Proceedings of the 39th IEEE
Conference on Decision and Control
Sydney, Australia • December, 2000

A New Hybrid Image-Based Visual Servo Control Scheme

Peter I. Corke
CSIRO Manufacturing Science & Technology
Pinjarra Hills
AUSTRALIA 4069.
pic@cat.csiro.au

Seth A. Hutchinson
Beckman Institute for Advanced Technology
University of Illinois at Urbana-Champaign
Urbana, Illinois, USA 61801
seth@uiuc.edu

Abstract

In image-based visual servo control, where control is effected with respect to the image, there is no direct control over the Cartesian velocities of the robot end effector. As a result, the robot executes trajectories that are desirable in the image, but which can be indirect and seemingly contorted in Cartesian space. In this paper we describe the cause of these phenomena, and introduce a new partitioned approach to visual servo control that overcomes the problem. In particular, we decouple the z-axis rotational and translational components of the control from the remaining degrees of freedom. Then, to guarantee that all features remain in the image throughout the entire trajectory, we incorporate a potential function that repels feature points from the boundary of the image plane. We illustrate our new control scheme with a variety of simulation results.

1 Introduction

In visual servo systems, information obtained from the vision system is used to control the motion of the robot in real-time, as opposed to older systems that used vision only to determine the initial state of the world, prior to task execution.

There are two basic approaches to visual servo control: Image-Based Visual Servo (IBVS), and, Position-Based Visual Servo (PBVS). In IBVS, which will be briefly reviewed in Section 2, an error signal is measured in the image, and is mapped directly to actuator commands (see, e.g., [16, 7]). There are however problems associated with IBVS systems since the control law involves a Jacobian mapping between image space velocities and velocities in the robot's workspace. These include singularities or poor conditioning in this Jacobian, and the lack of direct control over the Cartesian velocities of the robot end effector. Thus, trajectories that the robot executes can appear quite contorted in the Cartesian space.

These performance problems with IBVS systems have led to the recent introduction of several hybrid methods [12, 13, 4]. we present a new partitioned visual servo control scheme that overcomes a number of the performance problems faced by classical IBVS but with less computation than the proposed hybrid method. The basic idea is to decouple the z-axis motions (including both the translational component and rotational component) from the other degrees of freedom, and to derive separate controllers for these z-axis motions.

2 Traditional IBVS

In this section we present a very brief review of Image-Based Visual Servo control. Let $r = (x, y, z)^T$ represent coordinates of the end-effector, and $\dot{r} = (T_x, T_y, T_z, \omega_x, \omega_y, \omega_z)^T$ represent the corresponding end-effector velocity, composed of a linear velocity $\mathbf{v} = (T_x, T_y, T_z)^T$ and angular velocity $\omega = (\omega_x, \omega_y, \omega_z)^T$. Let $f = (u, v)^T$ be the image-plane coordinates of a point in the image and $\dot{f} = (\dot{u}, \dot{v})^T$ the corresponding velocities. The image Jacobian relationship is given by

$$\dot{f} = J(f, r)\dot{r}, \tag{1}$$

with

$$J = \begin{bmatrix} \frac{\lambda}{z} & 0 & \frac{-u}{z} & \frac{-uv}{\lambda} & \frac{\lambda^2 + u^2}{\lambda} & -v \\ 0 & \frac{\lambda}{z} & \frac{-v}{z} & \frac{-\lambda^2 - v^2}{\lambda} & \frac{uv}{\lambda} & u \end{bmatrix} \tag{2}$$

in which λ is the focal length for the camera. Derivations of this can be found in a number of references including [10, 1, 8].

The simplest approach to IBVS is to merely use (1) to construct the control law

$$\mathbf{u} = \Gamma J^{-1}(f, r)\dot{f} \tag{3}$$

in which $\dot{f}$ is the desired feature motion on the image plane, Γ is a gain matrix, and $\mathbf{u} = \dot{r}$ is the control input, an end-effector velocity. Of course this approach assumes that the image Jacobian is square and nonsingular, and when this is not the case, a generalized inverse, J^+, is used.

