

SINGLE-CAMERA VISUAL SERVOING¹

João P. Hespanha
hespanha@usc.edu

Dept. of Electrical Engineering–Systems, University of Southern California
3740 McClintock Avenue, Los Angeles, CA 90089-2563

Abstract

This paper deals with the problem of driving the position of a point-robot to a goal point, using measurements provided by a single camera. The camera can be moved to compensate for the ambiguity in depth. It is shown that, although the system consisting of the point robot together with the moving camera is observable, there is no output-feedback controller capable of asymptotically stabilizing an equilibrium point of the closed-loop system. However, it is possible to design a controller that drives the robot to the goal, provided that the position of the camera does not converge. We give a hybrid controller—combining logic-based switching with continuous dynamics—that accomplishes this. The stability of the controller is also analyzed when there is miscalibration between the robot and the camera.

1 Introduction

Consider the problem of moving a point-robot to a goal position, using measurements provided by a pin-hole camera. The camera is able to see both the point-robot and the goal position. It is well known that the measurements provided by a single fixed camera are not sufficient to solve this problem. In fact, a single camera does not provide sufficient information to determine whether or not the task has been accomplished since it cannot distinguish between points in a straight line passing through the optical center of the camera. This is often referred to as the *depth ambiguity*. To resolve it one can use a pair of fixed cameras or a single camera that is allowed to move. We pursue here the second approach.

We will see below that the use of a single camera leads to a more complex control problem, yet the image processing becomes faster and simpler. This is because (i) the number of images that have to be processed per unit of time is halved and (ii) the “image correspondence” problem is much simplified. Note that in a single-camera system one only has to establish corre-

spondences between images taken by the same camera at time instants separated by a few milliseconds. In this case, the shape of the objects does not change substantially from one image to the next and correlation-based tracking algorithms are very effective in solving the correspondence problem (cf. e.g., [1]). The simplification of the vision algorithms due to (i) and (ii) greatly outweighs the increase in complexity of the control algorithms.

For simplicity, in this paper we deal with the planar case and place the point-robot, the camera, and the goal position all in the same plane. It is important to emphasize that even in this planar case there is depth ambiguity because the camera is coplanar with the robot and the goal position. We will see that this ambiguity can be resolved by letting the camera move, e.g., along a straight line.

The setup considered here is shown in Figure 1. The robot can move in any direction to reach the goal position, but the camera is constrained to move along a straight line. Without loss of generality we place the first axis of the coordinate system aligned with the line over which the camera moves and the second axis perpendicular to the first and passing through the goal point. The Cartesian coordinates of the robot are denoted by (x_1, x_2) , the Cartesian coordinates of the camera by $(x_3, 0)$, and the Cartesian coordinates of the goal position by $(0, r)$. Both x_2 and r are assumed positive. The robot/camera system admits the following simple

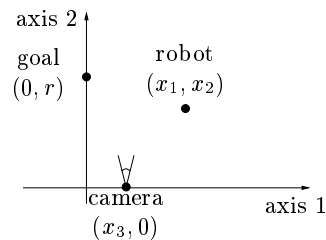


Figure 1: Planar Robot/Camera System

kinematic model

$$\dot{x} = u, \quad (1)$$

¹This research was supported by the Office of Naval Research.

where $x := [x_1 \ x_2 \ x_3]'$ and $u := [u_1 \ u_2 \ u_3]'$. The pin-hole camera is able to sense both the robot and the goal position. The corresponding measurements are

$$y_1 = \frac{x_1 - x_3}{x_2}, \quad y_2 = -\frac{x_3}{r}, \quad (2)$$

respectively. Here, we assumed that the camera's focal length is equal to 1 and that its optical axis is aligned with the second axis of the coordinate system. If this was not the case, the above variables could still be computed from the camera measurements.

The task is to move the robot to the goal position. This is equivalent to driving x_1 to zero and x_2 to r , using only the measurements $y := [y_1 \ y_2]'$ for feedback. To accomplish this task we propose a controller that combines logic with continuous dynamics. The resulting closed-loop system is therefore *hybrid* (cf. surveys [2, 3] and references therein). The motivation for this type of control law comes from the fact that, as we shall see shortly, standard asymptotic stabilization of the closed-loop system around an equilibrium point is not possible.

