Visual servoing and visual tracking

François Chaumette, Seth Hutchinson

Handbook of Robotics

Chapter 24: Visual servoing and visual tracking

François Chaumette
Inria/Irisa Rennes
Campus de Beaulieu
F 35042 Rennes-cedex, France
e-mail: François.Chaumette@irisa.fr

Seth Hutchinson
The Beckman Institute
University of Illinois at Urbana-Champaign
405 North Matthews Avenue
Urbana, IL 61801
e-mail: seth@uiuc.edu

Chapter 24

Visual servoing and visual tracking

This chapter introduces visual servo control, using computer vision data in the servo loop to control the motion of a robot. We first describe the basic techniques that are by now well established in the field. We give a general overview of the formulation of the visual servo control problem, and describe the two archetypal visual servo control schemes: image-based and position-based visual servo control. We then discuss performance and stability issues that pertain to these two schemes, motivating advanced techniques. Of the many advanced techniques that have been developed, we discuss 2 1/2 D, hybrid, partitioned and switched approaches. Having covered a variety of control schemes, we conclude by turning briefly to the problems of target tracking and controlling motion directly in the joint space.

24.1 Introduction

Visual servo control refers to the use of computer vision data to control the motion of a robot. The vision data may be acquired from a camera that is mounted directly on a robot manipulator or on a mobile robot, in which case motion of the robot induces camera motion, or the camera can be fixed in the workspace so that it can observe the robot motion from a stationary configuration. Other configurations can be considered, such as for instance several cameras mounted on pan-tilt heads observing the robot motion. The mathematical development of all these cases is similar, and in this chapter we will focus primarily on the former, so-called *eye-in-hand*, case.

Visual servo control relies on techniques from image processing, computer vision, and control theory. In the present chapter, we will deal primarily with the issues from control theory, making connections to previous chapters when appropriate.

24.2 The Basic Components of Visual Servoing

The aim of all vision-based control schemes is to minimize an error $\mathbf{e}(t)$ which is typically defined by

$$\mathbf{e}(t) = \mathbf{s}(\mathbf{m}(t), \mathbf{a}) - \mathbf{s}^* \tag{24.1}$$

This formulation is quite general, and it encompasses a wide variety approaches, as we will see below. The parameters in (24.1) are defined as follows. The vector $\mathbf{m}(t)$ is a set of image measurements (e.g., the image coordinates of interest points, or the parameters of a set of image segments). These image measurements are used to compute a vector of k visual features, $\mathbf{s}(\mathbf{m}(t), \mathbf{a})$, in which $\mathbf{a}$ is a set of parameters that represent potential additional knowledge about the system (e.g., coarse camera intrinsic parameters or 3D model of objects). The vector $\mathbf{s}^*$ contains the desired values of the features.

For now, we consider the case of a fixed goal pose and a motionless target, i.e., $\mathbf{s}^*$ is constant, and changes in $\mathbf{s}$ depend only on camera motion. Further, we consider here the case of controlling the motion of a camera with six degrees of freedom (e.g., a

camera attached to the end effector of a six degree-of-freedom arm). We will treat more general cases in later sections.

Visual servoing schemes mainly differ in the way that $\mathbf{s}$ is designed. In Sections 24.3 and 24.4, we describe classical approaches, including image-based visual servo control (IBVS), in which $\mathbf{s}$ consists of a set of features that are immediately available in the image data, and position-based visual servo control (PBVS), in which $\mathbf{s}$ consists of a set of 3D parameters, which must be estimated from image measurements. We also present in Section 24.5 several more advanced methods.

Once $\mathbf{s}$ is selected, the design of the control scheme can be quite simple. Perhaps the most straightforward approach is to design a velocity controller. To do this, we require the relationship between the time variation of $\mathbf{s}$ and the camera velocity. Let the spatial velocity of the camera be denoted by $\mathbf{v}_c = (\boldsymbol{v}_c, \boldsymbol{\omega}_c)$, with $\boldsymbol{v}_c$ the instantaneous linear velocity of the origin of the camera frame and $\boldsymbol{\omega}_c$ the instantaneous angular velocity of the camera frame The relationship between $\dot{\mathbf{s}}$ and $\mathbf{v}_c$ is given by

$$\dot{\mathbf{s}} = \mathbf{L}_\mathbf{s}\mathbf{v}_c \tag{24.2}$$

in which $\mathbf{L}_\mathbf{s} \in \mathbb{R}^{k\times 6}$ is named the *interaction matrix* related to $\mathbf{s}$ [1]. The term *feature Jacobian* is also used somewhat interchangeably in the visual servo literature [2], but in the present chapter we will use this last term to relate the time variation of the features to the joints velocity (see Section 24.9).

Using (24.1) and (24.2) we immediately obtain the relationship between camera velocity and the time variation of the error:

$$\dot{\mathbf{e}} = \mathbf{L}_\mathbf{e}\mathbf{v}_c \tag{24.3}$$

where $\mathbf{L}_\mathbf{e} = \mathbf{L}_\mathbf{s}$. Considering $\mathbf{v}_c$ as the input to the robot controller, and if we would like for instance to try to ensure an exponential decoupled decrease of the error (i.e., $\dot{\mathbf{e}} = -\lambda\mathbf{e}$), we obtain using (24.3):

$$\mathbf{v}_c = -\lambda\mathbf{L}_\mathbf{e}^+\mathbf{e} \tag{24.4}$$

where $\mathbf{L}_\mathbf{e}^+ \in \mathbb{R}^{6\times k}$ is chosen as the Moore-Penrose pseudo-inverse of $\mathbf{L}_\mathbf{e}$, that is $\mathbf{L}_\mathbf{e}^+ = (\mathbf{L}_\mathbf{e}^\top\mathbf{L}_\mathbf{e})^{-1}\mathbf{L}_\mathbf{e}^\top$ when $\mathbf{L}_\mathbf{e}$ is of full rank 6. This choice allows $\|\dot{\mathbf{e}} - \lambda\mathbf{L}_\mathbf{e}\mathbf{L}_\mathbf{e}^+\mathbf{e}\|$ and $\|\mathbf{v}_c\|$ to be minimal. When $k = 6$, if $\det\mathbf{L}_\mathbf{e} \neq 0$ it it possible to invert $\mathbf{L}_\mathbf{e}$, giving the control $\mathbf{v}_c = -\lambda\mathbf{L}_\mathbf{e}^{-1}\mathbf{e}$.

In real visual servo systems, it is impossible to know perfectly in practice either $\mathbf{L}_\mathbf{e}$ or $\mathbf{L}_\mathbf{e}^+$. So an approximation or an estimation of one of these two matrices must be realized. In the sequel, we denote both the pseudo-inverse of the approximation of the interaction matrix and the approximation of the pseudo-inverse of the interaction matrix by the symbol $\widehat{\mathbf{L}_\mathbf{e}^+}$. Using this notation, the control law is in fact:

$$\mathbf{v}_c = -\lambda\widehat{\mathbf{L}_\mathbf{e}^+}\mathbf{e} = -\lambda\widehat{\mathbf{L}_\mathbf{s}^+}(\mathbf{s} - \mathbf{s}^*) \tag{24.5}$$

Closing the loop and assuming the robot controller is able to realize perfectly $\mathbf{v}_c$ (that is injecting (24.5) in (24.3), we obtain:

$$\dot{\mathbf{e}} = -\lambda\mathbf{L}_\mathbf{e}\widehat{\mathbf{L}_\mathbf{e}^+}\mathbf{e} \tag{24.6}$$

This last equation characterizes the real behavior of the closed loop system, which is different from the specified one ($\dot{\mathbf{e}} = -\lambda\mathbf{e}$) as soon as $\mathbf{L}_\mathbf{e}\widehat{\mathbf{L}_\mathbf{e}^+} \neq \mathbf{I}_6$, where $\mathbf{I}_6$ is the 6×6 identity matrix. It is also at the basis of the stability analysis of the system using Lyapunov theory.

What we have presented above is the basic design implemented by most visual servo controllers. All that remains is to fill in the details: How should $\mathbf{s}$ be chosen? What then is the form of $\mathbf{L}_\mathbf{s}$? How should we estimate $\widehat{\mathbf{L}_\mathbf{e}^+}$? What are the performance characteristics of the resulting closed-loop system? These questions are addressed in the remainder of the chapter. We first describe the two basic approaches, IBVS and PBVS, whose principles have been proposed more than twenty years ago [3]. We then present more recent approaches which have improved their performances.

24.3 Image-Based Visual Servo

Traditional image-based control schemes [4], [3] use the image-plane coordinates of a set of points to define the set $\mathbf{s}$. The image measurements $\mathbf{m}$ are usually the pixel coordinates of the set of image points

(but this is not the only possible choice), and the parameters $\mathbf{a}$ in the definition of $\mathbf{s} = \mathbf{s}(\mathbf{m}, \mathbf{a})$ in (24.1) are nothing but the camera intrinsic parameters to go from image measurements expressed in pixels to the features.

24.3.1 The Interaction Matrix

For a 3D point with coordinates $\mathbf{X} = (X, Y, Z)$ in the camera frame which projects in the image as a 2D point with coordinates $\mathbf{x} = (x, y)$, we have:

$$\left\{ \begin{array}{rcl} x & = & X/Z = (u - c_u)/f\alpha \\ y & = & Y/Z = (v - c_v)/f \end{array} \right. \tag{24.7}$$

where $\mathbf{m} = (u, v)$ gives the coordinates of the image point expressed in pixel units, and $\mathbf{a} = (c_u, c_v, f, \alpha)$ is the set of camera intrinsic parameters as defined in Chapter 23: c_u and c_v are the coordinates of the principal point, f is the focal length and α is the ratio of the pixel dimensions. The intrinsic parameter β defined in Chapter 23 has been assumed to be 0 here. In this case, we take $\mathbf{s} = \mathbf{x} = (x, y)$, the image plane coordinates of the point. The details of imaging geometry and perspective projection can be found in many computer vision texts, including [5], [6].

Taking the time derivative of the projection equations (24.7), we obtain

$$\left\{ \begin{array}{rcccl} \dot{x} & = & \dot{X}/Z - X\dot{Z}/Z^2 & = & (\dot{X} - x\dot{Z})/Z \\ \dot{y} & = & \dot{Y}/Z - Y\dot{Z}/Z^2 & = & (\dot{Y} - y\dot{Z})/Z \end{array} \right. \tag{24.8}$$

We can relate the velocity of the 3D point to the camera spatial velocity using the well known equation

$$\dot{\mathbf{X}} = -\boldsymbol{v}_c - \boldsymbol{\omega}_c \times \mathbf{X} \Leftrightarrow \left\{ \begin{array}{l} \dot{X} = -v_x - \omega_y Z + \omega_z Y \\ \dot{Y} = -v_y - \omega_z X + \omega_x Z \\ \dot{Z} = -v_z - \omega_x Y + \omega_y X \end{array} \right. \tag{24.9}$$

Injecting (24.9) in (24.8), grouping terms and using (24.7), we obtain

$$\left\{ \begin{array}{l} \dot{x} = -v_x/Z + xv_z/Z + xy\omega_x - (1 + x^2)\omega_y + y\omega_z \\ \dot{y} = -v_y/Z + yv_z/Z + (1 + y^2)\omega_x - xy\omega_y - x\omega_z \end{array} \right. \tag{24.10}$$

which can be written

$$\dot{\mathbf{x}} = \mathbf{L}_\mathbf{x}\mathbf{v}_c \tag{24.11}$$

where the interaction matrix $\mathbf{L}_\mathbf{x}$ is given by

$$\mathbf{L}_\mathbf{x} = \left[\begin{array}{cccccc} -1/Z & 0 & x/Z & xy & -(1 + x^2) & y \\ 0 & -1/Z & y/Z & 1 + y^2 & -xy & -x \end{array} \right] \tag{24.12}$$

In the matrix $\mathbf{L}_\mathbf{x}$, the value Z is the depth of the point relative to the camera frame. Therefore, any control scheme that uses this form of the interaction matrix must estimate or approximate the value of Z. Similarly, the camera intrinsic parameters are involved in the computation of x and y. Thus $\mathbf{L}_\mathbf{x}^+$ can not be *directly* used in (24.4), and an estimation or an approximation $\widehat{\mathbf{L}_\mathbf{x}^+}$ must be used, as in (24.5). We discuss this in more detail below.

To control the six degrees of freedom, at least three points are necessary (i.e., we require $k \geq 6$). If we use the feature vector $\mathbf{x} = (\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3})$, by merely stacking interaction matrices for three points we obtain

$$\mathbf{L}_\mathbf{x} = \left[\begin{array}{c} \mathbf{L}_{\mathbf{x}_1} \\ \mathbf{L}_{\mathbf{x}_2} \\ \mathbf{L}_{\mathbf{x}_3} \end{array} \right]$$

In this case, there will exist some configurations for which $\mathbf{L}_\mathbf{x}$ is singular [7]. Furthermore, there exist four distinct camera poses for which $\mathbf{e} = 0$, i.e., four global minima exist, and it is impossible to differentiate them [8]. For these reasons, more than three points are usually considered.