0-7803-6638-7/00$10.00 © 2000 IEEE

3 Performance Issues

The two most prominent performance issues that confront IBVS systems, are the suboptimal camera motions during task execution and problems associated with estimating the distance to the target.

3.1 Suboptimal Cartesian Motions

A commonly mentioned criticism of IBVS is that the Cartesian paths often involve large camera motions, which are undesirable. Often the camera moves away from the target in a normal direction and then returns, a phenomenon we refer to as *camera retreat.* Such motion is not time optimal, requires large and possibly unachievable robot motion, and is a seemingly nonintuitive solution to the required image plane motion. Figure 1 illustrates the problem. In Figure 1(a), the feature points are seen to be driven on straight line trajectories to their goal positions, producing a large, and seeminly unnecessary, motion in the z-direction, seen in Figure 1(c).

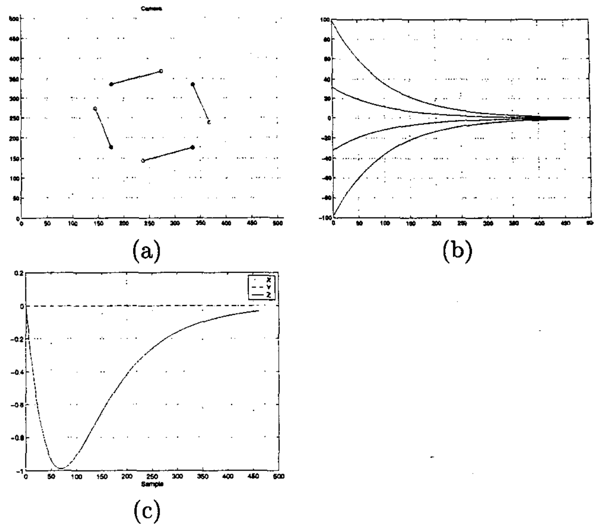


Figure 1: IBVS for pure target rotation (0.3 rad). (a) Image-plane feature motion (initial location is ○, desired location is ●), (b) Feature error trajectory, (c) Cartesian translation trajectory.

In [2], Chaumette introduced an extreme version of this problem, which we refer to as the Chaumette Conundrum, illustrated in Figure 2. Here, the desired camera pose corresponds to a pure rotation about the optic axis by π rad, i.e., the image feature point with initial coordinates (u, v) has the desired coordinates $(-u, -v)$. Since control laws of the form given in (3) drive the feature points in straight lines even for the case of pure target rotation, in this case the feature points are driven toward the origin, which corresponds to a singularity in the image Jacobian. The singularity arises because the feature points will reach the origin when the camera retreats to a distance of infinity, and no motion can be observed. We note that, as mentioned in [2], this problem cannot be detected by simply examining the image Jacboian, since the image Jacobian is well conditioned (at least initially). We use the term *IBVS failure* to refer to cases for which the system fails to achieve its goal.

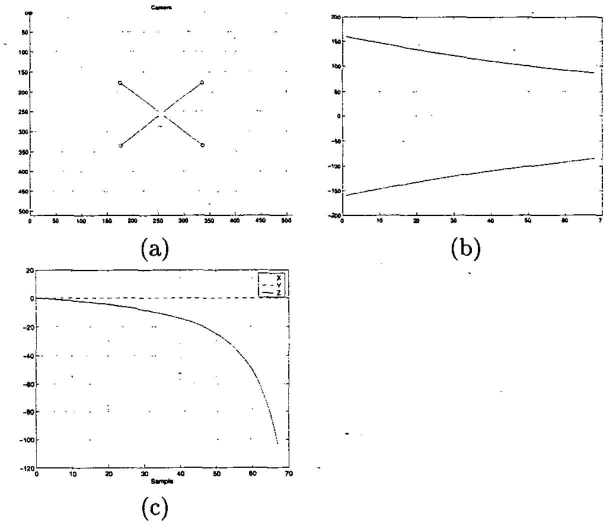


Figure 2: Performance of classical IBVS with the Chaumette example. (a) Image-plane feature motion (initial location is ○, desired location is ●), (b) Feature error trajectory, (c) Cartesian translation trajectory.