The problem addressed in this paper falls under what is often called *visual servoing* [4]. Although the use of vision in the control of robotic manipulators was proposed at least 20 years ago, only in the last few years it has been possible to satisfy the computational demands posed by the image processing it requires. Once fairly reliable real-time image processing became available, this field exhibited a large growth apparent in the survey papers [5, 6]. Most visual-based control of robotic manipulators reported in the literature employ a pair of cameras to resolve depth ambiguity [7, 8, 9, 10, 11, 12, 13]. In some cases a single camera is used when the robot moves on a plane and the optical center of the camera is placed away from that plane [14, 15]. However, this type of planar configuration does not suffer from depth ambiguity and the camera measurements are sufficient to uniquely determine the Cartesian coordinates of the points seen in the image space.

This paper is organized as follows. In Section 2 we present some fundamental limitations in the design of feedback controllers for robot positioning using a single camera. We show that, when utilizing output-feedback from a single camera, standard techniques based on asymptotic stabilization of the closed loop to an equilibrium point will fail. This, in spite of the fact that the system is observable through the single-camera measurements. Section 3 describes a hybrid controller that can be used to solve the visual servoing task and in Section 4 we show that this controller is robust with respect to hand/eye miscalibration. Concluding remarks and directions for future research are given in Section 5.

2 Limitations on asymptotic stabilizability

One could be tempted to solve the problem formulated above by linearizing the process (1)–(2) around some equilibrium point $\bar{x} := [0 \ r \ \alpha]$, $\alpha \in \mathbb{R}$, and designing a stabilizing controller for the linearized system. If this were possible, the resulting controller would asymptotically stabilize the closed-loop system (at least, locally) and therefore drive x_1 to zero and x_2 to r . However, this is not possible since the linearization of (1)–(2) around \bar{x} , which is given by

$$\dot{\tilde{x}} = u, \quad \dot{\tilde{y}} = \frac{1}{r} \begin{bmatrix} 1 & \alpha/r & -1 \\ 0 & 0 & -1 \end{bmatrix} \tilde{x},$$

is not detectable. It turns out that this is a manifestation of a more fundamental limitation on the class of output-feedback controllers that asymptotically stabilize (1)–(2).

Consider a n -dimensional dynamical system defined by

$$\dot{\bar{x}} = A(\bar{x}, t), \quad \bar{x} \in \mathbb{R}^n. \quad (3)$$

We say that a set $\mathcal{G} \subset \mathbb{R}^n$ is *uniformly stable for the system* (3), if for every $\epsilon > 0$ there is some $\delta > 0$ such that, for any initial time $t_0 \geq 0$ and any initial state¹ $\bar{x}(t_0) \in \mathbb{R}^n$ with $d(\bar{x}(t_0), \mathcal{G}) \leq \delta$, the solution to (3) exists globally and $d(\bar{x}(t), \mathcal{G}) \leq \epsilon$ for all $t \geq t_0$ [16]. Moreover, if $d(\bar{x}(t), \mathcal{G}) \rightarrow 0$ as $t \rightarrow \infty$ we say that \mathcal{G} is asymptotically stable.

We will see next that there is no output-feedback locally Lipschitz controller

$$\dot{z} = F(z, y, t), \quad u = G(z, y, t), \quad z \in \mathbb{R}^n, \quad (4)$$

that renders any set of the form

$$\mathcal{G}_\alpha := \{ [0 \ r \ \alpha \ z']' : z \in \mathbb{R}^n \}, \quad (5)$$

with $\alpha \in \mathbb{R}$, asymptotically stable for the closed-loop system defined by (1), (2), and (4). This means that there is no output-feedback locally Lipschitz controller that is capable of driving the state of (1) to an equilibrium point of this system for which our task is achieved. The following lemma, taken from [17], formalizes this observation.

Lemma 1 *There exists no integer n and functions $F, G : \mathbb{R}^n \times \mathbb{R}^2 \times [0, \infty) \rightarrow \mathbb{R}^n$ that are locally Lipschitz with respect to the first two arguments and continuous with respect to the third one, for which the set \mathcal{G}_α defined in (5) is uniformly asymptotically stable for the closed-loop system defined by (1), (2), and (4).*

¹Given a point $v \in \mathbb{R}^n$ and a subset \mathcal{V} of \mathbb{R}^n , we denote the distance from v to \mathcal{V} by $d(v, \mathcal{V})$, i.e., $d(v, \mathcal{V}) = \inf_{\bar{v} \in \mathcal{V}} \|v - \bar{v}\|$.

