] = \Omega^2 dt &(2)\\ y(t_k) &= x(t_k) + v \ ; \ v(t_k) \sim N[0, R_k] &(3)\end{aligned}$$

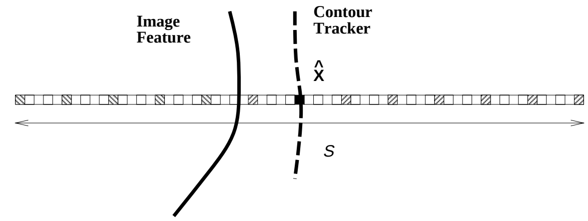

Figure 1: The width of the search region, S, should be large enough so that the target is found ($> 2\sigma$). Proportional resolution policies subsample that search region. The cartoon shows a of search window with the window size of 60 pixels and a resolution of 5 pixels. (For $P = \sigma^2 = 100$, we searched $\hat{x} \pm 3\sigma$ with a proportional resolution policy of $R = 0.05P = 5$ pixels. The processing (of only 10 pixels here), will produce an estimate $\hat{x}$ with accuracy limited by the resolution.

The Lyapanov equation for the covariance, P is

$$\frac{dP}{dt} = 2aP + \Omega^2$$

The Riccati equation with resolution $R_k = \alpha P_k$ is

$$P_{k+1} = \left(e^{2aT_k}\frac{\alpha}{1+\alpha}\right)P_k + \frac{\Omega^2}{2a}\left[e^{2aT_k} - 1\right] \tag{4}$$

where T_k is the sample time for the k^{th} measurement. The covariance evolves as

$$P_{k+1} = \gamma P_k + \beta.$$

The covariance decreases and hence the target can be tracked when $\gamma < 1$, or

$$e^{2aT_k}\frac{\alpha}{1+\alpha} < 1$$

This expression makes explicit the trade off between the sample time T_k, the ratio of resolution to covariance, α, and the system dynamics, a. If the system is sufficiently slow for the time scale used $a < a(T)$ then we can pick a resolution, αP, to ensure tracking. This is in agreement with the results of Blake *et al* [2] who also provides an algorithm for determining the spatio-temporal scale proportional to the estimated covariance.

The majority of Olivier's [14] work is for one dimensional tracking with linear system dynamics. Examples are presented for tracking Brownian motion, which can be considered a "slow" dynamical system. Blake *et al* [2] require approximate knowledge of the target dynamics and tracking is based on a constant acceleration dynamical model. These assumptions are invalid for non-linear, high speed motor systems such as robots.

3 Motion Control

For a given target speed, visual tracking can produce an estimate of the target position (at a particular accuracy and resolution). The visual system commands the motion system (which requires calibration between the vision and the motor system – see e.g. [5, 18]).

The speed with which motors can move and the delays associated with visual tracking make task times longer when using visual feedback tightly within the control loop. We compare the speed with which a tracker, as described in the previous section, can follow a manipulator motion to the speed with which an open loop motion can be performed.

As a simple example, consider the motion of a "1 link" manipulator (a simple rotor and link). We use a simple dynamical model describing the angular motion $\theta(t)$ for a given input torque $\tau(t)$

$$\tau = J\ddot{\theta}.$$

To displace the rotor by $\Delta\theta$ radians, given a torque limited motor ($\tau \le \tau_{max}$), and assuming bang-bang control, the minimum time for the motion is bounded from below and time is a function of the inertial properties of the link, J, the desired angular displacement, $\Delta\theta$, and the maximum torque, τ_{max}, ie. $T_{min} = T(\Delta\theta, J, \tau_{max})$. There is some overhead associated with computing the control parameters and "downloading" the motor commands to a motor controller. This time is typically independent of the size of the motion and can be assumed constant, a_o. With the execution time given by the time function the motion time for a displacement is:

$$MT \approx \underbrace{a_o}_{\text{overhead}} + \underbrace{b_o\sqrt{\frac{J\Delta\theta}{\tau_{max}}}}_{\text{task execution}} \tag{5}$$

If the control is "perfect" and there is no friction or other unmodeled phenomena, we can achieve this "optimal" motion time given by (5). In practice there will be errors from inadequate models of inertia, friction, etc. In addition, bang-bang control is only a theoretical model, instantaneous changes in acceleration are not realized. Due to these errors, the final position may not be the exact desired position. Feedback control can correct for unmodeled phenomena but the resulting time for the motion may exceed the "optimum" time. Feedback for such simple tasks usually relies on kinesthetic sensors (shaft encoders). We consider visual feedback here to demonstrate the timing issues for tracking a manipulator.