At first it might seem that some rotational motion of the camera about its optic axis should be induced for the Chaumette Conundrum; however, this is not the case. The ω_z component of (3) is given by

$$\omega_z = (J^+)_6 \dot{f} \qquad (4)$$

in which $(J^+)_6$ denotes the bottom row of the generalized inverse. In this particular case, even though $\dot{f} \neq 0$, the inner product is zero, i.e., the various contributions to rotational velocity cancel one another.

This camera retreat phenomenon can be explained in simple geometric terms, leading to a model that predicts the magnitude of the camera retreat motion. For the example of Figure 1, a pure rotational motion of the camera would cause the points to follow an arc from

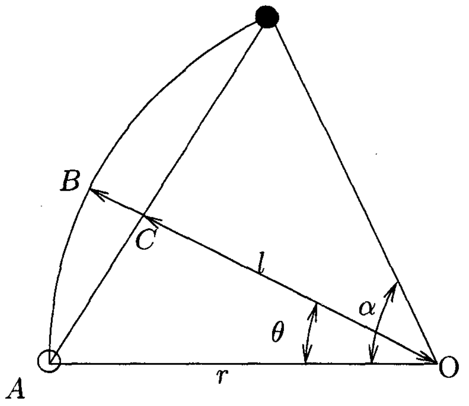


Figure 3: Camera retreat model.

point A to point B, as shown in Figure 3. In order for the points to follow a straight line, as specified by (3), the scale must be changed so as to move the point from B to C. The required change in scale is given simply by the ratio of the distances OC and OB. The scale reduction attains its maximum value at $\theta = \alpha/2$ for which

$$\left(\frac{OC}{OB}\right)_{max} = \cos\frac{\alpha}{2}. \tag{5}$$

In the IBVS the reduction in scale is achieved by moving the camera away from the target. The reduction in the apparent length of the line segment is inversely proportional to the distance that the camera retreats, and therefore,

$$\frac{OC}{OB} = \frac{d_{targ}}{d} \tag{6}$$

in which d is the current distance to the target, and d_{targ} is the desired target distance, and assuming the camera is moving normal to the target. The maximum reduction is thus given by

$$d_{max} = \frac{d_{targ}}{\cos\frac{\alpha}{2}}. \tag{7}$$

For the Chaumette Conundrum, in which $\alpha = \pi$, the model accurately predicts infinite camera retreat.

At first it might seem that the introduction of line segment features would solve the problem, since the orientation of such a segment is unambiguous. Chaumette notes that such an approach is not guaranteed to solve the performance problems [2], and our own simulation results support this conclusion. For the Chaumette Conundrum the addition of this feature does command *some* camera rotation, but the camera retreat is still very significant.

There are a variety of possible solutions to this camera retreat problem. The requirement that points move in a straight line to their goal positions could be relaxed, giving rise to an image feature trajectory planning problem. The target depth, z, could be underestimated, causing the points to move in an arc instead of a straight line, reducing the magnitude of camera retreat (this is mentioned in [2]), but this will still fail for the Chaumette Conundrum, in which no camera rotation occurs. Finally, hybrid approaches that combine aspects of IBVS and PBVS systems can be developed as discussed in Sections 4 and and 5.

3.2 How do we Estimate Depth?

As can be seen in equation (2), the image Jacobian is a function of the unknown depth, z. A number of researchers have proposed methods for dealing with this problem. The classical solution is to use standard computer vision techniques to estimate the value for z [7]; however, this approach amounts to performing a 3D reconstruction of the scene, and brings with it the same drawbacks faced by position-based visual servo schemes. A second approach is to estimate the value of z online, as demonstrated by Papanikolopoulos et al.[14] using adaptive control techniques. Finally, one can merely assume a constant value for the depth[5], an approach that is reasonable if the motion of the object remains approximately in a plane parallel to the image plane.