24.3.2 Approximating the Interaction Matrix

There are several choices available for constructing the estimate $\widehat{\mathbf{L}_\mathbf{e}^+}$ to be used in the control law. One popular scheme is of course to choose $\widehat{\mathbf{L}_\mathbf{e}^+} = \mathbf{L}_\mathbf{e}^+$ if $\mathbf{L}_\mathbf{e} = \mathbf{L}_\mathbf{x}$ is known, that is if the current depth Z of each point is available [2]. In practice, these parameters must be estimated at each iteration of the control scheme. The basic IBVS methods use classical pose estimation methods (see Chapter 23 and the beginning of Section 24.4). Another popular approach is to choose $\widehat{\mathbf{L}_\mathbf{e}^+} = \mathbf{L}_{\mathbf{e}^*}^+$ where $\mathbf{L}_{\mathbf{e}^*}$ is the value of $\mathbf{L}_\mathbf{e}$ for the desired position $\mathbf{e} = \mathbf{e}^* = 0$ [1]. In this case, $\widehat{\mathbf{L}_\mathbf{e}^+}$ is constant, and only the desired depth of each point has to be set, which means no varying 3D

parameters have to be estimated during the visual servo. Finally, the choice $\widehat{\mathbf{L}_\mathbf{e}^+} = (\mathbf{L}_\mathbf{e}/2 + \mathbf{L}_{\mathbf{e}^*}/2)^+$ has recently been proposed in [9]. Since $\mathbf{L}_\mathbf{e}$ is involved in this method, the current depth of each point has also to be available.

We illustrate the behavior of these control schemes with an example. The goal is to position the camera so that it observes a square as a centered square in the image (see Figure 24.1). We define $\mathbf{s}$ to include the x and y coordinates of the four points forming the square. Note that the initial camera pose has been selected far away from the desired pose, particularly with regard to the rotational motions, which are known to be the most problematic for IBVS. In the simulations presented in the following, no noise or modeling errors have been introduced in order to allow comparison of different behaviors in perfect conditions.

The results obtained by using $\widehat{\mathbf{L}_\mathbf{e}^+} = \mathbf{L}_{\mathbf{e}^*}^+$ are given in Figure 24.2. Note that despite the large displacement that is required the system converges. However, neither the behavior in the image, nor the computed camera velocity components, nor the 3D trajectory of the camera present desirable properties far from the convergence (i.e., for the first thirty or so iterations).

The results obtained using $\widehat{\mathbf{L}_\mathbf{e}^+} = \mathbf{L}_\mathbf{e}^+$ are given in Figure 24.3. In this case, the trajectories of the points in the image are almost straight lines, but the behavior induced in 3D is even less satisfactory than for the case of $\widehat{\mathbf{L}_\mathbf{e}^+} = \mathbf{L}_{\mathbf{e}^*}^+$ The large camera velocities at the beginning of the servo indicate that the condition number of $\widehat{\mathbf{L}_\mathbf{e}^+}$ is high at the start of the trajectory, and the camera trajectory is far from a straight line.

The choice $\widehat{\mathbf{L}_\mathbf{e}^+} = (\mathbf{L}_\mathbf{e}/2 + \mathbf{L}_{\mathbf{e}^*}/2)^+$ provides good performance in practice. Indeed, as can be seen in Figure 24.4, the camera velocity components do not include large oscillations, and provide a smooth trajectory in both the image and in 3D.

24.3.3 A Geometrical Interpretation of IBVS

It is quite easy to provide a geometric interpretation of the behavior of the control schemes defined above. The example illustrated in Figure 24.5 corresponds to a pure rotation around the optical axis from the initial configuration (shown in blue) to the desired configuration of four coplanar points parallel to the image plane (shown in red).

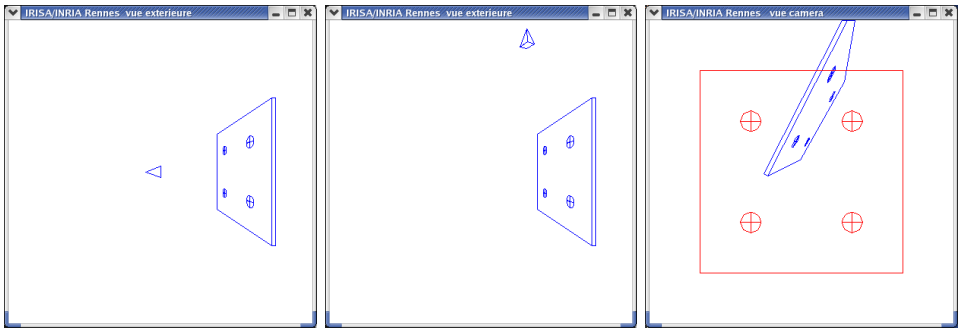
As explained above, using $\mathbf{L}_\mathbf{e}^+$ in the control scheme attempts to ensure an exponential decrease of the error $\mathbf{e}$. It means that when x and y image point coordinates compose this error, the points' trajectories in the image follow straight lines from their initial to their desired positions, when it is possible. This leads to the image motion plotted in green in the figure. The camera motion to realize this image motion can be easily deduced and is indeed composed of a rotational motion around the optical axis, but combined with a retreating translational motion along the optical axis [10]. This unexpected motion is due to the choice of the features, and to the coupling between the third and sixth columns in the interaction matrix. If the rotation between the initial and desired configurations is very large, this phenomenon is amplified, and leads to a particular case for a rotation of π rad where no rotational motion at all will be induced by the control scheme [11]. On the other hand, when the rotation is small, this phenomenon almost disappears. To conclude, the behavior is locally satisfactory (i.e., when the error is small), but it can be unsatisfactory when the error is large. As we will see below, these results are consistent with the local asymptotic stability results that can be obtained for IBVS.

If instead we use $\mathbf{L}_{\mathbf{e}^*}^+$ in the control scheme, the image motion generated can easily be shown to be the blue one plotted in Figure 24.5. Indeed, if we consider the same control scheme as before but starting from $\mathbf{s}^*$ to reach $\mathbf{s}$, we obtain:

$$\mathbf{v}_c = -\lambda \mathbf{L}_{\mathbf{e}^*}^+ (\mathbf{s}^* - \mathbf{s})$$

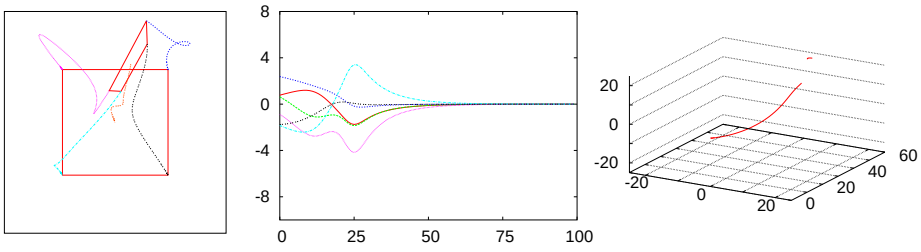
which again induces straight-line trajectories from the red points to the blue ones, causing image motion plotted in brown. Going back to our problem, the control scheme computes a camera velocity that is exactly the opposite one:

$$\mathbf{v}_c = -\lambda \mathbf{L}_{\mathbf{e}^*}^+ (\mathbf{s} - \mathbf{s}^*)$$



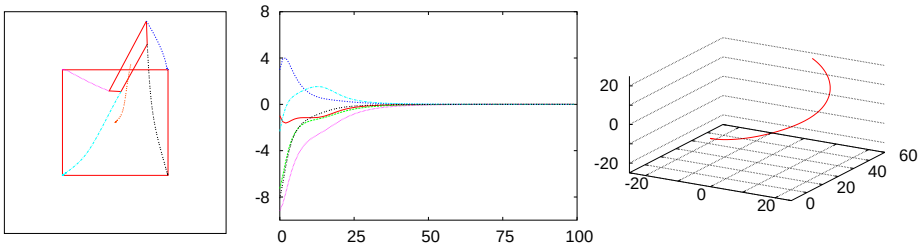
(a) (b) (c)

Figure 24.1: Example of positioning task: (a) the desired camera pose w.r.t. a simple target, (b) the initial camera pose, (c) the corresponding initial and desired image of the target.



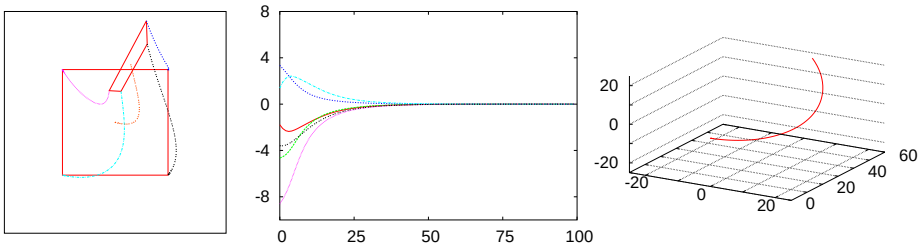
(a) (b) (c)

Figure 24.2: System behavior using $\mathbf{s} = (x_1, y_1, ..., x_4, y_4)$ and $\widehat{\mathbf{L}_\mathbf{e}^+} = \mathbf{L}_{\mathbf{e}^*}^+$: (a) image points trajectories including the trajectory of the center of the square which is not used in the control scheme, (b) $\mathbf{v}_c$ components (cm/s and dg/s) computed at each iteration of the control scheme, (c) 3D trajectory of the camera optical center expressed in $\mathcal{R}_{c^*}$ (cm).



(a) (b) (c)

Figure 24.3: System behavior using $\mathbf{s} = (x_1, y_1, ..., x_4, y_4)$ and $\widehat{\mathbf{L}_\mathbf{e}^+} = \mathbf{L}_\mathbf{e}^+$.



(a) (b) (c)

Figure 24.4: System behavior using $\mathbf{s} = (x_1, y_1, ..., x_4, y_4)$ and $\widehat{\mathbf{L}_\mathbf{e}^+} = (\mathbf{L}_\mathbf{e}/2 + \mathbf{L}_{\mathbf{e}^*}/2)^+$.

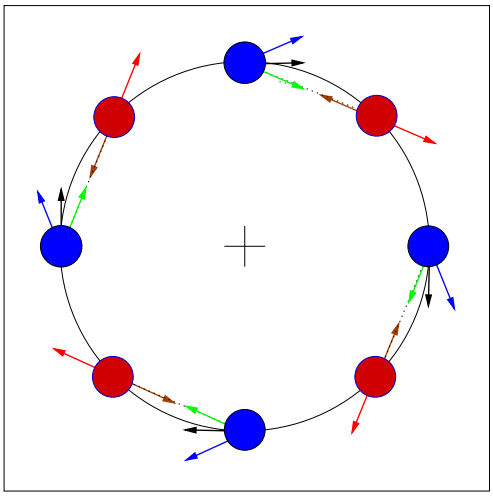


Figure 24.5: Geometrical interpretation of IBVS: going from the blue position to the red one. In red, image motion when $\mathbf{L}_\mathbf{e}^+$ is used in the control scheme; in blue, when $\mathbf{L}_{\mathbf{e}^*}^+$ is used; and in black when $(\mathbf{L}_\mathbf{e}/2 + \mathbf{L}_{\mathbf{e}^*}/2)^+$ is used (see text more more details).

and thus generates the image motion plotted in red at the red points. Transformed at the blue points, the camera velocity generates the blue image motion and corresponds once again to a rotational motion around the optical axis, combined now with an unexpected forward motion along the optical axis. The same analysis as before can be done, as for large or small errors. We can add that, as soon as the error decreases significantly, both control schemes get closer, and tend to the same one (since $\mathbf{L}_\mathbf{e} = \mathbf{L}_{\mathbf{e}^*}$ when $\mathbf{e} = \mathbf{e}^*$) with a nice behavior characterized with the image motion plotted in black and a camera motion composed of only a rotation around the optical axis when the error tends towards zero.

If we instead use $\widehat{\mathbf{L}_\mathbf{e}^+} = (\mathbf{L}_\mathbf{e}/2 + \mathbf{L}_{\mathbf{e}^*}/2)^+$, it is intuitively clear that considering the mean of $\mathbf{L}_\mathbf{e}$ and $\mathbf{L}_{\mathbf{e}^*}$ generates the image motion plotted in black, even when the error is large. In all cases but the rotation around π rad, the camera motion is now a pure rotation around the optical axis, without any unexpected translational motion.

24.3.4 Stability Analysis

We now consider the fundamental issues related to the stability of IBVS. To assess the stability of the closed-loop visual servo systems, we will use Lyapunov analysis. In particular, consider the candidate Lyapunov function defined by the squared error norm $\mathcal{L} = \frac{1}{2}\|\mathbf{e}(t)\|^2$, whose derivative is given by

$$\begin{aligned} \dot{\mathcal{L}} &= \mathbf{e}^\top \dot{\mathbf{e}} \\ &= -\lambda \mathbf{e}^\top \mathbf{L}_\mathbf{e} \widehat{\mathbf{L}_\mathbf{e}^+} \mathbf{e} \end{aligned}$$

since $\dot{\mathbf{e}}$ is given by (24.6). The global asymptotic stability of the system is thus obtained when the following sufficient condition is ensured:

$$\mathbf{L_e}\widehat{\mathbf{L_e}^+} > 0 \tag{24.13}$$

If the number of features is equal to the number of camera degrees of freedom (i.e., $k = 6$), and if the features are chosen and the control scheme designed so that $\mathbf{L_e}$ and $\widehat{\mathbf{L_e}^+}$ are of full rank 6, then condition (24.13) is ensured if the approximations involved in $\widehat{\mathbf{L_e}^+}$ are not too coarse.