Remark 1 Although Lemma 1 was proved for finite dimensional, time-varying, Lipschitz continuous controllers, it can be generalized to much larger classes of controllers. These include, e.g., infinite dimensional, hybrid, discontinuous², etc.

The difficulties in designing an output-feedback controller presented before do not arise from the lack of observability for the system (1)–(2). In fact, this system is observable in that any two initial states can be distinguished from each other by observing the output over a finite time interval [18]. This is proved below.

Lemma 2 For every $r > 0$ the system (1)–(2) is observable on \mathbb{R}^3 . Moreover, there exists a constant input which can be used to distinguish any two states and determine the value of r by observing the output over a finite time interval.

Proof: To prove this proposition it is sufficient to find a constant input to the system (1)–(2) which allows one to uniquely determine the initial state $x(t_0)$, $t_0 \geq 0$, and the value of $r > 0$ by measuring the output y on the finite interval $[t_0, t_0 + T]$, $T > 0$. One possible input with this property is

$$u_1(t) = 0, \quad u_2(t) = 0, \quad u_3(t) = 1, \quad t \in [t_0, t_0 + T].$$

It is straightforward to show that, with this input, we have that

$$-\frac{1}{x_2(t_0)} = -\frac{6}{T^2} \int_{t_0}^{t_0+T} y_1(\tau) d\tau + \frac{12}{T^3} \int_{t_0}^{t_0+T} \tau y_1(\tau) d\tau, \quad (6)$$

$$\frac{x_1(t_0) - x_3(t_0)}{x_2(t_0)} = \frac{4}{T} \int_{t_0}^{t_0+T} y_1(\tau) d\tau - \frac{6}{T^2} \int_{t_0}^{t_0+T} \tau y_1(\tau) d\tau, \quad (7)$$

$$-\frac{1}{r} = -\frac{6}{T^2} \int_{t_0}^{t_0+T} y_2(\tau) d\tau + \frac{12}{T^3} \int_{t_0}^{t_0+T} \tau y_2(\tau) d\tau, \quad (8)$$

$$-\frac{x_3(t_0)}{r} = \frac{4}{T} \int_{t_0}^{t_0+T} y_2(\tau) d\tau - \frac{6}{T^2} \int_{t_0}^{t_0+T} \tau y_2(\tau) d\tau, \quad (9)$$

from which $x(t_0)$ and r can be uniquely determined. The value of x at the final time $t_0 + T$ can also be

²When the conditions that guarantee uniqueness of solution are dropped, one needs to adapt the definition of stability by demanding that *all* solutions satisfy the usual continuity/convergence properties.

easily determined since

$$x_1(t_0 + T) = x_1(t_0), \quad (10)$$

$$x_2(t_0 + T) = x_2(t_0), \quad (11)$$

$$x_3(t_0 + T) = x_3(t_0) + T. \quad (12)$$

■

We have seen that it is not possible to build a feedback controller that asymptotically stabilizes the closed-loop system to an equilibrium point $[0 \ r \ \alpha]$, $\alpha \in \mathbb{R}$, of the process (1), i.e., to make the set \mathcal{G}_α asymptotically stable. Fortunately, to accomplish the task at hand this is not necessary since we do not require the camera to converge to some pre-specified position. In fact, it suffices to make x_1 converge to 0 and x_2 converge to r or, to be more precise, to make the set

$$\mathcal{G} := \bigcup_{\alpha \in \mathbb{R}} \mathcal{G}_\alpha = \{[0 \ r \ \alpha \ z]^T : z \in \mathbb{R}^n, \alpha \in \mathbb{R}\}, \quad (13)$$

asymptotically stable. In the next section we construct a controller that does precisely this. This controller will not make x_3 converge to an equilibrium value because this would violate Lemma 1.

Remark 2 In the prove of Lemma 2, to recover $x(t_0)$ and r from the value of y in the interval $[t_0, t_0 + T]$, it would be possibly to use only the initial and final values of the output, namely $y(t_0)$ and $y(t_0 + T)$. However, using integrals of the output over the whole interval, we obtain estimates for $x(t_0)$ and r that are less sensitive to measurement noise. This will become an issue below when we use (6)–(9) to construct an output feedback controller.

3 Hybrid controller

We have seen in Lemma 2 that it is possible to determine both the state x of the system and r by observing the output over a finite interval, while a constant input is applied to the system. A controller could then be designed in the following way:

1. Apply the constant input and estimate the state and r over an interval of length $T_{\text{look}} > 0$.
2. Apply an open-loop control that takes the system from the estimated state to $[0 \ r \ 0]$ in a finite time T_{move} .

