J Intell Robot Syst (2007) 49:219–236
DOI 10.1007/s10846-007-9135-z

Precise Positioning of Binocular Eye-to-Hand Robotic Manipulators

Wen-Chung Chang

Received: 29 August 2006 / Accepted: 21 January 2007 /
Published online: 27 February 2007

Abstract This article addresses the visual servoing of a rigid robotic manipulator equipped with a binocular vision system in eye-to-hand configuration. The control goal is to move the robot end-effector to a *visually determined* target position *precisely* without knowing the precise camera model. Many vision-based robotic positioning systems have been successfully implemented and validated by supporting experimental results. Nevertheless, this research aims at providing stability analysis for a class of robotic set-point control systems employing image-based feedback laws. Specifically, by exploring epipolar geometry of the binocular vision system, a *binocular visual constraint* is found to assist in establishing stability property of the feedback system. Any three-degree-of-freedom positioning task, if satisfying appropriate conditions with the *image-based* encoding approach, can be encoded in such a way that the encoded error, when driven to zero, implies that the original task has been accomplished with precision. The corresponding image-based control law is proposed to drive the encoded error to zero. The overall closed-loop system is exponentially stable provided that the binocular model imprecision is small.

Key words binocular vision · binocular visual constraint · epipolar geometry · exponential stability · regulation · robot control · task encoding · visual servoing

1 Introduction

It has long been a goal of researchers in robotics to develop systems capable of operating in environments that are dynamically changing and imprecisely calibrated.

W.-C. Chang (✉)
Department of Electrical Engineering, National Taipei University of Technology,
NTUT Box 2125, Taipei 106, Taiwan, Republic of China
e-mail: wchang@ee.ntut.edu.tw

Sensor-based robot control has been recognized as a way to overcome many difficulties when the process model is uncertain and the environments are unknown. Thus, robots would require sensing capability to adapt themselves to the environments for a long period of time without intervention of human beings.

Theoretical and practical issues in the design, analysis, and implementation of a system which interacts with the environment using visual data observed by video cameras is becoming a major research area in robotics. A robotic control system which possesses visual feedback is referred to as a *vision-based control system*. In the early 1970s, Shirai and Inoue [1] described how *visual servoing* [2], the use of vision in the feedback loop, can increase task accuracy. Following their work, considerable research effort (cf. a tutorial on visual servoing [3]) has been devoted to the development of vision-based control systems. However, a number of factors have caused the delay of the development of such systems. Vision systems are often slow compared with robot dynamics. The analytical complexity of robot control together with the integration of vision data make the vision-based control problem challenging. Recently, video cameras are being increasingly used as robotic sensors in feedback control systems since modern computer power enables the processing rate of visual data to be high enough such that control of robots employing visual feedback is practically realizable. Consequently, many vision-based robotic positioning systems have been successfully implemented and validated by supporting experimental results [4–9].

A fundamentally important characteristic of vision-based control systems is that *both* the process state (e.g., the position of the robot in its workspace) and the reference set-point (e.g., landmark determined target position) are typically observed through the same sensors (i.e., cameras). Because of this unusual architectural feature, it is sometimes possible to achieve *precise* positioning in the absence of measurement noise, despite sensor/actuator and process model imprecision [4, 10–12], just as it is in the case of a conventional set-point control system with a loop-integrator and fixed exogenous reference. But in contrast to a set-point control system where what to choose for an error is usually clear, in vision-based systems there are many choices for errors, each with different attributes. Control tasks are thus re-defined by encoding such an error using what can really be measured. This is referred to as *task encoding*. Some of these these issues are touched upon in work by Hager [13] and in the monograph by Samson et al. [14]. The concept of task encoding is discussed briefly in the Doctoral Thesis of Rizzi [15] which contains further references. Issues concerning task encoding in vision-based control systems have been discussed in a unified framework in [16–18]. An encoded error for a given task can be described as an equation of the form $e(y)$ where y denotes a list of observed feature coordinates in such a way that the original task is accomplished with precision if $e(y) = 0$. The given task is decidable [19] if there is an encoded error which can be used to verify that the original task has been accomplished. Compared with the other two encoding approaches, Cartesian-based and modified Cartesian-based [16–18], the image-based approach encodes control tasks using only available visual data without requiring knowledge of the camera model. This advantage is to some extent offset by the image-based control design problem which to some extent is more difficult than the corresponding problems using the Cartesian-based approaches. The discussion of image-based approaches can be found in [4, 11, 20–25].