As discussed above, for most IBVS approaches we have $k > 6$. Therefore condition (24.13) can never be ensured since $\mathbf{L_e}\widehat{\mathbf{L_e}^+} \in \mathbb{R}^{k\times k}$ is at most of rank 6, and thus $\mathbf{L_e}\widehat{\mathbf{L_e}^+}$ has a nontrivial null space. In this case, configurations such that $\mathbf{e} \in \text{Ker}\widehat{\mathbf{L_e}^+}$ correspond to local minima. Reaching such a local minimum is illustrated in Figure 24.6. As can be seen in Figure 24.6.d, each component of $\mathbf{e}$ has a nice exponential decrease with the same convergence speed, causing straight-line trajectories to be realized in the image, but the error reached is not exactly zero, and it is clear from Figure 24.6.c that the system has been attracted to a local minimum far away from the desired configuration. Thus, only local asymptotic stability can be obtained for IBVS.

To study local asymptotic stability when $k > 6$, let us first define a new error $\mathbf{e}'$ with $\mathbf{e}' = \widehat{\mathbf{L_e}^+}\mathbf{e}$. The time derivative of this error is given by

$$\begin{aligned}\dot{\mathbf{e}}' &= \widehat{\mathbf{L_e}^+}\dot{\mathbf{e}} + \dot{\widehat{\mathbf{L_e}^+}}\mathbf{e} \\ &= (\widehat{\mathbf{L_e}^+}\mathbf{L_e} + \mathbf{O})\mathbf{v}_c\end{aligned}$$

where $\mathbf{O} \in \mathbb{R}^{6\times 6}$ is equal to 0 when $\mathbf{e} = 0$, whatever the choice of $\widehat{\mathbf{L_e}^+}$ [12]. Using the control scheme (24.5), we obtain:

$$\dot{\mathbf{e}}' = -\lambda(\widehat{\mathbf{L_e}^+}\mathbf{L_e} + \mathbf{O})\mathbf{e}'$$

which is known to be locally asymptotically stable in a neighborhood of $\mathbf{e} = \mathbf{e}^* = 0$ if

$$\widehat{\mathbf{L_e}^+}\mathbf{L_e} > 0 \tag{24.14}$$

where $\widehat{\mathbf{L_e}^+}\mathbf{L_e} \in \mathbb{R}^{6\times 6}$. Indeed, only the linearized system $\dot{\mathbf{e}}' = -\lambda\widehat{\mathbf{L_e}^+}\mathbf{L_e}\mathbf{e}'$ has to be considered if we are interested in the local asymptotic stability [13].

Once again, if the features are chosen and the control scheme designed so that $\mathbf{L_e}$ and $\widehat{\mathbf{L_e}^+}$ are of full rank 6, then condition (24.14) is ensured if the approximations involved in $\widehat{\mathbf{L_e}^+}$ are not too coarse.

To end the demonstration of local asymptotic stability, we must show that there does not exist any configuration $\mathbf{e} \neq \mathbf{e}^*$ such that $\mathbf{e} \in \text{Ker}\ \widehat{\mathbf{L_e}^+}$ in a small neighborhood of $\mathbf{e}^*$ and in a small neighborhood of the corresponding pose $\mathbf{p}^*$. Such configurations correspond to local minima where $\mathbf{v}_c = 0$ and $\mathbf{e} \neq \mathbf{e}^*$. If such a pose $\mathbf{p}$ would exist, it is possible to restrict the neighborhood around $\mathbf{p}^*$ so that there exists a camera velocity $\mathbf{v}$ to reach $\mathbf{p}^*$ from $\mathbf{p}$. This camera velocity would imply a variation of the error $\dot{\mathbf{e}} = \mathbf{L_e}\mathbf{v}$. However, such a variation can not belong to Ker $\widehat{\mathbf{L_e}^+}$ since $\widehat{\mathbf{L_e}^+}\mathbf{L_e} > 0$. Therefore, we have $\mathbf{v}_c = 0$ if and only if $\dot{\mathbf{e}} = 0$, i.e., $\mathbf{e} = \mathbf{e}^*$, in a neighborhood of $\mathbf{p}^*$.

Even though local asymptotic stability can be ensured when $k > 6$, we recall that global asymptotic stability cannot be ensured. For instance, as illustrated in Figure 24.6, there may exist local minima corresponding to configurations where $\mathbf{e} \in$ Ker $\widehat{\mathbf{L_e}^+}$, which are outside of the neighborhood considered above. Determining the size of the neighborhood where the stability and the convergence are ensured is still an open issue, even if this neighborhood is surprisingly quite large in practice.

24.3.5 IBVS with a stereo-vision system

It is straightforward to extend the IBVS approach to a multi-camera system. If a stereo-vision system is used, and a 3D point is visible in both left and right images (see Figure 24.7), it is possible to use as visual features

$$\mathbf{s} = \mathbf{x}_s = (\mathbf{x}_l, \mathbf{x}_r) = (x_l, y_l, x_r, y_r)$$

i.e., to represent the point by just stacking in $\mathbf{s}$ the x and y coordinates of the observed point in the left

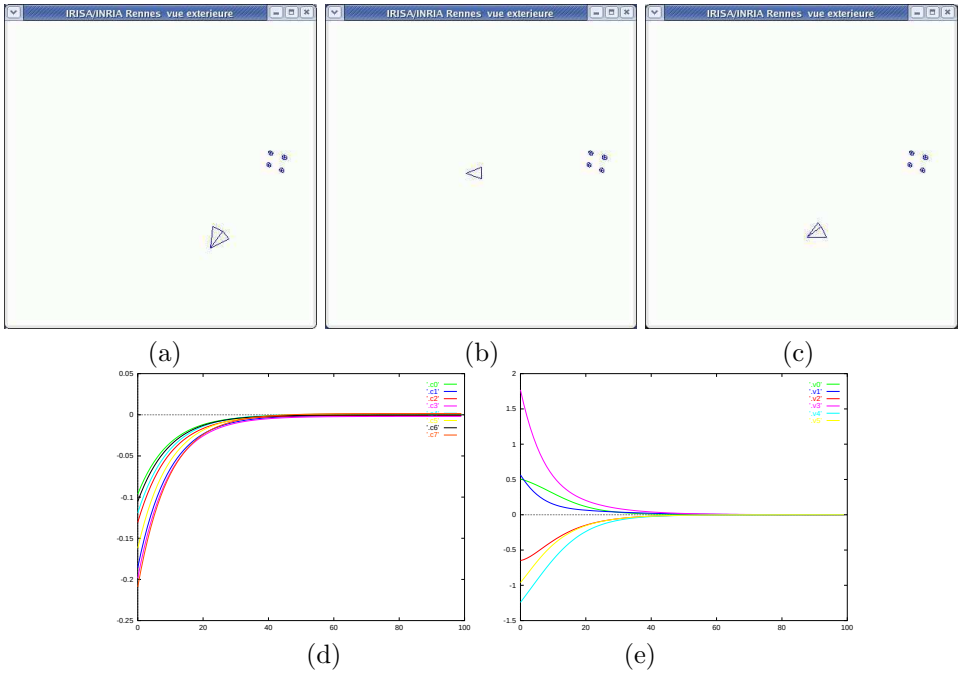


Figure 24.6: Reaching a local minimum using $\mathbf{s} = (x_1, y_1, ..., x_4, y_4)$and $\widehat{\mathbf{L}_\mathbf{e}^+} = \mathbf{L}_\mathbf{e}^+$: (a) initial configuration, (b) desired one, (c) configuration reached after the convergence of the control scheme, (d) evolution of the error $\mathbf{e}$ at each iteration of the control scheme, (f) evolution of the six components of the camera velocity $\mathbf{v}_c$.

and right images [14]. However, care must be taken when constructing the corresponding interaction matrix since the form given in (24.11) is expressed in either the left or right camera frame. More precisely, we have:

$$\left\{ \begin{array}{rcl} \dot{\mathbf{x}}_l & = & \mathbf{L}_{\mathbf{x}_l}\, \mathbf{v}_l \\ \dot{\mathbf{x}}_r & = & \mathbf{L}_{\mathbf{x}_r}\, \mathbf{v}_r \end{array} \right.$$

where $\mathbf{v}_l$ and $\mathbf{v}_r$ are the spatial velocity of the left and right camera respectively, and where the analytical form of $\mathbf{L}_{\mathbf{x}_l}$ and $\mathbf{L}_{\mathbf{x}_r}$ are given by (24.12).

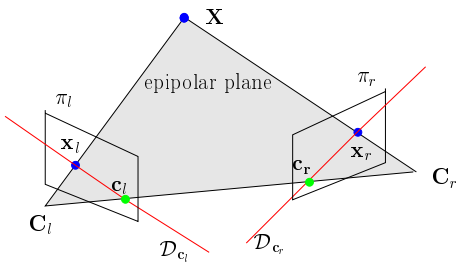


Figure 24.7: Stereo-vision system

By choosing a sensor frame rigidly linked to the stereo-vision system, we obtain:

$$\dot{\mathbf{x}}_s = \left[\begin{array}{c} \dot{\mathbf{x}}_l \\ \dot{\mathbf{x}}_r \end{array} \right] = \mathbf{L}_{\mathbf{x}_s} \mathbf{v}_s$$

where the interaction matrix related to $\mathbf{x}_s$ can be determined using the spatial motion transform matrix $\mathbf{V}$ defined in Chapter 1 to transform velocities expressed in the left or right cameras frames to the sensor frame. We recall that $\mathbf{V}$ is given by

$$\mathbf{V} = \left[\begin{array}{cc} \mathbf{R} & [\mathbf{t}]_\times \mathbf{R} \\ \mathbf{0} & \mathbf{R} \end{array} \right] \tag{24.15}$$

where $[\mathbf{t}]_\times$ is the skew symmetric matrix associated to the vector $\mathbf{t}$ and where $(\mathbf{R}, \mathbf{t}) \in SE(3)$ is the rigid body transformation from camera to sensor frame. The numerical values for these matrices are directly obtained from the calibration step of the stereo-vision system. Using this equation, we obtain

$$\mathbf{L}_{\mathbf{x}_s} = \left[\begin{array}{c} \mathbf{L}_{\mathbf{x}_l}{}^l\mathbf{V}_s \\ \mathbf{L}_{\mathbf{x}_r}{}^r\mathbf{V}_s \end{array} \right]$$

Note that $\mathbf{L}_{\mathbf{x}_s} \in \mathbb{R}^{4\times 6}$ is always of rank 3 because of the epipolar constraint that links the perspective projection of a 3D point in a stereo-vision system (see Figure 24.7). Another simple interpretation is that a 3D point is represented by three independent parameters, which makes it impossible to find more than three independent parameters using any sensor observing that point.

To control the six degrees of freedom of the system, it is necessary to consider at least three points, the rank of the interaction matrix by considering only two points being equal to 5.

Using a stereo-vision system, since the 3D coordinates of any point observed in both images can be easily estimated by a simple triangulation process it is possible and quite natural to use these 3D coordinates in the features set $\mathbf{s}$. Such an approach would be, strictly speaking, a position-based approach, since it would require 3D parameters in $\mathbf{s}$.

24.3.6 IBVS with cylindrical coordinates of image points

In the previous sections, we have considered the Cartesian coordinates of images points. As proposed in [15], it may be interesting to use the cylindrical coordinates (ρ, θ) of the image points instead of their Cartesian coordinates (x, y). They are given by:

$$\rho = \sqrt{x^2 + y^2}\ ,\ \theta = \arctan \frac{y}{x}$$

from which we deduce:

$$\dot{\rho} = (x\dot{x} + y\dot{y})/\rho\ ,\ \dot{\theta} = (x\dot{y} - y\dot{x})/\rho^2$$

Using (24.11) and then subsituting x by $\rho\cos\theta$ and y by $\rho\sin\theta$, we obtain immediately:

$$\begin{array}{rcl} \mathbf{L}_\rho & = & \left[\begin{array}{cccccc} \frac{-c}{Z} & \frac{-s}{Z} & \frac{\rho}{Z} & (1+\rho^2)s & -(1+\rho^2)c & 0 \end{array} \right] \\ \mathbf{L}_\theta & = & \left[\begin{array}{cccccc} \frac{s}{\rho Z} & \frac{-c}{\rho Z} & 0 & \frac{c}{\rho} & \frac{s}{\rho} & -1 \end{array} \right] \end{array}$$

where $c = \cos\theta$ and $s = \sin\theta$.