Figure 4 illustrates the effect of using a constant value $\hat{z}$ for the target depth. When the target depth is overestimated the trajectories bend inward, and when the target depth is underestimated the trajectories bend outward.

In our simulations we have inferred the necessary depth information from image features. However, any of the above approaches could be substituted.

4 Hybrid Approaches

A number of authors [12, 13, 4] have recently addressed the problems above by proposing hybrid control architectures. These methods rely on recent results in computing the epipolar geometry that relates a pair of images. In particular, the camera configurations that correspond to the initial and desired images are related by a homography matrix, which can be decomposed into the translational and rotational components of the motion between the two camera configurations. This homography matrix can be computed from a set of corresponding points in the initial and desired images.

The homography must then be decomposed to extract the rotational component and the problem of non-unique solutions must be dealt with. This method

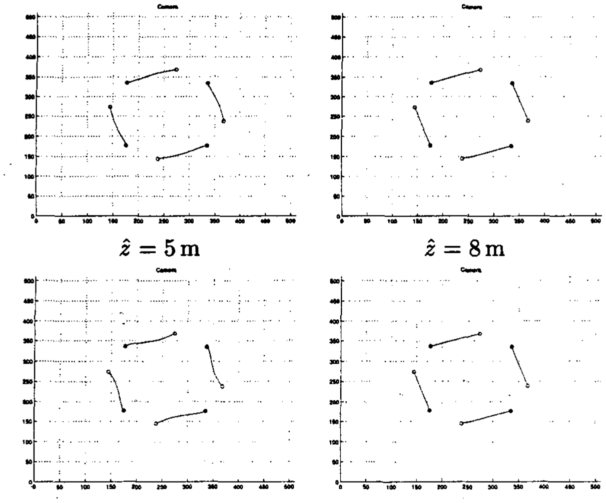

$\hat{z} = 5\,\text{m}$ $\hat{z} = 8\,\text{m}$
$\hat{z} = 15\,\text{m}$ $\hat{z} = z$

Figure 4: Performance of classical IBVS for pure rotation of 0.3 rad about Z-axis with different z estimates.

is computationally complex, though tractable in realtime, and requires coplanar feature points. We now describe our new approach, which does not exploit the epipolar geometry of the desired and initial images, and does not use any explicit 3D information.

5 A New Partitioned IBVS Scheme

Our approach is based on the observation that while IBVS works well for small motions, problems arise with large motions and particularly those involving rotation about the z axis. Our proposed partitioned scheme singles out just Z-axis translation and rotation for special treatment, unlike the hybrid approaches mentioned above which treat all three rotational degrees of freedom specially.

We partition the classical IBVS of (1) so that

$$\dot{f} = J_{xy}\dot{r}_{xy} + J_z\dot{r}_z \tag{8}$$

where $\dot{r}_{xy} = [Tx\ Ty\ \omega_x\ \omega_y]$, $\dot{r}_z = [T_z\ \omega_z]$, and J_{xy} and J_z are respectively columns $\{1, 2, 4, 5\}$ and $\{3, 6\}$ of J. Since $\dot{r}_z$ will be computed separately we can write (8) as

$$\dot{r}_{xy} = J_{xy}^{+}\left\{\dot{f} - J_z\dot{r}_z\right\} \tag{9}$$

where $\dot{f}$ is the feature point coordinate error as in the traditional IBVS scheme.