The aim of this paper is to present theoretical issues in the design and analysis of a robot set-point control system employing image-based control law for three-degree-of-freedom motion. Specifically, the control task is to *precisely* drive the robot end-effector to a visually determined target position by using an approximately calibrated binocular vision system in eye-to-hand configuration. The visual tracking system for the robot and the target features is assumed to be fast and reliable such that correct visual data can be measured. Image-based encoding is chosen for encoding an error in the eye-to-hand set-point control system. Moreover, the encoded error, if driven to zero, implies that the original 3-DOF positioning task is accomplished with precision as well. It is shown that the proposed image-based control law guarantees exponential convergence of the feedback system provided that camera modelling error is small.

The paper is organized as follows. The preliminary definitions and problem formulation are described in Section 2. Section 3 then introduces the proposed image-based control law. The stability analysis of the feedback system is introduced in Section 4. Section 5 reports the conclusion where the results and the perspectives of the research are addressed.

2 Preliminaries

2.1 Notation

Let $\measuredangle$ denote measured angle, overhead $\rightarrow$ denote ray (half-line), bold uppercase denote matrix, typewriter alphabet denote ordered list, calligraphic alphabet denote set, bold lowercase denote vector, and prime denote transpose. Moreover:

- $\mathcal{X}, \mathcal{V}, \mathcal{Y}, \mathcal{F}$: the robot workspace $\mathcal{X} \subset \mathbb{R}^3$, the binocular-camera field of view $\mathcal{V} \subset \mathbb{R}^3$, the image space $\mathcal{Y} \triangleq \mathbb{R}^2 \oplus \mathbb{R}^2$, and the composite feature space respectively;
- $\bar{\mathcal{V}}, \bar{\mathcal{Y}}$: the power sets of $\mathcal{V}$ and $\mathcal{Y}$ respectively;
- $G : \mathcal{V} \rightarrow \mathcal{Y}$: binocular-camera model;
- $\mathtt{F}$, $\mathtt{Z}$: a list of target features $\mathtt{F} \in \mathcal{F}$ and its G-image respectively;
- T, $\widetilde{T}$: a target map $T : \mathcal{F} \rightarrow \bar{\mathcal{V}}$ and the corresponding target map in image space;
- $\mathbf{r}, \mathbf{r}^*$: the position of the robot and the target in $\mathcal{X}$, $\mathbf{r}^* = T(\mathtt{F})$;
- $\mathbf{y}, \mathbf{y}^*$: the position of the robot and the target in $\mathcal{Y}$, $\mathbf{y}^* = \widetilde{T}(\mathtt{Z})$;
- $\mathbf{u} \in \mathbb{R}^3$: the robot control input;
- $\mathbf{c}_n \in \mathbb{R}^3$, f_n: the position of the optical center and focal length of camera n respectively, $n = 1, 2$;
- SO^3: Special Orthogonal group of order 3;
- $[\mathbf{i}_n, \mathbf{j}_n, \mathbf{k}_n]' \in SO^3$: the rotation matrices of camera n, $n = 1, 2$;
- I_n: the image plane of camera n, $n = 1, 2$;
- $\mathbf{o}_n$: the image centers of I_n, $n = 1, 2$;
- $\mathbf{y}_n \in \mathbb{R}^2$: the robot position in I_n, $\mathbf{y}_n \triangleq [y_{ni}, y_{nj}]'$, $n = 1, 2$;
- $\mathbf{y}_n^* \in \mathbb{R}^2$: the target position in I_n, $\mathbf{y}_n^* \triangleq \left[y_{ni}^*, y_{nj}^*\right]'$, $n = 1, 2$;
- $\bar{\mathbf{y}}_n$: visual error in I_n, $\bar{\mathbf{y}}_n \triangleq \mathbf{y}_n - \mathbf{y}_n^* = [\bar{y}_{ni}, \bar{y}_{nj}]'$, $n = 1, 2$;
- $\mathbf{w}_n$: the unit visual error vector in I_n, $n = 1, 2$.

2.2 System Description

The problem of interest is to control the position of a rigid robot in a prescribed *workspace* $\mathcal{X} \subset \mathbb{R}^3$ using data observed by two *approximately calibrated* video cameras which together comprise a binocular vision system. Specifically, it consists of driving a robot to a target position in $\mathcal{X}$ determined by a list of observed target features in the *binocular-camera field of view* $\mathcal{V} \subset \mathbb{R}^3$. The observed data consists of the robot position in $\mathcal{X}$ as well as various geometrical features of significance which appear in $\mathcal{V}$. Invariably $\mathcal{X} \subset \mathcal{V}$ and both $\mathcal{X}$ and $\mathcal{V}$ are compact subsets of $\mathbb{R}^3$. The system is illustrated in Fig. 1.