The behavior obtained in that case is quite satisfactory. Indeed, by considering the same example as before, and using now $\mathbf{s} = (\rho_1, \theta_1, ..., \rho_4, \theta_4)$ and $\widehat{\mathbf{L}_\mathbf{e}^+} = \mathbf{L}_{\mathbf{e}^*}^+$ (that is a constant matrix), the system

behavior shown on Figure 24.8 has the same nice properties than using $\widehat{\mathbf{L}_\mathbf{e}^+} = (\mathbf{L}_\mathbf{e}/2 + \mathbf{L}_{\mathbf{e}^*}/2)^+$ with $\mathbf{s} = (x_1, y_1, ..., x_4, y_4)$. If we go back to the example depicted in Figure 24.5, the behavior obtained will be also the expected one thanks to the decoupling between the third and sixth columns of the interaction matrix.

24.3.7 IBVS with other geometrical features

In the previous sections, we have only considered image points coordinates in $\mathbf{s}$. Other geometrical primitives can of course be used. There are several reasons to do so. First, the scene observed by the camera cannot always be described merely by a collection of points, in which case the image processing provides other types of measurements, such as a set of straight lines or the contours of an object. Second, richer geometric primitives may ameliorate the decoupling and linearizing issues that motivate the design of partitioned systems. Finally, the robotic task to be achieved may be expressed in terms of virtual linkages (or fixtures) between the camera and the observed objects [16], [17], sometimes expressed directly by constraints between primitives, such as for instance point-to-line [18] (which means that an observed point must lie on a specified line).

For the first point, it is possible to determine the interaction matrix related to the perspective projection of a large class of geometrical primitives, such as segments, straight lines, spheres, circles, and cylinders. The results are given in [1] and [16]. Recently, the analytical form of the interaction matrix related to any image moments corresponding to planar objects has been computed. This makes possible to consider planar objects of any shape [19]. If a collection of points is measured in the image, moments can also be used [20]. In both cases, moments allow the use of intuitive geometrical features, such as the center of gravity or the orientation of an object. By selecting an adequate combination of moments, it is then possible to determine partitioned systems with good decoupling and linearizing properties [19], [20].

Note that for all these features (geometrical primitives, moments), the depth of the primitive or of the object considered appears in the coefficients of the interaction matrix related to the translational degrees of freedom, as was the case for the image points. An estimation of this depth is thus generally still necessary. Few exceptions occur, using for instance an adequate normalization of moments, which allows in some particular cases to make only the constant desired depth appear in the interaction matrix [20].

24.3.8 Direct estimation

In the previous sections, we have focused on the analytical form of the interaction matrix. It is also possible to estimate directly its numerical value using either an off-line learning step, or an on-line estimation scheme.

All the methods proposed to estimate numerically the interaction matrix rely on the observation of a variation of the features due to a known or measured camera motion. More precisely, if we measure a feature's variation $\Delta\mathbf{s}$ due to a camera motion $\Delta\mathbf{v}_c$, we have from (24.2):

$$\mathbf{L}_\mathbf{s}\Delta\mathbf{v}_c = \Delta\mathbf{s}$$

which provides k equations while we have $k \times 6$ unknown values in $\mathbf{L}_\mathbf{s}$. Using a set of N independent camera motions with $N > 6$, it is thus possible to estimate $\mathbf{L}_\mathbf{s}$ by solving

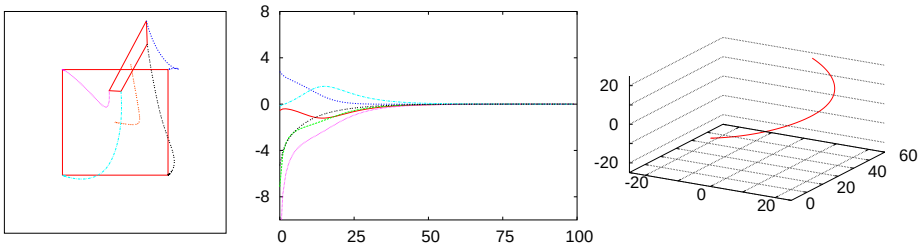
$$\mathbf{L}_\mathbf{s}\mathbf{A} = \mathbf{B}$$

where the columns of $\mathbf{A} \in \mathbb{R}^{6\times N}$ and $\mathbf{B} \in \mathbb{R}^{k\times N}$ are respectively formed with the set of camera motions and the set of corresponding features variations. The least square solution is of course given by

$$\widehat{\mathbf{L}_\mathbf{s}} = \mathbf{B}\mathbf{A}^+ \tag{24.16}$$

Methods based on neural networks have also been developed to estimate $\mathbf{L}_\mathbf{s}$ [21], [22]. It is also possible to estimate directly the numerical value of $\mathbf{L}_\mathbf{s}^+$, which provides in practice a better behavior [23]. In that case, the basic relation is

$$\mathbf{L}_\mathbf{s}^+\Delta\mathbf{s} = \Delta\mathbf{v}_c$$



(a) (b) (c)

Figure 24.8: System behavior using $\mathbf{s} = (\rho_1, \theta_1, ..., \rho_4, \theta_4)$ and $\widehat{\mathbf{L}_\mathbf{e}^+} = \mathbf{L}_{\mathbf{e}^*}^+$.

which provides 6 equations. Using a set of N measurements, with $N > k$, we now obtain:

$$\widehat{\mathbf{L}_\mathbf{s}^+} = \mathbf{A}\mathbf{B}^+ \tag{24.17}$$

In the first case (24.16), the six columns of $\mathbf{L_s}$ are estimated by solving six linear systems, while in the second case (24.17), the k columns of $\mathbf{L}_\mathbf{s}^+$ are estimated by solving k linear systems, which explains the difference in the results.

Estimating the interaction matrix online can be viewed as an optimization problem, and consequently a number of researchers have investigated approaches that derive from optimization methods. These methods typically discretize the system equation (24.2), and use an iterative updating scheme to refine the estimate of $\widehat{\mathbf{L_s}}$ at each stage. One such on-line and iterative formulation uses the Broyden update rule given by [24], [25]:

$$\widehat{\mathbf{L}_\mathbf{s}}(t+1) = \widehat{\mathbf{L}_\mathbf{s}}(t) + \frac{\alpha}{\Delta\mathbf{v}_c^\top \Delta\mathbf{v}_c} \left(\Delta\mathbf{x} - \widehat{\mathbf{L}_\mathbf{s}}(t)\Delta\mathbf{v}_c\right) \Delta\mathbf{v}_c^\top$$

where α defines the update speed. This method has been generalized to the case of moving objects in [26].

The main interest of using such numerical estimations in the control scheme is that it avoids all the modeling and calibration steps. It is particularly useful when using features whose interaction matrix is not available in analytical form. For instance, in [27], the main eigenvalues of the Principal Component Analysis of an image have been considered in a visual servoing scheme. The drawbacks of these methods is that no theoretical stability and robustness analysis can be made.

24.4 Position-Based Visual Servo

Position-based control schemes (PBVS) [3], [28], [29] use the pose of the camera with respect to some reference coordinate frame to define $\mathbf{s}$. Computing that pose from a set of measurements in one image necessitates the camera intrinsic parameters and the 3D model of the object observed to be known. This classical computer vision problem is called the 3D localization problem. While this problem is beyond the scope of the present chapter, many solutions have been presented in the literature (see, e.g., [30], [31]) and its basic principles are recalled in Chapter 23.

It is then typical to define $\mathbf{s}$ in terms of the parameterization used to represent the camera pose. Note that the parameters $\mathbf{a}$ involved in the definition (24.1) of $\mathbf{s}$ are now the camera intrinsic parameters and the 3D model of the object.

It is convenient to consider three coordinate frames: the current camera frame $\mathcal{F}_c$, the desired camera frame $\mathcal{F}_{c^*}$, and a reference frame $\mathcal{F}_o$ attached to the object. We adopt here the standard notation of using a leading superscript to denote the frame with respect to which a set of coordinates is defined. Thus, the coordinate vectors ${}^c\mathbf{t}_o$ and ${}^{c^*}\mathbf{t}_o$ give the coordinates of the origin of the object frame expressed relative to the current camera frame, and relative to the desired camera frame, respectively. Furthermore, let $\mathbf{R} = {}^{c^*}\mathbf{R}_c$ be the rotation matrix that gives the orientation of the current camera frame relative to the desired frame.

We can define $\mathbf{s}$ to be $(\mathbf{t}, \theta\mathbf{u})$, in which $\mathbf{t}$ is a translation vector, and $\theta\mathbf{u}$ gives the angle/axis parameterization for the rotation. We now discuss two choices for $\mathbf{t}$, and give the corresponding control laws.

If $\mathbf{t}$ is defined relative to the object frame $\mathcal{F}_o$, we obtain $\mathbf{s} = ({}^c\mathbf{t}_o, \theta\mathbf{u})$, $\mathbf{s}^* = ({}^{c^*}\mathbf{t}_o, \mathbf{0})$, and $\mathbf{e} = ({}^c\mathbf{t}_o - {}^{c^*}\mathbf{t}_o, \theta\mathbf{u})$. In this case, the interaction matrix related to $\mathbf{e}$ is given by

$$\mathbf{L_e} = \begin{bmatrix} -\mathbf{I}_3 & [{}^c\mathbf{t}_o]_\times \\ \mathbf{0} & \mathbf{L}_{\theta\mathbf{u}} \end{bmatrix} \tag{24.18}$$

in which $\mathbf{I}_3$ is the 3×3 identity matrix and $\mathbf{L}_{\theta\mathbf{u}}$ is given by [32]

$$\mathbf{L}_{\theta\mathbf{u}} = \mathbf{I}_3 - \frac{\theta}{2}[\mathbf{u}]_\times + \left(1 - \frac{\operatorname{sinc}\theta}{\operatorname{sinc}^2\frac{\theta}{2}}\right)[\mathbf{u}]_\times^2 \tag{24.19}$$

where $\operatorname{sinc} x$ is the sinus cardinal defined such that $x \operatorname{sinc} x = \sin x$ and $\operatorname{sinc} 0 = 1$.

Following the development in Section 24.2, we obtain the control scheme

$$\mathbf{v}_c = -\lambda\widehat{\mathbf{L}_\mathbf{e}^{-1}}\mathbf{e}$$

since the dimension k of $\mathbf{s}$ is 6, that is the number of camera degrees of freedom. By setting:

$$\widehat{\mathbf{L}_\mathbf{e}^{-1}} = \begin{bmatrix} -\mathbf{I}_3 & [{}^c\mathbf{t}_o]_\times \mathbf{L}_{\theta\mathbf{u}}^{-1} \\ \mathbf{0} & \mathbf{L}_{\theta\mathbf{u}}^{-1} \end{bmatrix} \tag{24.20}$$

we obtain after simple developments:

$$\begin{cases} \boldsymbol{v}_c &= -\lambda(({}^{c^*}\mathbf{t}_o - {}^c\mathbf{t}_o) + [{}^c\mathbf{t}_o]_\times \theta\mathbf{u}) \\ \boldsymbol{\omega}_c &= -\lambda\theta\mathbf{u} \end{cases} \tag{24.21}$$

since $\mathbf{L}_{\theta\mathbf{u}}$ is such that $\mathbf{L}_{\theta\mathbf{u}}^{-1}\theta\mathbf{u} = \theta\mathbf{u}$.

Ideally, that is if the pose parameters are perfectly estimated, the behavior of $\mathbf{e}$ will be the expected one ($\dot{\mathbf{e}} = -\lambda\mathbf{e}$). The choice of $\mathbf{e}$ causes the rotational motion to follow a geodesic with an exponential decreasing speed and causes the translational parameters involved in $\mathbf{s}$ to decrease with the same speed. This explains the nice exponential decrease of the camera velocity components in Figure 24.9. Furthermore, the trajectory in the image of the origin of the object frame follows a pure straight line (here the center of the four points has been selected as this origin). On the other hand, the camera trajectory does not follow a straight line.

Another PBVS scheme can be designed by using $\mathbf{s} = ({}^{c^*}\mathbf{t}_c, \theta\mathbf{u})$. In that case, we have $\mathbf{s}^* = \mathbf{0}$, $\mathbf{e} = \mathbf{s}$, and

$$\mathbf{L_e} = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{\theta\mathbf{u}} \end{bmatrix} \tag{24.22}$$

Note the decoupling between translational and rotational motions, which allows us to obtain a simple control scheme:

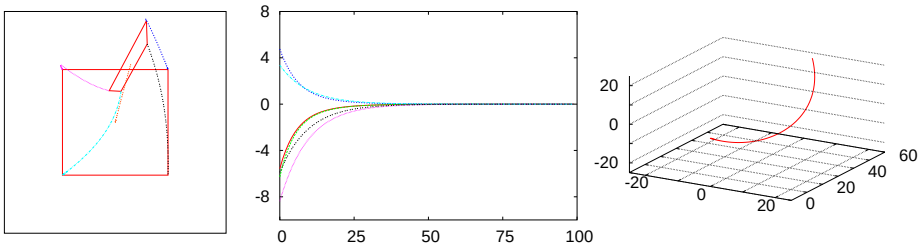
$$\begin{cases} \boldsymbol{v}_c &= -\lambda\mathbf{R}^\top {}^{c^*}\mathbf{t}_c \\ \boldsymbol{\omega}_c &= -\lambda\,\theta\mathbf{u} \end{cases} \tag{24.23}$$

In this case, as can be seen in Figure 24.10, if the pose parameters involved in (24.23) are perfectly estimated, the camera trajectory is a pure straight line, while the image trajectories are less satisfactory than before. Some particular configurations can even be found so that some points leave the camera field of view.