The Z-axis velocity, $\dot{r}_z$, is based directly on two new image features that are simple and computationally inexpensive to compute. The first image feature, $0 \leq \theta_{ij} < 2\pi$, is the angle between the u-axis of the image plane and the directed line segment joining feature points i and j, see Figure 5. For numerical conditioning it is advantageous to select the longest line segment. The rotational rate is simply

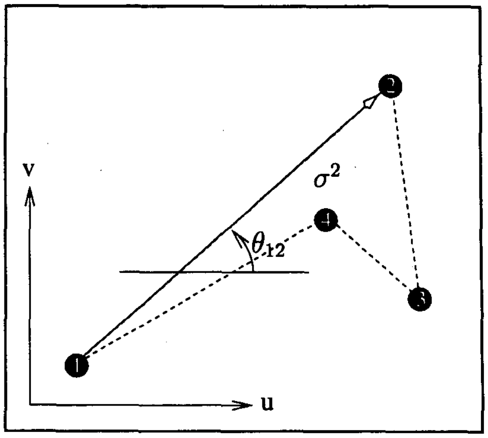


Figure 5: Image features for new partitioned IBVS control.

$$\omega_z = \gamma_{\omega_z}(\theta_{ij}^{*} - \theta_{ij})$$

in which γ_{ω_z} is a scalar gain coefficient. This form allows explicit control over the direction of rotation, which may be important to avoid mechanical motion limits.

The second new image feature that we use is a function of the area of the regular polygon whose vertices are the image feature points, see Figure 5. The advantages of this measure are that (1) it is a scalar; (2) it is rotation invariant thus decoupling camera rotation from Z-axis translation; (3) it can be cheaply computed. The feature that we choose to use is the square root of area

$$\sigma = \sqrt{m_{00}}$$

which has the dimension of length. The camera z-axis translation rate is thus given by

$$T_z = \gamma_{T_z}(\sigma^{*} - \sigma). \tag{10}$$

An example that involves complex translational and rotational motion is shown in Figure 6. The new features decrease monotonically, but the error in f does not decrease monotonically and the points follow complex curves on the image plane. Figure 7 compares the Cartesian camera motion for the two IBVS methods. The proposed partitioned method has eliminated the camera retreat and also exhibits better behavior for the X- and Y-axis motion. However the consequence is much more complex image plane feature motion that admits the possibility of the points leaving the field of view.

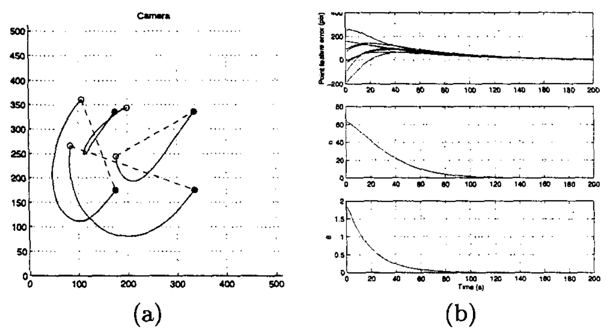
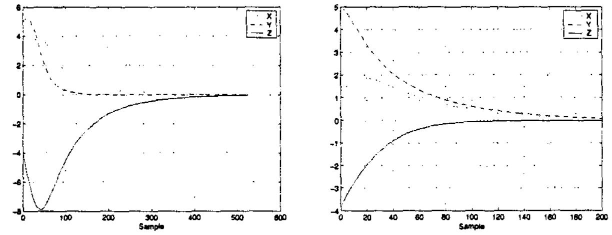


Figure 6: Proposed partitioned IBVS for general target motion. (a) Image-plane feature motion (dashed line shows straight line motion for classical IBVS), (b) Feature error trajectory.



IBVS

Proposed partitioned IBVS

Figure 7: Comparison of Cartesian camera motion for classic and new partitioned IBVS for general target motion.

6 Keeping Features in the Image Plane

In order to keep all feature points inside the viewable portion of the image plane at all times, we borrow collision avoidance techniques from the robot motion planning community. In particular, we establish a repulsive potential at the boundary of the viewable portion of the image, and incorporate the gradient of this potential into the control law. We use the simple potential given by

$$U_{rep}(u,v) = \begin{cases} \frac{1}{2}\eta\left(\frac{1}{\rho(u,v)} - \frac{1}{\rho_0}\right) & : \ \rho(u,v) \le \rho_0 \\ 0 & : \ \rho(u,v) > \rho_0 \end{cases} \tag{11}$$

in which $\rho(u,v)$ is the shortest distance to the edge of the image plane from the image point with coordinates (u,v). The value ρ_0 specifies the zone of the image in which U_{rep} affects the control; if the feature point is not with distance ρ_0 of the boundary, then the corresponding motion is not affected by U_{rep}. The value of η is a scalar gain coefficient.