The robot is assumed to admit a simple kinematic model of the form

$$\dot{\mathbf{r}} = \mathbf{u} \tag{1}$$

where $\mathbf{r}$ is a state vector whose components are the Cartesian coordinates of the robot in $\mathcal{X}$, and $\mathbf{u}$ is a control vector taking values in $\mathbb{R}^3$. Visually identifiable landmarks are modelled by subsets of the binocular-camera field of view $\mathcal{V}$ (e.g., a line segment, a circle, etc.) and are referred to as *features*. The robot position in $\mathcal{X}$ as well as various features of interest in $\mathcal{V}$ are seen in an *image space* $\mathcal{Y} \triangleq \mathbb{R}^2 \oplus \mathbb{R}^2$ through a fixed but imprecisely known readout function or *binocular-camera model* $G : \mathcal{V} \to \mathcal{Y}$ which describes a binocular-camera vision system. Thus the robot position in image space is a measured output $\mathbf{y}$ related to $\mathbf{r}$ by the formula[1]

$$\mathbf{y} = G(\mathbf{r}). \tag{2}$$

Correspondingly, a specific feature ϕ in $\mathcal{V}$ is seen in $\mathcal{Y}$ as

$$\zeta = G(\phi). \tag{3}$$

Note that because ζ is a subset of $\mathcal{Y}$, it is also an element of the power set[2] of $\mathcal{Y}$ which we henceforth denote by $\bar{\mathcal{Y}}$. Similarly ϕ is an element of the power set of $\mathcal{V}$, which we write as $\bar{\mathcal{V}}$. Equations (1), (2) and (3) thus model the vision-based robotic system of interest.

Generally speaking one would need G to be at least injective in order for the vision subsystem to provide the requisite information for positioning. On the other hand, one would never expect to know precisely what G is. In fact, G can never be precisely identified. Accordingly, we will assume that G is a fixed but unknown element of a prescribed set $\mathcal{G}$ of injective functions each mapping $\mathcal{V}$ into $\mathcal{Y}$. It will be useful to index $\mathcal{G}$ by a parameter p taking values in some appropriate index set $\mathcal{P}$. In other words

$$\mathcal{G} = \{G_p : \quad p \in \mathcal{P}\} \tag{4}$$

and

$$G = G_{p^*} \tag{5}$$

where p^* is some fixed but unknown element of $\mathcal{P}$.

[1] With $\mathcal{A}$ and $\mathcal{B}$ sets, $\mathcal{C} \subset \mathcal{A}$, and $f : \mathcal{A} \to \mathcal{B}$, we write $f(\mathcal{C})$ for the f-image of $\mathcal{C}$ which is defined by $\{f(a) \mid a \in \mathcal{C}\}$.

[2] Given a set $\mathcal{A}$, the power set of $\mathcal{A}$ is the set of all subsets of $\mathcal{A}$.

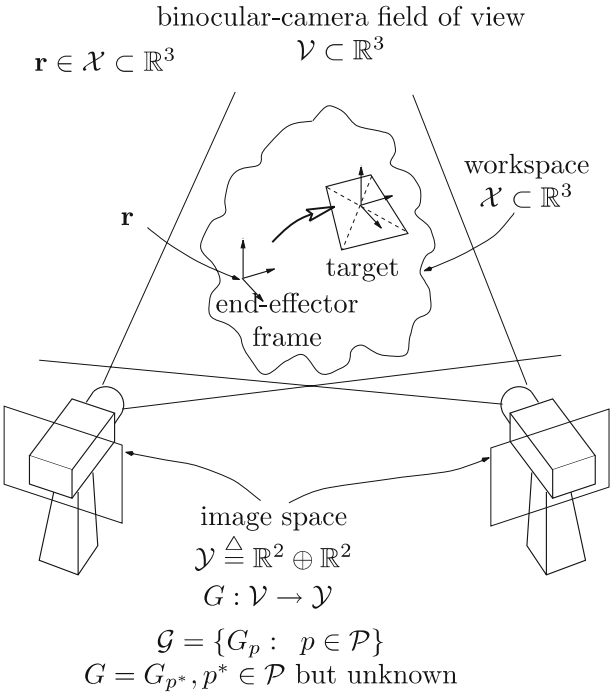

Fig. 1 The system: an example task is to drive the position of the end-effector of a robot to a target position which is the center of a table defined by the four observed corner features