The stability properties of PBVS seem quite attractive. Since $\mathbf{L}_{\theta\mathbf{u}}$ given in (24.19) is nonsingular when $\theta \neq 2k\pi$, we obtain from (24.13) the global asymptotic stability of the system since $\mathbf{L_e}\widehat{\mathbf{L}_\mathbf{e}^{-1}} = \mathbf{I}_6$, under the strong hypothesis that all the pose parameters are perfect. This is true for both methods presented above, since the interactions matrices given in (24.18) and (24.22) are full rank when $\mathbf{L}_{\theta\mathbf{u}}$ is nonsingular.

With regard to robustness, feedback is computed using *estimated* quantities that are a function of the image measurements and the system calibration parameters. For the first method presented in Section 24.4 (the analysis for the second method is analogous), the interaction matrix given in (24.18) corresponds to perfectly estimated pose parameters, while the real one is unknown since the estimated pose parameters may be biased due to calibration errors, or inaccurate and unstable due to noise [11]. The true positivity condition (24.13) should be in fact written:

$$\mathbf{L}_{\widehat{\mathbf{e}}}\widehat{\mathbf{L}_{\widehat{\mathbf{e}}}^{-1}} > 0 \tag{24.24}$$



(a) (b) (c)

Figure 24.9: System behavior using $\mathbf{s} = ({}^{c}\mathbf{t}_0, \theta\mathbf{u})$.

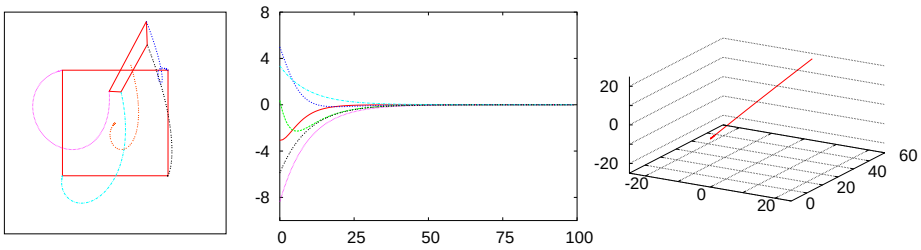


Figure 24.10: System behavior using $\mathbf{s} = ({}^{c^*}\mathbf{t}_c, \theta\mathbf{u})$.

where $\widehat{\mathbf{L}_{\widehat{\mathbf{e}}}^{-1}}$ is given by (24.20) but where $\mathbf{L}_{\widehat{\mathbf{e}}}$ is unknown, and not given by (24.18). Indeed, even small errors in computing the points position in the image can lead to pose errors that can impact significantly the accuracy and the stability of the system (see Figure 24.11).

24.5 Advanced approaches

24.5.1 Hybrid VS

Suppose we have access to a clever control law for $\boldsymbol{\omega}_c$, such as the one used in PBVS (see (24.21) or (24.23)):

$$\boldsymbol{\omega}_c = -\lambda\, \theta\mathbf{u}. \tag{24.25}$$

How could we use this in conjunction with traditional IBVS?

Considering a feature vector $\mathbf{s}_t$ and an error $\mathbf{e}_t$ devoted to control the translational degrees of freedom, we can partition the interaction matrix as follows:

$$\begin{aligned}\dot{\mathbf{s}}_t &= \mathbf{L}_{\mathbf{s}_t}\mathbf{v}_c \\ &= \begin{bmatrix} \mathbf{L}_{\boldsymbol{v}} & \mathbf{L}_{\boldsymbol{\omega}} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_c \\ \boldsymbol{\omega}_c \end{bmatrix} \\ &= \mathbf{L}_{\boldsymbol{v}}\boldsymbol{v}_c + \mathbf{L}_{\boldsymbol{\omega}}\boldsymbol{\omega}_c\end{aligned}$$

Now, setting $\dot{\mathbf{e}}_t = -\lambda\mathbf{e}_t$, we can solve for the desired translational control input as

$$\begin{aligned}-\lambda\mathbf{e}_t = \dot{\mathbf{e}}_t = \dot{\mathbf{s}}_t &= \mathbf{L}_{\boldsymbol{v}}\boldsymbol{v}_c + \mathbf{L}_{\boldsymbol{\omega}}\boldsymbol{\omega}_c \\ \Rightarrow \boldsymbol{v}_c &= -\mathbf{L}_{\boldsymbol{v}}^{+}\left(\lambda\mathbf{e}_t + \mathbf{L}_{\boldsymbol{\omega}}\boldsymbol{\omega}_c\right)\end{aligned} \tag{24.26}$$

We can think of the quantity $(\lambda\mathbf{e}_t + \mathbf{L}_{\boldsymbol{\omega}}\boldsymbol{\omega}_c)$ as a modified error term, one that combines the original error with the error that would be induced by the rotational motion due to $\boldsymbol{\omega}_c$. The translational control input $\boldsymbol{v}_c = -\mathbf{L}_{\boldsymbol{v}}^{+}\left(\lambda\mathbf{e}_t + \mathbf{L}_{\boldsymbol{\omega}}\boldsymbol{\omega}_c\right)$ will drive this error to zero.

The method known as 2 1/2 D Visual Servo [32] was the first to exploit such a partitioning in combining IBVS and PBVS. More precisely, in [32], $\mathbf{s}_t$

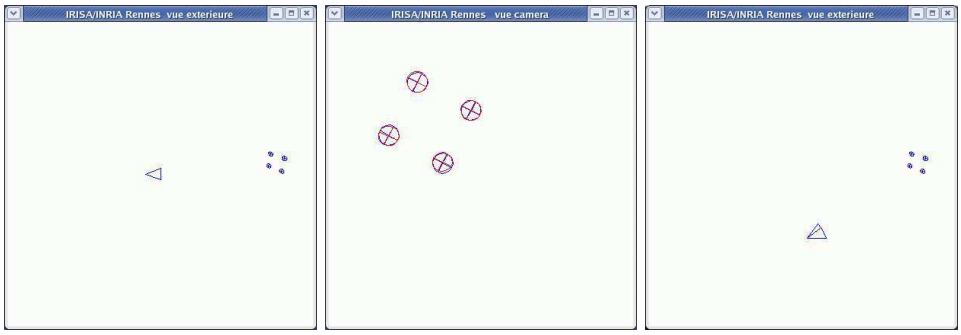


Figure 24.11: Two different camera poses (left and right) that provide almost the same image of four coplanar points (middle)

has been selected as the coordinates of an image point, and the logarithm of its depth, so that $\mathbf{L}_{\boldsymbol{v}}$ is a triangular always invertible matrix. More precisely, we have $\mathbf{s}_t = (\mathbf{x}, \log Z)$, $\mathbf{s}_t^* = (\mathbf{x}^*, \log Z^*)$, $\mathbf{e}_t = (\mathbf{x} - \mathbf{x}^*, \log \rho_Z)$ where $\rho_Z = Z/Z^*$, and

$$\mathbf{L}_{\boldsymbol{v}} = \frac{1}{Z^* \rho_Z} \begin{bmatrix} -1 & 0 & x \\ 0 & -1 & y \\ 0 & 0 & -1 \end{bmatrix}$$

$$\mathbf{L}_{\boldsymbol{\omega}} = \begin{bmatrix} xy & -(1+x^2) & y \\ 1+y^2 & -xy & -x \\ -y & x & 0 \end{bmatrix}$$

Note that the ratio ρ_Z can be obtained directly from the partial pose estimation algorithm that will be described in Section 24.7.

If we come back to the usual global representation of visual servo control schemes, we have $\mathbf{e} = (\mathbf{e}_t, \theta\mathbf{u})$ and $\mathbf{L}_{\mathbf{e}}$ given by:

$$\mathbf{L}_{\mathbf{e}} = \begin{bmatrix} \mathbf{L}_{\boldsymbol{v}} & \mathbf{L}_{\boldsymbol{\omega}} \\ \mathbf{0} & \mathbf{L}_{\theta\mathbf{u}} \end{bmatrix}$$

from which we obtain immediately the control law (24.25) and (24.26) by applying (24.5).

The behavior obtained using this choice for $\mathbf{s}_t$ is given on Figure 24.12. Here, the point that has been considered in $\mathbf{s}_t$ is the center of gravity $\mathbf{x}_g$ of the target. We can note the image trajectory of that point, which is a straight line as expected, and the nice decreasing of the camera velocity components, which makes this scheme very near from the first PBVS one.

As for stability, it is clear that this scheme is globally asymptotically stable in perfect conditions (refer to the stability analysis provided in the Part I of the tutorial: we are here in the simple case $k = 6$). Furthermore, thanks to the triangular form of the interaction matrix $\mathbf{L}_{\mathbf{e}}$, it has been possible to analyze the stability of this scheme in the presence of calibration errors using the partial pose estimation algorithm that will be described in Section 24.7 [33]. Finally, the only unknown constant parameter involved in this scheme, that is Z^*, can be estimated on line using adaptive techniques [34].

Other hybrid schemes can be designed. For instance, in [35], the third component of $\mathbf{s}_t$ is different and has been selected so that all the target points remain in the camera field of view as far as possible. Another example has been proposed in [36]. In that case, $\mathbf{s}$ is selected as $\mathbf{s} = ({}^{c^*}\mathbf{t}_c, \mathbf{x}_g, \theta u_z)$ which provides with a block-triangular interaction matrix of the form:

$$\mathbf{L}_{\mathbf{e}} = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{L}'_{\boldsymbol{v}} & \mathbf{L}'_{\boldsymbol{\omega}} \end{bmatrix}$$

where $\mathbf{L}'_{\boldsymbol{v}}$ and $\mathbf{L}'_{\boldsymbol{\omega}}$ can easily be computed. This scheme is so that, in perfect conditions, the camera trajectory is a straight line (since ${}^{c^*}\mathbf{t}_c$ is a part of $\mathbf{s}$), and the image trajectory of the center of gravity of the object is also a straight line (since $\mathbf{x}_g$ is also a part of $\mathbf{s}$). The translational camera degrees of freedom are devoted to realize the 3D straigth line, while

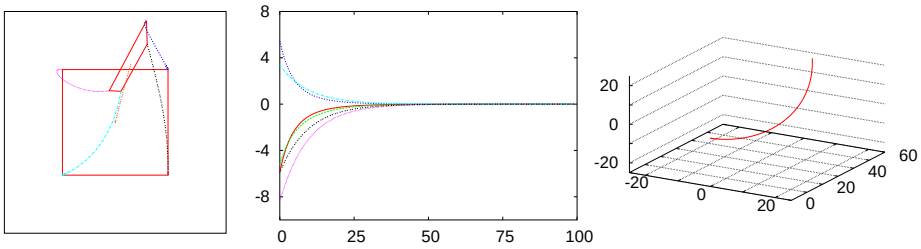


Figure 24.12: System behavior using $\mathbf{s} = (\mathbf{x}_g, \log(Z_g), \theta\mathbf{u})$.

the rotational camera degrees of freedom are devoted to realize the 2D straight line and compensate also the 2D motion of $\mathbf{x}_g$ due to the transational motion. As can be seen on Figure 24.13, this scheme is particularly satisfactory in practice.

Finally, it is possible to combine differently 2D and 3D features. For instance, in [37], it has been proposed to use in $\mathbf{s}$ the 2D homogeneous coordinates of a set of image points expressed in pixels multiplied by their corresponding depth: $\mathbf{s} = (u_1 Z_1, v_1 Z_1, Z_1, \cdots, u_n Z_n, v_n Z_n, Z_n)$. As for classical IBVS, we obtain in that case a set of redundant features, since at least three points have to be used to control the six camera degrees of freedom (we here have $k \geq 9$). However, it has been demonstrated in [38] that this selection of redundant features is free of attractive local minima.

24.5.2 Partitioned VS

The hybrid visual servo schemes described above have been designed to decouple the rotational motions from the translational ones by selecting adequate visual features defined in part in 2D, and in part in 3D (that is why they have been called 2 1/2 D visual servoing). This work has inspired some researchers to find features that exhibit similar decoupling properties but using only features expressed directly in the image. More precisely, the goal is to find six features so that each of them is related to only one degree of freedom (in which case the interaction matrix is a diagonal matrix), the Grail being to find a diagonal interaction matrix whose elements are constant, as near as possible of the identity matrix, leading to a pure, direct, and simple linear control problem.

The first work in this area partitioned the interaction matrix so as to isolate motion related to the optic axis [10]. Indeed, whatever the choice of $\mathbf{s}$, we have

$$\begin{aligned} \dot{\mathbf{s}} &= \mathbf{L}_\mathbf{s}\mathbf{v}_c \\ &= \mathbf{L}_{xy}\mathbf{v}_{xy} + \mathbf{L}_z\mathbf{v}_z \\ &= \dot{\mathbf{s}}_{xy} + \dot{\mathbf{s}}_z \end{aligned}$$

in which $\mathbf{L}_{xy}$ includes the first, second, fourth and fifth columns of $\mathbf{L}_\mathbf{s}$, and $\mathbf{L}_z$ includes the third and sixth columns of $\mathbf{L}_\mathbf{s}$. Similarly, $\mathbf{v}_{xy} = (v_x, v_y, \omega_x, \omega_y)$ and $\mathbf{v}_z = (v_z, \omega_z)$. Here, $\dot{\mathbf{s}}_z = \mathbf{L}_z\mathbf{v}_z$ gives the component of $\dot{\mathbf{s}}$ due to the camera motion along and rotation about the optic axis, while $\dot{\mathbf{s}}_{xy} = \mathbf{L}_{xy}\mathbf{v}_{xy}$ gives the component of $\dot{\mathbf{s}}$ due to velocity along and rotation about the camera x and y axes.