In an ideal situation, after the second step, the robot would be precisely in the desired position. However, this is unlikely to happen because of noise and unmodeled dynamics in the manipulator. To “robustify” this algorithm we make this procedure iterative by adding a third step

3. Goto 1.

The resulting controller is hybrid with a discrete state variable σ that takes two distinct values, one for each step. The set where σ takes values is called the discrete state-space and we take it to be $\mathcal{S} := \{\text{“look”}, \text{“move”}\}$. This controller also requires five continuous states: a variable τ that will function as a timer to define the transitions between the two discrete states and four variables z_1, z_2, z_3, z_4 that will be used to keep track of the integrals in (6)–(9) needed to estimate the state. Formally, the hybrid controller—which we denote by C_H —is defined by the following rules:

- The controller is initialized with

$$\begin{aligned} \sigma(0) &= \text{“look”}, & \tau(0) &= 0, \\ z_i(0) &= 0, & i &\in \{1, 2, 3, 4\}. \end{aligned}$$

- While $\sigma(t) = \text{“look”}$ we have

$$u_1(t) = 0, \quad u_2(t) = 0, \quad u_3(t) = 1,$$

and

$$\begin{aligned} \dot{\tau} &= 1, & \dot{z}_1 &= y_1(t), & \dot{z}_2 &= ty_1(t), \\ \dot{z}_3 &= y_2(t), & \dot{z}_4 &= ty_2(t). \end{aligned}$$

- While $\sigma(t) = \text{“move”}$ we have

$$\begin{aligned} \dot{\tau} &= 1, & \dot{z}_1 &= 0, & \dot{z}_2 &= 0, \\ \dot{z}_3 &= 0, & \dot{z}_4 &= 0. \end{aligned}$$

and u_1, u_2, u_3 constant so that x will reach $[0 \ r \ 0]^T$ in an interval of length T_{move} , starting from the position estimated through (6)–(12). The values of the integrals in (6)–(9) are obtained directly from the z_i .

- σ switches from “look” to “move” when τ reaches T_{look} . At the same time τ is reset to zero.

σ switches from “move” to “look” when τ reaches T_{move} . At the same time τ , and the z_i are reset to zero.

It would be straightforward to express this hybrid system, for example, in the formalism of [3]. However, the simple description above is sufficient for our purposes. We can now state the following:

Theorem 1 *The hybrid closed-loop system, defined by the dynamical process (1)–(2) and the hybrid controller C_H , renders the set \mathcal{G} in (13) asymptotically stable³.*

³There is some abuse of notation here since set stability was defined for dynamical systems but we are using it now for a hybrid system. However, the extension is straightforward.

This theorem is a simple consequence of the fact that, while the controller is in the “look” discrete state, both x_1 and x_2 remain unchanged, and when the controller switches to the “move” discrete state, (x_1, x_2) go to $(0, r)$ along a straight line. This means that the distance from the state to \mathcal{G} never increases and actually converges to zero in finite time (at least in the idealized situation addressed in this section). Figure 2 shows the evolution of the closed-loop system for $r = 1$, $T_{\text{look}} = 1/2$, and $T_{\text{move}} = 2$.

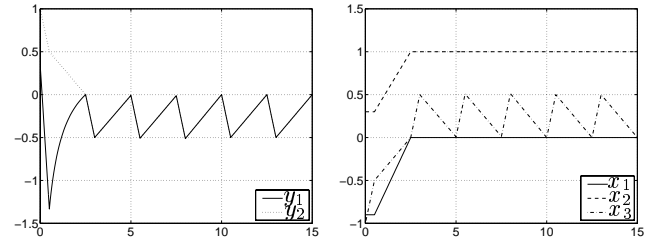


Figure 2: Closed-loop response with the hybrid controller.

4 Robustness with respect to hand/eye miscalibration

Theorem 1 assumes that there is no noise and the calibration between the robot (hand) and the camera (eye) is perfect. In practice, this will never be exactly true. We now analyze the effect of miscalibration between the robot and the camera in the stability of the closed-loop system.