For each binocular-camera model $G_p \in \mathcal{G}$, there is a naturally induced injective function $\bar{G}_p$ from $\bar{\mathcal{V}}$ to $\bar{\mathcal{Y}}$ defined by the rule $\bar{x} \longmapsto G_p(\bar{x})$. The $\bar{G}_p$ are injective because the G_p are. Henceforth, $\bar{G} \triangleq \bar{G}_\omega$. In this notation Eq. (3) can also be written as

$$\zeta = \bar{G}(\phi). \tag{6}$$

In the following sections, notation (3) is preferred, but where it may lead to confusion Eq. (6) is used instead.

A typical positioning problem for the type of system we are considering consists of driving a robot to a 'target' in $\mathcal{X}$ determined by a list of observed features in $\mathcal{V}$. In practice, targets are determined not by just one observed feature, but rather by an ordered set of n simultaneously observed features $\{\phi_1, \phi_2, \ldots, \phi_n\}$ where each ϕ_i is a member of some specific class $\mathcal{F}_i$ of like or similar objects. For example, $\mathcal{F}_1$ might be a set of line segments, $\mathcal{F}_2$ a set of cubes, etc. Henceforth, $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n$ are n such feature classes and $\mathcal{F}$ is a given *composite feature space* consisting of all lists $\{\phi_1, \phi_2, \ldots, \phi_n\}$ of interest, where each $\phi_i \in \mathcal{F}_i$.

The relationship between observed lists of features (e.g., four corner features of a square table each consisting of a single point) and targets (e.g., the center of a table defined by its four corners) is typically characterized by a given function $T : \mathcal{F} \to \bar{\mathcal{V}}$ called a *target map*. Each list of target features $\mathtt{F}$ in $\mathcal{F}$ specifies a set-point

$$\mathbf{r}^* \triangleq T(\mathtt{F}) \in \mathcal{X} \tag{7}$$

to which we would like to move the robot. In other words, for a given list of target features $\mathtt{F} = \{\phi_1, \phi_2, \ldots, \phi_n\} \in \mathcal{F}$, the *task* to be accomplished is

$$\mathbf{r} = T(\mathtt{F}). \tag{8}$$

The target position in the image space, which may or may not be computable, is an output $\mathbf{y}^*$ related to the target or desired set-point $\mathbf{r}^*$ by the formula

$$\mathbf{y}^* = G(\mathbf{r}^*). \tag{9}$$

Nevertheless, the measured list of observed target features $\mathtt{Z}$ is related to the list of target features by the formula

$$\mathtt{Z} = \widetilde{G}(\mathtt{F}) \tag{10}$$

where the map $\widetilde{G} : \mathcal{F} \to \widetilde{\mathcal{Y}} \triangleq \underbrace{\bar{\mathcal{Y}} \times \bar{\mathcal{Y}} \times \cdots \times \bar{\mathcal{Y}}}_{n \text{ times}}$ by the rule

$$\{\phi_1, \phi_2, \ldots, \phi_n\} \longmapsto \big\{G(\phi_1), G(\phi_2), \ldots, G(\phi_n)\big\}. \tag{11}$$

The state equation (1) which describes the robot kinematics is clearly controllable. However, the state $\mathbf{r}$ is not measurable whereas the G-images of various geometrical features which determine positions of the robot and the target are measurable. Nevertheless, the vision-based robotic system is observable because G is assumed to be injective. Thus, one could design a control law $\mathbf{u}$ to accomplish a positioning task. Clearly then, driving $\mathbf{r}$ to $\mathbf{r}^*$ is equivalent to driving $\mathbf{y}$ to $\mathbf{y}^*$. Were G to be known precisely (i.e., were the cameras to be precisely calibrated), then one could accomplish the latter in any one of a number of well known ways. But, since G is not presumed known with precision, neither $\mathbf{r}$ nor the features ϕ_i can be determined without error. Therefore the desired task (8) may not be achieved as it stands.

The objective of this paper is to discuss a class of tasks which can be implemented without knowing G precisely and which, if achieved, implies that the original task (8) is achieved as well. Moreover, under the assumption that robot and target features in image space can be correctly measured by a fast and reliable visual tracking system, what this paper is concerned with is how one might define control laws to to achieve precise positioning

$$\mathbf{r} \longrightarrow \mathbf{r}^* \tag{12}$$

when the camera model G is not known precisely. The control problem is thus to develop a causal algorithm, with inputs $\mathbf{y}$ and $\{G(\phi_1), G(\phi_2), \ldots, G(\phi_n)\}$, which causes the robot to achieve the positioning task exponentially.