Proceeding as above, by setting $\dot{\mathbf{e}} = -\lambda\mathbf{e}$ we obtain

$$-\lambda\mathbf{e} = \dot{\mathbf{e}} = \dot{\mathbf{s}} = \mathbf{L}_{xy}\mathbf{v}_{xy} + \mathbf{L}_z\mathbf{v}_z,$$

which leads to

$$\mathbf{v}_{xy} = -\mathbf{L}_{xy}^{+}\left(\lambda\mathbf{e}(t) + \mathbf{L}_z\mathbf{v}_z\right).$$

As before, we can consider $(\lambda\mathbf{e}(t) + \mathbf{L}_\mathbf{z}\mathbf{v}_z)$ as a modified error that incorporates the original error while taking into account the error that will be induced by $\mathbf{v}_z$.

Given this result, all that remains is to choose $\mathbf{s}$ and $\mathbf{v}_z$. As for basic IBVS, the coordinates of a collection of image points can be used in $\mathbf{s}$, while two new image features can be defined to determine $\mathbf{v}_z$.

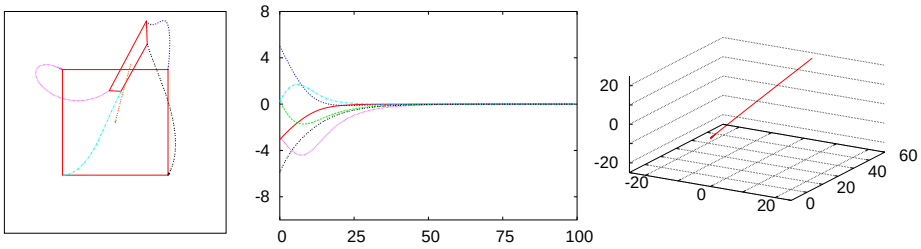


Figure 24.13: System behavior using $\mathbf{s} = ({}^{c^*}\mathbf{t}_c, \mathbf{x}_g, \theta u_z)$.

- Define α, with $0 \leq \alpha < 2\pi$ as the angle between the horizontal axis of the image plane and the directed line segment joining two feature points. It is clear that α is closely related to the rotation around the optic axis.
- Define σ^2 to be the area of the polygon defined by these points. Similarly, σ^2 is closely related to the translation along the optic axis.

Using these features, $\mathbf{v}_z$ has been defined in [10] as

$$\begin{cases} v_z &= \lambda_{v_z} \ln \frac{\sigma^*}{\sigma} \\ \omega_z &= \lambda_{\omega_z}(\alpha^* - \alpha) \end{cases}$$

24.6 Performance Optimization and Planning

In some sense, partitioned methods represent an effort to optimize system performance by assigning distinct features and controllers to individual degrees of freedom. In this way, the designer performs a sort of off-line optimization when allocating controllers to degrees of freedom. It is also possible to explicitly design controllers that optimize various system performance measures. We describe a few of these in this section.

24.6.1 Optimal control and redundancy framework

An example of such an approach is given in [39] and [40], in which LQG control design is used to choose gains that minimize a linear combination of state and control input. This approach explicitly balances the trade-off between tracking errors (since the controller attempts to drive $\mathbf{s} - \mathbf{s}^*$ to zero) and robot motion. A similar control approach is proposed in [41] where joint limit avoidance is considered simultaneously with the positioning task.

It is also possible to formulate optimality criteria that explicitly express the observability of robot motion in the image. For example, the singular value decomposition of the interaction matrix reveals which degrees of freedom are most apparent and can thus be easily controlled, while the condition number of the interaction matrix gives a kind of global measure of the visibility of motion. This concept has been called resolvability in [42] and motion perceptibility in [43]. By selecting features and designing controllers that maximize these measures, either along specific degrees of freedom or globally, the performance of the visual servo system can be improved.

The constraints considered to design the control scheme using the optimal control approach may be contradictory in some cases, leading the system to fail due to local minima in the objective function to be minimized. For example, it may happen that the motion produced to move away from a robot joint limit is exactly the opposite of the motion produced to near the desired pose, which results in a zero global motion. To avoid this potential problem, it is possible to use the Gradient Projection Method, which is classical in robotics. Applying this method to visual servoing has been proposed in [1] and [17]. The approach consists in projecting the secondary constraints $\mathbf{e}_s$ on the null space of the vision-based task $\mathbf{e}$ so that they have no effect on the regulation of $\mathbf{e}$

to 0:

$$\mathbf{e}_g = \widehat{\mathbf{L}_\mathbf{e}^+}\mathbf{e} + \mathbf{P}_\mathbf{e}\ \mathbf{e}_s$$

where $\mathbf{e}_g$ is the new global task considered and $\mathbf{P}_\mathbf{e} = (\mathbf{I}_6 - \widehat{\mathbf{L}_\mathbf{e}^+}\widehat{\mathbf{L}_\mathbf{e}})$ is such that $\widehat{\mathbf{L}_\mathbf{e}}\mathbf{P}_\mathbf{e}\mathbf{e}_s = 0, \forall \mathbf{e}_s$. Avoiding the robot joint limits using this approach has been presented in [44]. However, when the vision-based task constrains all the camera degrees of freedom, the secondary constraints cannot be considered since, when $\widehat{\mathbf{L}_\mathbf{e}}$ is of full rank 6, we have $\mathbf{P}_\mathbf{e}\mathbf{e}_s = 0, \forall \mathbf{e}_s$. In that case, it is necessary to inject the constraints in a global objective function, such as navigation functions which are free of local minima [45], [46].

24.6.2 Switching schemes

The partitioned methods described previously attempt to optimize performance by assigning individual controllers to specific degrees of freedom. Another way to use multiple controllers to optimize performance is to design switching schemes that select at each moment in time which controller to use based on criteria to be optimized.

A simple switching controller can be designed using an IBVS and a PBVS controller as follows [47]. Let the system begin by using the IBVS controller. Consider the Lyapunov function for the PBVS controller given by $\mathcal{L} = \frac{1}{2}\|\mathbf{e}(t)\|^2$, with $\mathbf{e}(t) = ({}^c\mathbf{t}_o - {}^{c^*}\mathbf{t}_o, \theta\mathbf{u})$. If at any time the value of this Lyapunov function exceeds a threshold γ_P, the system switches to the PBVS controller. While using the PBVS controller, if at any time the value of the Lyapunov function for the IBVS controller exceeds a threshold, $\mathcal{L} = \frac{1}{2}\|\mathbf{e}(t)\|^2 > \gamma_I$, the system switches to the IBVS controller. With this scheme, when the Lyapunov function for a particular controller exceeds a threshold, that controller is invoked, which in turn reduces the value of the corresponding Lyapunov function. If the switching thresholds are selected appropriately, the system is able to exploit the relative advantages of IBVS and PBVS, while avoiding their shortcomings.

An example for this system is shown in Figure 24.14, for the case of a rotation by 160° about the optical axis. Note that the system begins in IBVS mode and the features initially move on straight lines toward their goal positions in the image. However, as the camera retreats, the system switches to PBVS, which allows the camera to reach its desired position by combining a rotational motion around its optic axis and a forward translational motion, producing the circular trajectories observed in the image.

Other examples of temporal switching schemes can be found, such as for instance the one developed in [48] to ensure the visibility of the target observed.

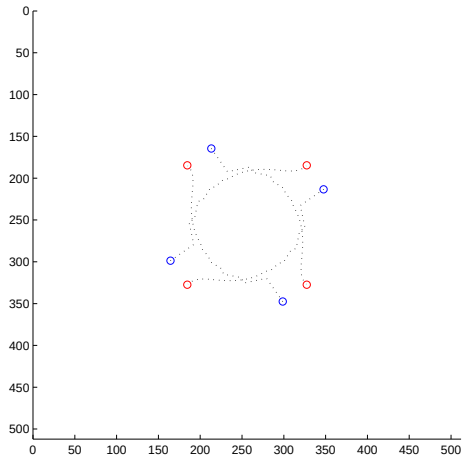


Figure 24.14: Image feature trajectories for a rotation of 160° about the optical axis using a switched control scheme (initial points position in blue, and desired points position in red).

24.6.3 Feature trajectory planning

It is also possible to treat the optimization problem offline, during a planning stage. In that case, several constraints can be simultaneously taken into account, such as obstacle avoidance [49], joint limit and occlusions avoidance, and ensuring the visibility of the target [50]. The feature trajectories $\mathbf{s}^*(t)$ that allow the camera to reach its desired pose while ensuring the constraints are satisfied are determined using path planning techniques, such as the well known potential field approach.

Coupling path planning with trajectory following also allows to improve significantly the robustness of the visual servo with respect to modeling errors. Indeed, modeling errors may have large effects when the error $\mathbf{s}-\mathbf{s}^*$ is large, but have few effects when $\mathbf{s}-\mathbf{s}^*$ is small. Once desired features trajectories $\mathbf{s}^*(t)$ such that $\mathbf{s}^*(0)=\mathbf{s}(0)$ have been designed during the planning stage, it is easy to adapt the control scheme to take into account the fact that $\mathbf{s}^*$ is varying, and to make the error $\mathbf{s}-\mathbf{s}^*$ remain small. More precisely, we now have

$$\dot{\mathbf{e}} = \dot{\mathbf{s}} - \dot{\mathbf{s}}^* = \mathbf{L}_\mathbf{e}\mathbf{v}_c - \dot{\mathbf{s}}^*,$$

from which we deduce, by selecting as usual $\dot{\mathbf{e}} = -\lambda\mathbf{e}$ as desired behavior,

$$\mathbf{v}_c = -\lambda\widehat{\mathbf{L}_\mathbf{e}^+}\mathbf{e} + \widehat{\mathbf{L}_\mathbf{e}^+}\dot{\mathbf{s}}^*.$$

The new second term of this control law anticipates the variation of $\mathbf{s}^*$, removing the tracking error it would produce. We will see in Section 24.8 that the form of the control law is similar when tracking a moving target is considered.

24.7 Estimation of 3D parameters

All the control schemes described in the previous sections use 3D parameters that are not directely available from the image measurements. As for IBVS, we recall that the range of the object with respect to the camera appears in the coefficients of the interaction matrix related to the translational degrees of freedom. Noticeable exceptions are the schemes based on a numerical estimation of $\mathbf{L}_\mathbf{e}$ or of $\mathbf{L}_\mathbf{e}^+$ (see Section 24.3.8). Another exception is the IBVS scheme that uses the constant matrix $\widehat{\mathbf{L}_{\mathbf{e}^*}^+}$ in the control scheme. In that case, only the depth for the desired pose is required, which is not so difficult to obtain in practice. As for PBVS and hybrid schemes that combine 2D and 3D data in $\mathbf{e}$, 3D parameters appear both in the error $\mathbf{e}$ to be regulated to $\mathbf{0}$ and in the interaction matrix. A correct estimation of the 3D parameters involved is thus important for IBVS since they will have an effect on the camera motion during the task execution (they appear in the stability conditions (24.13) and (24.14)), while a correct estimation is crucial in PBVS and hybrid schemes since they will have also an effect on the accuracy of the pose reached after convergence.

If a calibrated stereo vision system is used, all 3D parameters can be easily determined by triangulation, as evoked in Section 24.3.5 and described in Chapter 23. Similarly, if a 3D model of the object is known, all 3D parameters can be computed from a pose estimation algorithm. However, we recall that such an estimation can be quite unstable due to image noise. It is also possible to estimate 3D parameters by using the epipolar geometry that relates the images of the same scene observed from different viewpoints. Indeed, in visual servoing, two images are generally available: the current one and the desired one.

Given a set of matches between the image measurements in the current image and in the desired one, the fundamental matrix, or the esential matrix if the camera is calibrated, can be recovered [6], and then used in visual servoing [51]. Indeed, from the essential matrix, the rotation and the translation up to a scalar factor between the two views can be estimated. However, near the convergence of the visual servo, that is when the current and desired images are similar, the epipolar geometry becomes degenerate and it is not possible to estimate accurately the partial pose between the two views. For this reason, using homography is generally prefered.

Let $\mathbf{x}_i$ and $\mathbf{x}_i^*$ denote the homogeneous image coordinates for a point in the current and desired images. Then $\mathbf{x}_i$ is related to $\mathbf{x}_i^*$ by

$$\mathbf{x}_i = \mathbf{H}_i\mathbf{x}_i^*$$

in which $\mathbf{H}_i$ is a homography matrix.