When the coordinate systems of the robot and the camera are not perfectly aligned, the velocity commands u_1 and u_2 sent to the robot will make it move along a direction different than the one expected. In this case, the process dynamics (1) must be replaced by

$$\dot{x}_{12} = u_{12} + \Delta(u_{12}), \quad \dot{x}_3 = u_3, \quad (14)$$

where $x_{12} := [x_1 \ x_2]^T$, $u_{12} := [u_1 \ u_2]^T$, and $\Delta : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ denotes a possibly nonlinear function that models the miscalibration. We assume that the miscalibration does not introduce bias and therefore that $\Delta(0) = 0$.

Suppose that at time $t_n \geq 0$ the robot starts the “look” phase. During this whole phase, no velocity commands are sent to the robot and therefore

$$x_{12}(t_n + T_{\text{look}}) = x_{12}(t_n).$$

Moreover, since Δ plays no role here, the estimation of the state is still accurate. During the “move” phase, the controller will send to the robot a constant velocity of

$$u_{12}(t) = \frac{1}{T_{\text{move}}}(-\tilde{x}_{12}(t_n)),$$

$$\forall t \in (t_n + T_{\text{look}}, t_n + T_{\text{look}} + T_{\text{move}}),$$

where $\tilde{x}_{12} := x_{12} - [0 \ r]'$, to make it move to the position $(0, r)$. However, due to the miscalibration, we have that

$$x_{12}(t_n + T_{\text{look}} + T_{\text{move}}) = x_{12}(t_n)$$

$$+ \int_{t_n + T_{\text{look}}}^{t_n + T_{\text{look}} + T_{\text{move}}} \left(-\frac{1}{T_{\text{move}}} \tilde{x}_{12}(t_n) + \Delta \left(-\frac{1}{T_{\text{move}}} \tilde{x}_{12}(t_n) \right) \right) dt,$$

and therefore

$$\tilde{x}_{12}(t_n + T_{\text{look}} + T_{\text{move}}) = T_{\text{move}} \Delta \left(-\frac{1}{T_{\text{move}}} \tilde{x}_{12}(t_n) \right). \quad (15)$$

Suppose now that

$$\gamma := \sup_{z \in \mathbb{R}^2} \frac{\|\Delta(z)\|}{\|z\|} < 1,$$

and therefore that

$$\sup_{z \in \mathbb{R}^2} \frac{\|T_{\text{move}} \Delta \left(-\frac{1}{T_{\text{move}}} z \right)\|}{\|z\|} = \sup_{z \in \mathbb{R}^2} \frac{\|\Delta(z)\|}{\|z\|} = \gamma < 1.$$

Because of (15) we then obtain

$$\|\tilde{x}_{12}(t_n + T_{\text{look}} + T_{\text{move}})\| \leq \gamma \|\tilde{x}_{12}(t_n)\|.$$

Iterating the above inequality, one concludes that \tilde{x}_{12} converges to zero exponentially fast as time goes to infinity. The following was proved.

Theorem 2 *The hybrid closed-loop system, defined by the miscalibrated dynamical process (14), (2) and the hybrid controller C_H , renders the set \mathcal{G} in (13) asymptotically stable, provided that*

$$\sup_{z \in \mathbb{R}^2} \frac{\|\Delta(z)\|}{\|z\|} < 1.$$

Figure 3 shows the evolution of the closed-loop system (14), (2) with the hybrid controller C_H , for $r = 1$, $T_{\text{look}} = 1/2$, $T_{\text{move}} = 2$, and $\Delta := I_{2 \times 2} - \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 0.9 \end{bmatrix}$, with $\theta := 20$ degrees. Figure 4 shows a similar simulation but with the camera measurements corrupted by white noise.

5 Conclusions

In this paper we propose a hybrid controller to drive the position of a point-robot to a goal point, using

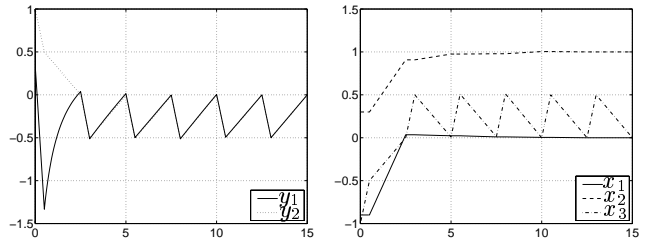


Figure 3: Closed-loop response of the miscalibrated system without noise.