For this simple system, there are two points of note. First of all, when using bang-bang control the motion commands produce abrupt changes in the angular acceleration. The manipulator is initially at rest. When the motor commands are executed the manipulator moves with high accelerations. Observability constraints for visual tracking *require* sufficiently slow dynamics. The second point is that the accelerations are not constant, as most visual tracking models assume. In joint space the accelerations are constant but the resulting motions in workspace are highly non-linear. The dynamics governing motion of a rotor are linear in the angular position variable and hence PID control is highly successful. However, controlling the motion with visual feedback requires expressing the system dynamics in workspace (or camera space) variables resulting in non linear dynamics.

4 Reaching a target

To avoid having to synchronize the visual tracking to rapid manipulator motions (and because visual tracking may not be feasible for rapid manipulator motions), the control strategy must be modified. For many tasks it is acceptable to slow down the robot motion to facilitate visual tracking. However, there are tasks for which the actual motion does not *need* to be monitored. In reaching tasks, the motors systems can be commanded to move (independent of the visual system). The vision system can "saccade" to the expected the final manipulator position and re-locate the manipulator. *To do so the fingers have to be readily located in the image.* After the reach motion is completed, the vision system evaluates success or failure and may re-initiate a reach task. New motor commands are issued and the cycle repeats. Segmented motor control can use visual feedback for periodic correction or modulation of the pattern of movements, as is thought to occur in humans. As described earlier, this is the form of visual-motor interaction that is thought in human motor control. The control is broken into a sequence of steps which may be iterated.

Step 1 Estimate (if not known) the current manipulator position. Estimate the target position with respect to the camera. Without prior knowledge, the entire image needs to be searched. The initial covariance P_o is commensurate with the image size. The resolution, however, can be coarse to determine a rough estimate of the target location. Kinematic information from calibration or the image jacobian [18] converts the image location to a desired manipulator position.

Step 2 Determine the required motor commands to perform the motion. The motion is executed and we assume that the final position is sufficiently close to the commanded position. (This is not an unrealistic assumption as most well-tuned control strategies can achieve high accuracy. Examples in the literature in-

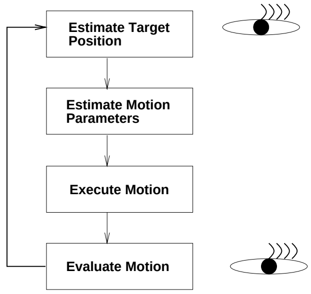

Figure 2: Graphical depiction of the vision/motion control strategy. Vision is used to monitor the motion, not as servo feedback.

clude [3, 20]).

Step 3 The resulting position for the manipulator will depend on many factors, such as the accuracy of the image jacobian calibration, the accuracy of the motion control strategy used, the models used for the manipulator dynamics, friction effects, etc. With respect to this work the relevant parameter is the covariance of the target estimate obtained from the image analysis. The initial covariance of the target estimate and the resolution at which the processing was performed give an approximation to the accuracy of the reaching task. If the covariance is small enough, and the target width is wide enough, one motion segment will suffice to "reach" the target. If the target width is less than the covariance, the process must be repeated (from step 1).

It can be shown that for each motion, the time to move is proportional to the square root of the distance moved

$$\text{time}_k \sim a_o + \lambda\sqrt{\Delta\theta} \sim a_o + \lambda\sqrt[4]{P_k}.$$

To achieve the desired accuracy may involve M motions (of decreasing size). The total motion time will be:

$$T_o + \sum_{k=1}^{M} \lambda\sqrt[4]{P_k}$$

where M is the iteration when the covariance is smaller than the target width ($R_M \approx W^2$ so $\alpha P_M \approx W^2$).

5 Experimental Setup

We performed some simple reaching experiments (figure 3). A plot of the motion time *vs.* the positional accuracy for a visual guided reach is shown in figure 4. As expected the curve

Figure 3: Robotic manipulator used for reaching experiments

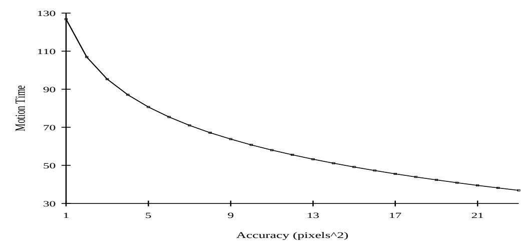

Figure 4: Plot of motion time vs. the positional accuracy.