In the following section is discussed the image-based approach to encode the task defined by Eq. (8) into tasks depending only on known information. Corresponding feedback controller is proposed based on the image-based task encoding approach with the goal of achieving *precise* positioning; i.e., satisfying Eq. (12).

3 Binocular Image-based Control with Precision

The starting point for the image-based task encoding is the requirement that the target map T be "invariant" on $\mathcal{G}$. Invariance is defined as follows.

A target map T is said to be *invariant* on $\mathcal{G}$ if there exists a function $\widetilde{T} : \widetilde{\mathcal{Y}} \to \bar{\mathcal{Y}}$ such that

$$\bar{G}_p \circ T = \widetilde{T} \circ \widetilde{G}_p, \qquad p \in \mathcal{P}. \tag{13}$$

$\widetilde{T}$ can be seen as a function that maps the binocular-camera image of a list of observed target features into the binocular-camera image of the corresponding target. It is straightforward to verify that for each $p \in \mathcal{P}$, the equation

$$\bar{G}_p \circ T = \widetilde{T}_p \circ \widetilde{G}_p \tag{14}$$

holds with

$$\widetilde{T}_p = \bar{G}_p \circ T \circ \widetilde{G}_p^{-1}. \tag{15}$$

Thus T is invariant on $\mathcal{G}$ if it is possible to find a single function $\widetilde{T}_p$, not depending on p for which Eq. (14) holds for all $G_p \in \mathcal{G}$.

To proceed, assume that the invariance property is satisfied by some computable function $\widetilde{T}$ for a certain set-point control task. To accomplish the task defined by Eq. (8), the image-based encoding approach seeks to achieve

$$\mathbf{y} = \widetilde{T}(\mathtt{Z}). \tag{16}$$

The approach is justified by the following theorem.

Theorem 3.1 *Suppose that $\widetilde{T}$ is a fixed function for which the invariance property (13) holds. For each $\mathbf{r} \in \mathcal{X}$ and each $\mathtt{F} \in \mathcal{F}$, Eq. (16) implies Eq. (8).*

Proof Since $\mathtt{Z} = \widetilde{G}(\mathtt{F})$ and $\widetilde{G} = \widetilde{G}_\omega$, we can write $\mathtt{Z} = \widetilde{G}_\omega(\mathtt{F})$. From this and Eq. (14) it follows that

$$\widetilde{T}(\mathtt{Z}) = \big[\widetilde{T} \circ \widetilde{G}_\omega\big](\mathtt{F}) = \big[\bar{G}_\omega \circ T\big](\mathtt{F}). \tag{17}$$

But $\bar{G}_\omega = \bar{G}$ and $\bar{G}(\mathtt{F}) = G(\mathtt{F})$ so

$$\widetilde{T}(\mathtt{Z}) = G\big(T(\mathtt{F})\big). \tag{18}$$

Using this and the readout equation $\mathbf{y} = G(\mathbf{r})$, it is now possible to rewrite Eq. (16) as

$$G(\mathbf{r}) = G\big(T(\mathtt{F})\big). \tag{19}$$

But G is injective because G belongs to $\mathcal{G}$ by assumption. From the injectivity of G and Eq. (19) it thus follows that Eq. (8) is true. □

Achieving the task defined by Eq. (16) clearly causes the task defined by Eq. (8) to be achieved. Clearly, the image-based approach does not require the selection of a candidate binocular-camera model G_p in $\mathcal{G}$ to serve as an estimate of G. However, in practice, designing a controller that achieves Eq. (16) may require some estimate of G. Specifically, the binocular-camera model employed in the analysis is introduced as follows.

Camera coordinate directions are established as follows: for camera 1, $\mathbf{i}_1$ points to the right and $\mathbf{j}_1$ points downward in camera 1's image plane I_1, and $\mathbf{k}_1 \triangleq \mathbf{i}_1 \times \mathbf{j}_1$ points outward along the camera optical axis. Camera coordinate directions for camera 2

are established similarly. Using *perspective projection* camera model [26], G is the nonlinear function which maps from $\mathcal{V}$ to $\mathcal{Y}$ and is defined as follows.

$$G(\mathbf{r}) \triangleq \begin{bmatrix} f_1 \frac{\mathbf{i}_1'\mathbf{r}}{\mathbf{k}_1'\mathbf{r}} \\ f_1 \frac{\mathbf{j}_1'\mathbf{r}}{\mathbf{k}_1'\mathbf{r}} \\ f_2 \frac{\mathbf{i}_2'(\mathbf{r}+\mathbf{l})}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})} \\ f_2 \frac{\mathbf{j}_2'(\mathbf{r}+\mathbf{l})}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})} \end{bmatrix}, \quad \mathbf{l} \triangleq \mathbf{c}_1 - \mathbf{c}_2 \tag{20}$$

where $\mathbf{r}$ is the robot position in $\mathcal{X}$ relative to $\mathbf{c}_1$.