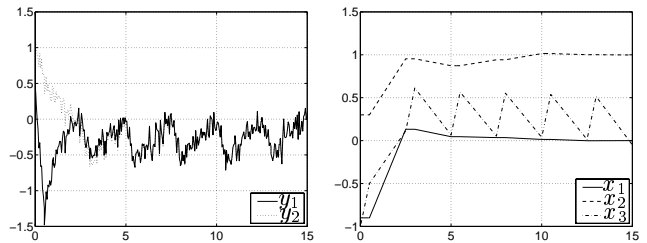


Figure 4: Closed-loop response of the miscalibrated system with measurement noise.

measurements provided by a single camera. Under this controller the position of the camera does not converge asymptotically. This is shown to be necessary for the system at hand.

We provide a stability analysis for the hybrid closed-loop system when there is miscalibration between the robot and the camera. We are in the process of also analyzing the robustness of the system with respect to noise and other types of unmodeled dynamics, e.g., miscalibration in the intrinsic parameters of the camera. Extensions of the results here to three dimensions and to a dynamical model for the robot seem straightforward.

References

- [1] G. D. Hager, “The “X-vision” system: A general purpose substructure for real-time vision-based robotic,” in *Proc. Workshop on Vision for Robots*, pp. 56–63, 1995.
- [2] A. S. Morse, “Control using logic-based switching,” in *Trends in Control: A European Perspective* (A. Isidori, ed.), pp. 69–113, London: Springer-Verlag, 1995.
- [3] M. D. Lemmon, K. X. He, and I. Markovskiy, “Supervisory hybrid systems,” *IEEE Contr. Syst. Mag.*, vol. 19, pp. 42–55, Aug. 1999.
- [4] J. Hill and W. T. Park, “Real time control of a robot with a mobile camera,” in *Proc. of the 9th ISIR*, pp. 233–246, Mar. 1979.

- [5] P. I. Corke, "Visual control of robot manipulators—a review," in *Visual Servoing* (K. Hashimoto, ed.), pp. 1–32, World Scientific, 1994.
- [6] S. A. Hutchinson, G. D. Hager, and P. I. Corke, "A tutorial on visual servo control," *IEEE Trans. Robot. Automat.*, vol. 12, pp. 651–670, Oct. 1996.
- [7] A. C. Sanderson and L. E. Weiss, "Adaptive visual servo control of robots," in *Robot Vision* (A. Pugh, ed.), pp. 107–116, IFS Ltd./Springer-Verlag, 1983.
- [8] K. Hashimoto, T. Kimoto, T. Ebine, and H. Kimura, "Manipulator control with image-based visual servo," in *Proc. IEEE Int. Conf. on Robot. and Automat.*, pp. 2267–2272, May 1991.
- [9] N. Hollinghurst and R. Cipolla, "Uncalibrated stereo hand-eye coordination," *Image and Vision Computing*, vol. 12, pp. 187–192, Apr. 1994.
- [10] G. D. Hager, W.-C. Chang, and A. S. Morse, "Robot hand-eye coordination based on stereo vision," *IEEE Contr. Syst. Mag.*, vol. 15, pp. 30–39, Feb. 1995.
- [11] A. A. Rizzi and D. E. Koditschek, "An active visual estimator for dextrous manipulation," *IEEE Trans. Robot. Automat.*, vol. 12, pp. 697–713, Oct. 1996.
- [12] G. D. Hager, "A modular system for robust hand-eye coordination," *IEEE Trans. Robot. Automat.*, vol. 13, no. 4, pp. 582–595, 1997.
- [13] J. P. Hespanha, Z. Dodds, G. D. Hager, and A. S. Morse, "What tasks can be performed with an uncalibrated stereo vision system?," Special Issue on Vision Research at Yale of the *Int. J. of Computer Vision*, vol. 35, pp. 65–85, Nov. 1999.
- [14] S. W. Wijesoma, D. F. H. Wolfe, and R. J. Richards, "Eye-to-hand coordination for vision-guided robot control applications," *Int. J. of Robot. Res.*, vol. 12, pp. 65–78, Feb. 1993.
- [15] R. Kelly, "Robust asymptotically stable visual servoing of planar robots," *IEEE Trans. Robot. Automat.*, vol. 12, pp. 759–766, Oct. 1996.
- [16] J. P. LaSalle, *The Stability of Dynamical Systems*. No. 25 in Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, 1976.
- [17] J. P. Hespanha, "Single-camera visual servoing," tech. rep., EE-Systems Dept., University of Southern California, Los Angeles, CA, Feb. 2000.
- [18] A. Isidori, *Nonlinear Control Systems*. Communications and Control Engineering Series, Berlin: Springer-Verlag, 3rd ed., 1995.