6 Discussion

We are developing a framework in which one may decouple the vision system from the motor control system. Control loops that are tightly coupled to *slow* sensors, such as vision, are inherently limited. We explore whether tasks, such as reaching, can be achieved in an open-loop manner, much like the human system. As the task difficulty increases (narrower target), then the system must use intermediate feedback to reach the target (M is larger in section 3) and the motion time increases. For smaller target width, W, a finer resolution must be used and hence the reaching task time increases. Further work is necessary to refine our framework and account for all sources of uncertainty. In order to isolate the errors due to uncertainty in the motor control *vs.* errors in the image analysis, the system must be carefully calibrated. Our initial experiments assumed "perfect" motion control (lumping all uncertainty into the visual system). Further analysis is necessary to formulate the problem with explicit uncertainty parameters for the motor control.

Clearly some tasks will not lend themselves to this form of control strategy, however for various assembly tasks, the ability to separate the visual control from the motor will facilitate the incorporation of visual sensors in robotic devices.

References

[1] P. Allen, B. Yoshimi, and A. Timcenko. Real–time visual servoing. In *Proc. IEEE Int. Conf. Robotics and Automation*, volume 1, pages 851–856, 1991.

[2] A. Blake, R. Curwen, and A. Zisserman. A framework for spatio-temporal control in the tracking of

visual contours. *Int. Journal of Computer Vision*, 1993.

[3] J. Bobrow, S. Dubowsky, and J. S. Gibson. Time-optimal control of robotic manipulators along specified paths. *Int. Journal of Robotics Research*, 4(3):3–17, 1976.

[4] A. Bryson and Y. Ho. *Applied Optimal Control*. Hemisphere Publishing Corp., New York, 1975.

[5] A. Castano and S. Hutchinson. Visual compliance: Task-directed visual servo control. *IEEE Transactions on Robotics and Automation*, 10(3):334–342, 1994.

[6] J. Clark and N. Ferrier. Modal control of visual attention. In *Proceedings of the International Conference on Computer Vision*, pages 514–531, Tarpon Springs, Florida, December 1988.

[7] N. Ferrier. Achieving a fitts law relationship for visual guided reaching. In *Proc. 1998 IEEE International Conf. on Computer Vision*, Mumbai, India, 1998.

[8] N. Ferrier and J. Clark. The harvard binocular head. *International Journal of Pattern Recognition and AI*, March 1993.

[9] A. Gelb and the Technical Staff, Analytic Sciences Corporation. *Applied optimal estimation*. MIT Press, Cambridge, MA, 1974.

[10] S. Keele. *Attention and Human Performanace*. Goodyear Publishing Co., 1973.

[11] J. Kelso, editor. *Human Motor Behavior*. Lawrence Erlbaum Associates, 1982.

[12] R. Klein. Attention and movement. In G. Stelmach, editor, *Motor Control: Issues and Trends*. Academic Press, 1976.

[13] B. Nelson, N. Papanikolopoulos, and P. Khosla. Visual servoing for robotic assembly. In *Visual Servoing-Real-Time Control of Robot Manipulators Based on Visual Sensory Feedback*, pages 139–164. World Scientific Press, 1993.

[14] C. Olivier. Real-time observatiblity of targets with constrained processing power. *IEEE Transactions on Automatic Control*, 41(5):689–701, 1996.

[15] K. Pahlavan and J.-O. Eklundh. Heads, eyes and head-eye systems. In H. Christensen, K. W. Bowyer, and H. Bunke, editors, *Active Robot Vision*. World Scientific Press, 1993.

[16] J. Roberts and D. Charnley. Parallel attentive visual tracking. *Engineering Applications of Artificial Intelligence*, 7(2):205–215, 1994.

[17] R. Schmidt. More on motor programs. In J. Kelso, editor, *Human Motor Behavior*. Lawrence Erlbaum Associates, 1982.

[18] T. Schnackertz and R. Grupen. A control basis for visual servoing tasks. In *Proc. 1995 IEEE Conf. on Robotics and Automation*, Nagoya, Japan, 1995.

[19] J. Son, R. Howe, J. Wang, and G. Hager. Preliminary results on grasping with vision and touch. In *Proceedings of the IEEE Int. Conference on Intelligent Robots and Systems*, 1996.

[20] H. Yang and J.-J. Slotine. Fast algorithms for near-minimum-time control of robot manipulators. *The International Journal of Robotics Research*, 13(6):521–532, 1994.