If all feature points lie on a 3D plane, then there is a single homography matrix $\mathbf{H}$ such that $\mathbf{x}_i = \mathbf{H}\mathbf{x}_i^*$ for all i. This homography can be estimated using the position of four matched points in the desired and the current images. If all the features points do not belong to the same 3D plane, then three points can be used to define such a plane and five supplementary points are needed to estimate $\mathbf{H}$ [52].

Once $\mathbf{H}$ is available, it can be decomposed as

$$\mathbf{H} = \mathbf{R} + \frac{\mathbf{t}}{d^*}\mathbf{n}^{*\top}, \tag{24.27}$$

in which $\mathbf{R}$ is the rotation matrix relating the orientation of the current and desired camera frames, $\mathbf{n}^*$ is the normal to the chosen 3D plane expressed in the desired frame, d^* is the distance to the 3D plane from the desired frame, and $\mathbf{t}$ is the translation between current and desired frames. From $\mathbf{H}$, it is thus posible to recover $\mathbf{R}$, $\mathbf{t}/d^*$, and $\mathbf{n}$. In fact, two solutions for these quantities exist [53], but it is quite easy to select the correct one using some knowledge about the desired pose. It is also possible to estimate the depth of any target point up to a common scale factor [50]. The unknown depth of each point that appears in classical IBVS can thus be expressed as a function of a single constant parameter. Similarly, the pose parameters required by PBVS can be recovered up to a scalar factor as for the translation term. The PBVS schemes described previously can thus be revisited using this approach, with the new error defined as the translation up to a scalar factor and the angle/axis parameterization of the rotation. Finally, this approach has also been used for the hybrid visual servoing schemes described in Section 24.5.1. In that case, using such homography estimation, it has been possible to analyse the stability of hybrid visual servoing schemes in the presence of calibration errors [32].

24.8 Target tracking

We now consider the case of a moving target and a constant desired value $\mathbf{s}^*$ for the features, the generalization to varying desired features $\mathbf{s}^*(t)$ being immediate. The time variation of the error is now given by:

$$\dot{\mathbf{e}} = \mathbf{L}_\mathbf{e}\mathbf{v}_c + \frac{\partial \mathbf{e}}{\partial t} \tag{24.28}$$

where the term $\frac{\partial \mathbf{e}}{\partial t}$ expresses the time variation of $\mathbf{e}$ due to the generally unknown target motion. If the control law is still designed to try to ensure an exponential decoupled decrease of $\mathbf{e}$ (that is, once again $\dot{\mathbf{e}} = -\lambda\mathbf{e}$), we now obtain using (24.28):

$$\mathbf{v}_c = -\lambda\widehat{\mathbf{L}_\mathbf{e}^+}\mathbf{e} - \widehat{\mathbf{L}_\mathbf{e}^+}\widehat{\frac{\partial \mathbf{e}}{\partial t}} \tag{24.29}$$

where $\widehat{\frac{\partial \mathbf{e}}{\partial t}}$ is an estimation or an approximation of $\frac{\partial \mathbf{e}}{\partial t}$. This term must be introduced in the control law to compensate for the target motion.

Closing the loop, that is injecting (24.29) in (24.28), we obtain

$$\dot{\mathbf{e}} = -\lambda\mathbf{L}_\mathbf{e}\widehat{\mathbf{L}_\mathbf{e}^+}\mathbf{e} - \mathbf{L}_\mathbf{e}\widehat{\mathbf{L}_\mathbf{e}^+}\widehat{\frac{\partial \mathbf{e}}{\partial t}} + \frac{\partial \mathbf{e}}{\partial t} \tag{24.30}$$

Even if $\mathbf{L}_\mathbf{e}\widehat{\mathbf{L}_\mathbf{e}^+} > 0$, the error will converge to zero only if the estimation $\widehat{\frac{\partial \mathbf{e}}{\partial t}}$ is sufficiently accurate so that

$$\mathbf{L}_\mathbf{e}\widehat{\mathbf{L}_\mathbf{e}^+}\widehat{\frac{\partial \mathbf{e}}{\partial t}} = \frac{\partial \mathbf{e}}{\partial t}. \tag{24.31}$$

Otherwise, tracking errors will be observed. Indeed, by just solving the scalar differential equation $\dot{e} = -\lambda e + b$, which is a simplification of (24.30), we obtain $e(t) = e(0)\exp(-\lambda t) + b/\lambda$, which converges towards b/λ. On one hand, setting a high gain λ will reduce the tracking error, but on the other hand, setting the gain too high can make the system unstable. It is thus necessary to make b as small as possible.

Of course, if the system is known to be such that $\frac{\partial \mathbf{e}}{\partial t} = 0$ (that is the camera observes a motionless object, as it has been described in Section 24.2), no tracking error will appear with the most simple estimation given by $\widehat{\frac{\partial \mathbf{e}}{\partial t}} = 0$. Otherwise, a classical method in automatic control to cancel tracking errors consists in compensating the target motion through an integral term in the control law. In that case, we have

$$\widehat{\frac{\partial \mathbf{e}}{\partial t}} = \mu \sum_j \mathbf{e}(j)$$

where μ is the integral gain that has to be tuned. This scheme allows to cancel the tracking errors only if the target has a constant velocity. Other methods, based on feedforward control, estimate directly the term $\widehat{\frac{\partial \mathbf{e}}{\partial t}}$ through the image measurements and the camera velocity, when it is available. Indeed, from (24.28), we

get:

$$\widehat{\frac{\partial \mathbf{e}}{\partial t}} = \widehat{\dot{\mathbf{e}}} - \widehat{\mathbf{L_e}}\widehat{\mathbf{v}_c}$$

where $\widehat{\dot{\mathbf{e}}}$ can for instance be obtained as $\widehat{\dot{\mathbf{e}}}(t) = (\mathbf{e}(t) - \mathbf{e}_(t - \Delta t))/\Delta t$, Δt being the duration of the control loop. A Kalman filter [54] or more elaborate filtering methods [55] can then be used to improve the estimated values obtained. If some knowledge about the target velocity or the target trajectory is available, it can of course be used to smooth or predict the motion [56], [57], [58]. For instance, in [59], the periodic motion of the heart and of the breath are compensated for an application of visual servoing in medical robotics. Finally, other methods have been developed to remove as fast as possible the perturbations induced by the target motion [39], using for instance predictive controllers [60].

24.9 Eye-in-hand and eye-to-hand systems controlled in the joint space

In the previous sections, we have considered the six components of the camera velocity as input of the robot controller. As soon as the robot is not able to realize this motion, for instance because it has less than six degrees of freedom, the control scheme must be expressed in the joint space. In this section, we describe how this can be done, and in the process develop a formulation for eye-to-hand systems.

In the joint space, the system equations for both the eye-to-hand configuration and the eye-in-hand configuration have the same form

$$\dot{\mathbf{s}} = \mathbf{J_s}\,\dot{\mathbf{q}} + \frac{\partial \mathbf{s}}{\partial t} \tag{24.32}$$

Here, $\mathbf{J_s} \in \mathbb{R}^{k \times n}$ is the feature Jacobian matrix, which can be linked to the interaction matrix, and n is the number of robot joints.

For an eye-in-hand system (see Figure 24.15.a), $\frac{\partial \mathbf{s}}{\partial t}$ is the time variation of $\mathbf{s}$ due to a potential object motion, and $\mathbf{J_s}$ is given by

$$\mathbf{J_s} = \mathbf{L_s}\,{}^c\mathbf{X}_N\mathbf{J}(\mathbf{q}) \tag{24.33}$$

where

- ${}^c\mathbf{X}_N$ is the spatial motion transform matrix (as defined in Chapter 1 and recalled in (24.15)) from the vision sensor frame to the end effector frame. It is usually a constant matrix (as soon as the vision sensor is rigidly attached to the end effector). Thanks to the robustness of closed loop control schemes, a coarse approximation of this transform matrix is sufficient in visual servoing. If needed, an accurate estimation is possible through classical hand-eye calibration methods [61].
- $\mathbf{J}(\mathbf{q})$ is the robot Jacobian expressed in the end effector frame (as defined in Chapter 1)

For an eye-to-hand system (see Figure 24.15.b), $\frac{\partial \mathbf{s}}{\partial t}$ is now the time variation of $\mathbf{s}$ due to a potential vision sensor motion and $\mathbf{J_s}$ can be expressed as:

$$\mathbf{J_s} = -\mathbf{L_s}\,{}^c\mathbf{X}_N{}^N\mathbf{J}(\mathbf{q}) \tag{24.34}$$

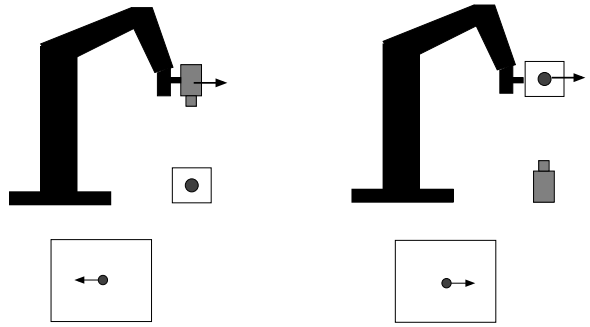
$$= -\mathbf{L_s}\,{}^c\mathbf{X}_0{}^0\mathbf{J}(\mathbf{q}) \tag{24.35}$$

In (24.34), the classical robot Jacobian ${}^N\mathbf{J}(\mathbf{q})$ expressed in the end effector frame is used but the spatial motion transform matrix ${}^c\mathbf{X}_N$ from the vision sensor frame to the end effector frame changes all along the servo, and it has to be estimated at each iteration of the control scheme, usually using pose estimation methods.

In (24.35), the robot Jacobian ${}^0\mathbf{J}(\mathbf{q})$ is expressed in the robot reference frame and the spatial motion transform matrix ${}^c\mathbf{X}_0$ from the vision sensor frame to that reference frame is constant as long as the camera does not move. In that case, which is convenient in practice, a coarse approximation of ${}^c\mathbf{X}_0$ is usually sufficient.

Once the modeling step is finished, it is quite easy to follow the procedure that has been used above to design a control scheme expressed in the joint space, and to determine sufficient condition to ensure the stability of the control scheme. We obtain, considering again $\mathbf{e} = \mathbf{s} - \mathbf{s}^*$, and an exponential decoupled decrease of $\mathbf{e}$:

$$\dot{\mathbf{q}} = -\lambda\widehat{\mathbf{J}_\mathbf{e}^+}\mathbf{e} - \widehat{\mathbf{J}_\mathbf{e}^+}\widehat{\frac{\partial \mathbf{e}}{\partial t}} \tag{24.36}$$



(a) (b)

Figure 24.15: on the top: a) eye-in-hand system, b) eye-to-hand system; on the bottom: oppsite image motion produced by the same robot motion

If $k = n$, considering as in Section 24.2 the Lyapunov function $\mathcal{L} = \frac{1}{2}\|\mathbf{e}(t)\|^2$, a sufficient condition to ensure the global asymptotic stability is given by:

$$\mathbf{J_e}\widehat{\mathbf{J_e^+}} > 0 \tag{24.37}$$

If $k > n$, we obtain similarly as in Section 24.2

$$\widehat{\mathbf{J_e^+}}\mathbf{J_e} > 0 \tag{24.38}$$

to ensure the local asymptotic stability of the system. Note that the actual extrinsic camera parameters appears in $\mathbf{J_e}$ while the estimated ones are used in $\widehat{\mathbf{J_e^+}}$. It is thus possible to analyse the robustness of the control scheme with respect to the camera extrinsic parameters. It is also possible to estimate directly the numerical value of $\mathbf{J_e}$ or $\mathbf{J_e^+}$ using the methods described in Section 24.3.8.

Finally, to remove tracking errors, we have to ensure that:

$$\mathbf{J_e}\widehat{\mathbf{J_e^+}}\widehat{\frac{\partial \mathbf{e}}{\partial t}} = \frac{\partial \mathbf{e}}{\partial t}$$

Finally, let us note that, even if the robot has six degrees of freedom, it is generally not equivalent to first compute $\mathbf{v}_c$ using (24.5) and then deduce $\dot{\mathbf{q}}$ using the robot inverse Jacobian, and to compute directly $\dot{\mathbf{q}}$ using (24.36). Indeed, it may occur that the robot Jacobian $\mathbf{J}(\mathbf{q})$ is singular while the feature Jacobian $\mathbf{J_s}$ is not (that may occur when $k < n$). Furthermore, the properties of the pseudo-inverse ensure that using (24.5), $\|\mathbf{v}_c\|$ is minimal while using (24.36), $\|\dot{\mathbf{q}}\|$ is minimal. As soon as $\mathbf{J_e^+} \neq \mathbf{J}^+(\mathbf{q})^N\mathbf{X}_c\mathbf{L_e^+}$, the control schemes will be different and will induce different robot trajectories. The choice of the state space is thus important.

24.10 Conclusions

In this chapter, we have only considered velocity controllers. It is convenient for most of classical robot arms. However, the dynamics of the robot must of course be taken into account for high speed tasks, or when we deal with mobile nonholonomic or underactuated robots. As for the sensor, we have only considered geometrical features coming from a classical perspective camera. Features related to the image motion or coming from other vision sensors (fish-eye camera, catadioptric camera, echographic probes, ...) necessitate to revisit the modeling issues to select adequate visual features. Finally, fusing visual features with data coming from other sensors (force sensor, proximetry sensors, ...) at the level of the control scheme will allow to address new research topics. Numerous ambitious applications of visual servoing can also be considered, as well as for mobile robots in indoor or outdoor environments, for aerial, space, and submarine robots, and in medical robotics. The end of fruitful research in the field of visual servo is thus nowhere yet in sight.