For an $N_r \times N_c$ image, the value of ρ is easily computed as

$$\rho(u,v) = \min\{u, v, N_r - u, N_c - v\}. \tag{12}$$

If $\mathbf{n}$ is the unit vector directed from the nearest boundary to image feature point with coordinates (u,v), then $\nabla U_{rep} = F\mathbf{n}$, with F given by

$$F(u,v) = \begin{cases} \eta\left(\frac{1}{\rho(u,v)} - \frac{1}{\rho_0}\right)\frac{1}{\rho^2(u,v)} & : \ \rho(u,v) \le \rho_0 \\ 0 & : \ \rho(u,v) > \rho_0 \end{cases} \tag{13}$$

Since a pure translation in the negative z-direction will cause feature points to move toward the center of the image, the value of F is mapped directly to the T_z component of the velocity command by combining it with the control given in (10). Because of chatter effects (where the feature points oscillate in and out of the potential field), we smooth and clip the resulting T_z, yielding the discrete-time controller

$$\begin{aligned} T_z'(k) &= \mu T_z'(k-1) + (1-\mu)(\sigma^\star - \sigma - F) & (14) \\ T_z &= \min\{\max\{T_z'(k), T_{z_{min}}\}, T_{z_{max}}\}. & (15) \end{aligned}$$

In simulation we found it advantageous to use asymmetric velocity clipping where $|T_{z_{max}}| < |T_{z_{min}}|$, that is, the camera can retreat faster than it can approach the target. This reduces the magnitude of the "bounces" off the boundaries of the image plane when points first enter the potential field. In practice this smoothing and clipping may not need to be explicitly implemented, since the real robot will have finite bandwidth and velocity capability.

The use of a potential field raises the issue of local minima in the field, but in our case, these issues do not arise. The potential field is used merely to force a camera retreat, and since it will be possible for the system to achieve the goal when this retreat is effected (in this case we merely approach the performance of the classical IBVS system). Of course this assumes that no goal feature point locations lie within the influence of the potential field. Should this not be the case, then ρ_0 must be adjusted accordingly.

Results of the new partitioned IBVS with collision avoidance are shown in Figure 8. The target is larger than before, so that as the camera rotates the feature points move into the potential field.

7 Conclusion

We have investigated some problems with classical image-based visual servoing and proposed a new partitioned visual servoing scheme that inexpensively overcomes these limitations. We have also provided simple geometric insight into the root cause of the undesirable camera retreat phenomenon, and the pathological case we have termed IBVS failure.

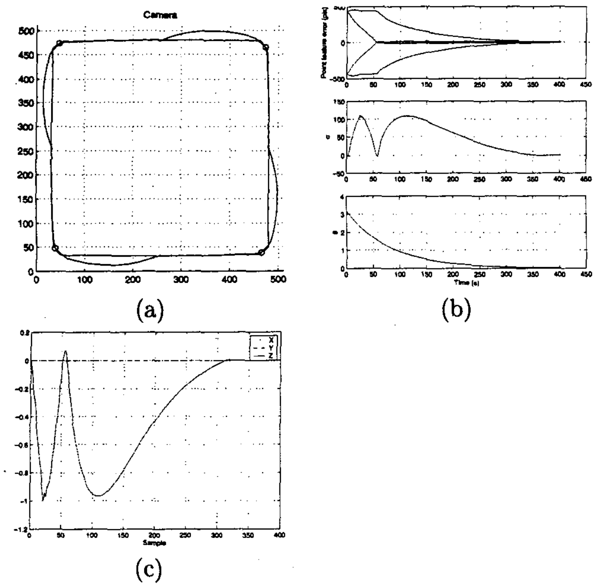


Figure 8: Proposed partitioned IBVS with collision avoidance for pure target rotation (π rad), $\eta = 5 \times 10^6$ and $\mu = 0.8$. (a) Image-plane feature motion (initial location is $\circ$, desired location is $\bullet$), (b) Feature error trajectory, (c) Cartesian translation trajectory.

Other hybrid IBVS schemes have been recently proposed and are based on decoupling camera translational and rotational degrees of freedom. We have proposed a different decoupling and servo Z-axis rotation and translation using decoupled controllers based on two easily computed image features.

All hybrid schemes admit the possibility of points leaving the image plane, as does the approach that we described in Section 5. In this paper we consider this as a collision avoidance problem and employ potential field techniques to repel the feature points from the image plane boundary.

References

[1] J. Aloimonos and D. P. Tsakiris. On the mathematics of visual tracking. *Image and Vision Computing*, 9(4):235–251, August 1991.

[2] F. Chaumette. Potential problems of stability and convergence in image-based and position-based visual servoing. In D. Kriegman, G. Hager, and S. Morse, editors, *The confluence of vision and control*, volume 237 of *Lecture Notes in Control and Information Sciences*, pages 66–78. Springer-Verlag, 1998.

[3] Francois Chaumette. *La relation vision-commande: théorie et application à des tâches robotiques.* PhD thesis, L'Univesité de Rennes I, 1990.

[4] K. Deguchi. Optimal motion control for image-based visual servoing by decoupling translation and rotation. In *Proc. Int. Conf. Intelligent Robots and Systems*, pages 705–711, October 1998.

[5] B. Espiau, F. Chaumette, and P. Rives. A New Approach to Visual Servoing in Robotics. *IEEE Trans. on Robotics and Automation*, 8:313–326, 1992.

[6] O.D. Faugeras. *Three-Dimensional Computer Vision.* MIT Press, Cambridge, MA, 1993.

[7] J.T. Feddema and O.R. Mitchell. Vision-guided servoing with feature-based trajectory generation. *IEEE Trans. Robot. Autom.*, 5(5):691–700, October 1989.

[8] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision.* Addison Wesley, 1993.

[9] Koh Hosoda and Minoru Asada. Versatile visual servoing without knowledge of true Jacobian. In *Proc. IROS*, September 1994.

[10] S. Hutchinson, G. Hager, and P. Corke. A tutorial on visual servo control. *IEEE Trans. Robot. Autom.*, 12(5):651–670, October 1996.

[11] Martin Jägersand, Olac Fuentes, and Randal Nelson. Experimental evaluation of uncalibrated visual servoing for precision manipulation. In *Proc. IEEE Int. Conf. Robotics and Automation*, 1996.

[12] E. Malis, F. Chaumette, and S. Boudet. 2-1/2-d visual servoing. *IEEE Trans. Robot. Autom.*, 15(2):238–250, April 1999.

[13] G. Morel, T. Liebezeit, J. Szewczyk, S. Boudet, and J. Pot. Explicit incorporation of 2d constraints in vision based control of robot manipulators. In Peter Corke and James Trevelyan, editors, *Experimental Robotics VI*, volume 250 of *Lecture Notes in Control and Information Sciences*, pages 99–108. Springer-Verlag, 2000. ISBN: 1 85233 210 7.

[14] N. P. Papanikolopoulos and P. K. Khosla. Adaptive Robot Visual Tracking: Theory and Experiments. *IEEE Trans. on Automatic Control*, 38(3):429–445, 1993.

[15] J.A. Piepmeier, G. McMurray, and H. Lipkin. A dynamic quasi-newton method for uncalibrated visual servoing. In *Proc. IEEE Int. Conf. Robotics and Automation*, pages 1595–1600, 1999.

[16] A. C. Sanderson, L. E. Weiss, and C. P. Neuman. Dynamic sensor-based control of robots with visual feedback. *IEEE Trans. Robot. Autom.*, RA-3(5):404–417, October 1987.