Bibliography

[1] B. Espiau, F. Chaumette, and P. Rives, "A new approach to visual servoing in robotics," *IEEE Trans. on Robotics and Automation*, vol. 8, pp. 313–326, Jun. 1992.

[2] S. Hutchinson, G. Hager, and P. Corke, "A tutorial on visual servo control," *IEEE Trans. on Robotics and Automation*, vol. 12, pp. 651–670, Oct. 1996.

[3] L. Weiss, A. Sanderson, and C. Neuman, "Dynamic sensor-based control of robots with visual feedback," *IEEE J. on Robotics and Automation*, vol. 3, pp. 404–417, Oct. 1987.

[4] J. Feddema and O. Mitchell, "Vision-guided servoing with feature-based trajectory generation," *IEEE Trans. on Robotics and Automation*, vol. 5, pp. 691–700, Oct. 1989.

[5] D. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Upper Saddle River, NJ: Prentice Hall, 2003.

[6] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry, *An Invitation to 3-D Vision: From Images to Geometric Models*. New York: Springer-Verlag, 2003.

[7] H. Michel and P. Rives, "Singularities in the determination of the situation of a robot effector from the perspective view of three points," Tech. Rep. 1850, INRIA Research Report, Feb. 1993.

[8] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, pp. 381–395, Jun. 1981.

[9] E. Malis, "Improving vision-based control using efficient second-order minimization techniques," in *IEEE Int. Conf. on Robotics and Automation*, (New Orleans), pp. 1843–1848, Apr. 2004.

[10] P. Corke and S. Hutchinson, "A new partitioned approach to image-based visual servo control," *IEEE Trans. on Robotics and Automation*, vol. 17, pp. 507–515, Aug. 2001.

[11] F. Chaumette, "Potential problems of stability and convergence in image-based and position-based visual servoing," in *The Confluence of Vision and Control* (D. Kriegman, G. Hager, and S. Morse, eds.), vol. 237 of *Lecture Notes in Control and Information Sciences*, pp. 66–78, Springer-Verlag, 1998.

[12] E. Malis, "Visual servoing invariant to changes in camera intrinsic parameters," *IEEE Trans. on Robotics and Automation*, vol. 20, pp. 72–81, Feb. 2004.

[13] A. Isidori, *Nonlinear Control Systems, 3rd edition*. Springer-Verlag, 1995.

[14] G. Hager, W. Chang, and A. Morse, "Robot feedback control based on stereo vision: Towards calibration-free hand-eye coordination," *IEEE Control Systems Magazine*, vol. 15, pp. 30–39, Feb. 1995.

[15] M. Iwatsuki and N. Okiyama, "A new formulation of visual servoing based on cylindrical coordinate system," *IEEE Trans. on Robotics and Automation*, vol. 21, pp. 266–273, Apr. 2005.

[16] F. Chaumette, P. Rives, and B. Espiau, "Classification and realization of the different vision-based tasks," in *Visual Servoing* (K. Hashimoto,

ed.), vol. 7 of *Robotics and Automated Systems*, pp. 199–228, World Scientific, 1993.

[17] A. Castano and S. Hutchinson, "Visual compliance: task directed visual servo control," *IEEE Trans. on Robotics and Automation*, vol. 10, pp. 334–342, Jun. 1994.

[18] G. Hager, "A modular system for robust positioning using feedback from stereo vision," *IEEE Trans. on Robotics and Automation*, vol. 13, pp. 582–595, Aug. 1997.

[19] F. Chaumette, "Image moments: a general and useful set of features for visual servoing," *IEEE Trans. on Robotics and Automation*, vol. 20, pp. 713–723, Aug. 2004.

[20] O. Tahri and F. Chaumette, "Point-based and region-based image moments for visual servoing of planar objects," *IEEE Trans. on Robotics*, vol. 21, pp. 1116–1127, Dec. 2005.

[21] I. Suh, "Visual servoing of robot manipulators by fuzzy membership function based neural networks," in *Visual Servoing* (K. Hashimoto, ed.), vol. 7 of *World Scientific Series in Robotics and Automated Systems*, pp. 285–315, World Scientific Press, 1993.

[22] G. Wells, C. Venaille, and C. Torras, "Vision-based robot positioning using neural networks," *Image and Vision Computing*, vol. 14, pp. 75–732, Dec. 1996.

[23] J.-T. Lapresté, F. Jurie, and F. Chaumette, "An efficient method to compute the inverse jacobian matrix in visual servoing," in *IEEE Int. Conf. on Robotics and Automation*, (New Orleans), pp. 727–732, Apr. 2004.

[24] K. Hosada and M. Asada, "Versatile visual servoing without knowledge of true jacobian," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, (Munchen), pp. 186–193, Sep. 1994.

[25] M. Jägersand, O. Fuentes, and R. Nelson, "Experimental evaluation of uncalibrated visual servoing for precision manipulation," in *IEEE Int. Conf. on Robotics and Automation*, (Albuquerque), pp. 2874–2880, Apr. 1997.

[26] J. Piepmeier, G. M. Murray, and H. Lipkin, "Uncalibrated dynamic visual servoing," *IEEE Trans. on Robotics and Automation*, vol. 20, pp. 143–147, Feb. 2004.

[27] K. Deguchi, "Direct interpretation of dynamic images and camera motion for visual servoing without image feature correspondence," *J. of Robotics and Mechatronics*, vol. 9, no. 2, pp. 104–110, 1997.

[28] W. Wilson, C. Hulls, and G. Bell, "Relative end-effector control using cartesian position based visual servoing," *IEEE Trans. on Robotics and Automation*, vol. 12, pp. 684–696, Oct. 1996.

[29] B. Thuilot, P. Martinet, L. Cordesses, and J. Gallice, "Position based visual servoing: Keeping the object in the field of vision," in *IEEE Int. Conf. on Robotics and Automation*, (Washington D.C.), pp. 1624–1629, May 2002.

[30] D. Dementhon and L. Davis, "Model-based object pose in 25 lines of code," *Int. J. of Computer Vision*, vol. 15, pp. 123–141, Jun. 1995.

[31] D. Lowe, "Three-dimensional object recognition from single two-dimensional images," *Artificial Intelligence*, vol. 31, no. 3, pp. 355–395, 1987.

[32] E. Malis, F. Chaumette, and S. Boudet, "2-1/2D visual servoing," *IEEE Trans. on Robotics and Automation*, vol. 15, pp. 238–250, Apr. 1999.

[33] E. Malis and F. Chaumette, "Theoretical improvements in the stability analysis of a new class of model-free visual servoing methods," *IEEE Trans. on Robotics and Automation*, vol. 18, pp. 176–186, Apr. 2002.

[34] J. Chen, D. Dawson, W. Dixon, and A. Behal, "Adaptive homography-based visual servo tracking for fixed and camera-in-hand configurations," *IEEE Trans. on Control Systems Technology*, vol. 13, pp. 814–825, Sep. 2005.

[35] G. Morel, T. Leibezeit, J. Szewczyk, S. Boudet, and J. Pot, "Explicit incorporation of 2D constraints in vision-based control of robot manipulators," in *Int. Symp. on Experimental Robotics, vol. 250 of LNCIS Series*, pp. 99–108, 2000.

[36] F. Chaumette and E. Malis, "2 1/2 D visual servoing: a possible solution to improve image-based and position-based visual servoings," in *IEEE Int. Conf. on Robotics and Automation*, (San Fransisco), pp. 630–635, Apr. 2000.

[37] E. Cervera, A. D. Pobil, F. Berry, and P. Martinet, "Improving image-based visual servoing with three-dimensional features," *Int. J. of Robotics Research*, vol. 22, pp. 821–840, Oct. 2004.

[38] F. Schramm, G. Morel, A. Micaelli, and A. Lottin, "Extended-2D visual servoing," in *IEEE Int. Conf. on Robotics and Automation*, (New Orleans), pp. 267–273, Apr. 2004.

[39] N. Papanikolopoulos, P. Khosla, and T. Kanade, "Visual tracking of a moving target by a camera mounted on a robot: A combination of vision and control," *IEEE Trans. on Robotics and Automation*, vol. 9, pp. 14–35, Feb. 1993.

[40] K. Hashimoto and H. Kimura, "LQ optimal and nonlinear approaches to visual servoing," in *Visual Servoing* (K. Hashimoto, ed.), vol. 7 of *Robotics and Automated Systems*, pp. 165–198, World Scientific, 1993.

[41] B. Nelson and P. Khosla, "Strategies for increasing the tracking region of an eye-in-hand system by singularity and joint limit avoidance," *Int. J. of Robotics Research*, vol. 14, pp. 225–269, Jun. 1995.

[42] B. Nelson and P. Khosla, "Force and vision resolvability for assimilating disparate sensory feedback," *IEEE Trans. on Robotics and Automation*, vol. 12, pp. pp. 714 – 731, Oct. 1996.

[43] R. Sharma and S. Hutchinson, "Motion perceptibility and its application to active vision-based servo control," *IEEE Trans. on Robotics and Automation*, vol. 13, pp. 607–617, Aug. 1997.

[44] E. Marchand, F. Chaumette, and A. Rizzo, "Using the task function approach to avoid robot joint limits and kinematic singularities in visual servoing," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, (Osaka), pp. 1083–1090, Nov. 1996.

[45] E. Marchand and G. Hager, "Dynamic sensor planning in visual servoing," in *IEEE Int. Conf. on Robotics and Automation*, (Leuven), pp. 1988–1993, May 1998.

[46] N. Cowan, J. Weingarten, and D. Koditschek, "Visual servoing via navigation functions," *IEEE Trans. on Robotics and Automation*, vol. 18, pp. 521–533, Aug. 2002.

[47] N. Gans and S. Hutchinson, "An asymptotically stable switched system visual controller for eye in hand robots," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, (Las Vegas), pp. 735 – 742, Oct. 2003.

[48] G. Chesi, K. Hashimoto, D. Prattichizio, and A. Vicino, "Keeping features in the field of view in eye-in-hand visual servoing: a switching approach," *IEEE Trans. on Robotics and Automation*, vol. 20, pp. 908–913, Oct. 2004.

[49] K. Hosoda, K. Sakamato, and M. Asada, "Trajectory generation for obstacle avoidance of uncalibrated stereo visual servoing without 3D reconstruction," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, vol. 3, (Pittsburgh), pp. 29–34, Aug. 1995.

[50] Y. Mezouar and F. Chaumette, "Path planning for robust image-based control," *IEEE Trans. on Robotics and Automation*, vol. 18, pp. 534–549, Aug. 2002.

[51] R. Basri, E. Rivlin, and I. Shimshoni, "Visual homing: Surfing on the epipoles," *Int. J. of Computer Vision*, vol. 33, pp. 117–137, Sep. 1999.

[52] E. Malis, F. Chaumette, and S. Boudet, "2 1/2 D visual servoing with respect to unknown objects through a new estimation scheme of camera displacement," *Int. J. of Computer Vision*, vol. 37, pp. 79–97, Jun. 2000.

[53] O. Faugeras, *Three-dimensional computer vision: a geometric viewpoint*. MIT Press, 1993.

[54] P. Corke and M. Goods, "Controller design for high performance visual servoing," in *12th World Congress IFAC'93*, (Sydney), pp. 395–398, Jul. 1993.

[55] F. Bensalah and F. Chaumette, "Compensation of abrupt motion changes in target tracking by visual servoing," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, (Pittsburgh), pp. 181–187, Aug. 1995.

[56] P. Allen, B. Yoshimi, A. Timcenko, and P. Michelman, "Automated tracking and grasping of a moving object with a robotic hand-eye system," *IEEE Trans. on Robotics and Automation*, vol. 9, pp. 152–165, Apr. 1993.

[57] K. Hashimoto and H. Kimura, "Visual servoing with non linear observer," in *IEEE Int. Conf. on Robotics and Automation*, (Nagoya), pp. 484–489, Apr. 1995.

[58] A. Rizzi and D. Koditschek, "An active visual estimator for dexterous manipulation," *IEEE Trans. on Robotics and Automation*, vol. 12, pp. 697–713, Oct. 1996.

[59] R. Ginhoux, J. Gangloff, M. de Mathelin, L. Soler, M. A. Sanchez, and J. Marescaux, "Active filtering of physiological motion in robotized surgery using predictive control," *IEEE Trans. on Robotics*, vol. 21, pp. 67–79, Feb. 2005.

[60] J. Gangloff and M. de Mathelin, "Visual servoing of a 6 dof manipulator for unknown 3D profile following," *IEEE Trans. on Robotics and Automation*, vol. 18, pp. 511–520, Aug. 2002.

[61] R. Tsai and R. Lenz, "A new technique for fully autonomous and efficient 3D robotics hand-eye calibration," *IEEE Trans. on Robotics and Automation*, vol. 5, pp. 345–358, Jun. 1989.

Index