Differentiating Eq. (2) with respect to time, we have

$$\dot{\mathbf{y}} = \mathbf{J}(\mathbf{r})\,\mathbf{u} \tag{21}$$

where the Jacobian of the nonlinear map G is defined as

$$\mathbf{J}(\mathbf{r}) \triangleq \frac{\partial G(\mathbf{r})}{\partial \mathbf{r}} = \begin{bmatrix} \frac{f_1}{\mathbf{k}_1'\mathbf{r}} \left(\mathbf{i}_1' - \frac{\mathbf{i}_1'\mathbf{r}}{\mathbf{k}_1'\mathbf{r}} \mathbf{k}_1' \right) \\ \frac{f_1}{\mathbf{k}_1'\mathbf{r}} \left(\mathbf{j}_1' - \frac{\mathbf{j}_1'\mathbf{r}}{\mathbf{k}_1'\mathbf{r}} \mathbf{k}_1' \right) \\ \frac{f_2}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})} \left(\mathbf{i}_2' - \frac{\mathbf{i}_2'(\mathbf{r}+\mathbf{l})}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})} \mathbf{k}_2' \right) \\ \frac{f_2}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})} \left(\mathbf{j}_2' - \frac{\mathbf{j}_2'(\mathbf{r}+\mathbf{l})}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})} \mathbf{k}_2' \right) \end{bmatrix}. \tag{22}$$

From the invariance property (13) and the readout equation $\mathtt{Z} = \widetilde{G}(\mathtt{F})$, one can see that

$$\mathbf{y}^* \triangleq \widetilde{T}(\mathtt{Z}) = \left[\widetilde{T} \circ \widetilde{G}\right](\mathtt{F}) = \left[G \circ T\right](\mathtt{F}) = G(\mathbf{r}^*). \tag{23}$$

Based on the image-based encoding approach (16), the encoded error for the positioning task (8) is defined as

$$\bar{\mathbf{y}} = \mathbf{y} - \widetilde{T}(\mathtt{Z}) \tag{24}$$

where $\bar{\mathbf{y}}$ can be computed based only on available information. Thus, one can design image-based control laws to assure exponential convergence of vision-based control systems. Were G to be known precisely (i.e., were the cameras to be precisely calibrated), with $\mathbf{Q}(\cdot)_{3\times 3}$ a symmetric positive definite matrix[2], G^{-1} a continuously differentiable left inverse of G, and $\eta(\mathbf{y}, \mathtt{Z}) \triangleq \frac{\mathrm{v}_{\max}}{\max\left\{ \left\| \left(\left[\mathbf{Q} \circ G^{-1}\right](\mathbf{y}) \right) \cdot \left(\left[\mathbf{J} \circ G^{-1}\right](\mathbf{y}) \right)' \cdot \bar{\mathbf{y}} \right\|, \mathrm{v}_{\max} \right\}}$ a saturation function ($\mathrm{v}_{\max}$ is the maximum speed of the robot), the feedback law

$$\mathbf{u} = -\eta(\mathbf{y}, \mathtt{Z}) \cdot \left(\left[\mathbf{Q} \circ G^{-1}\right](\mathbf{y}) \right) \cdot \left(\left[\mathbf{J} \circ G^{-1}\right](\mathbf{y}) \right)' \cdot \bar{\mathbf{y}} \tag{25}$$

would suffice. In fact, the control law (25) can drive $\bar{\mathbf{y}}$ to zero exponentially.

[2]For example, $\left[\mathbf{Q} \circ G^{-1}\right](\mathbf{y})$ can be chosen as the commonly used pseudo-inverse $\left(\left(\left[\mathbf{J} \circ G^{-1}\right](\mathbf{y}) \right)' \left(\left[\mathbf{J} \circ G^{-1}\right](\mathbf{y}) \right) \right)^{-1}$.

Even if G were modelled only approximately by some perspective projection function $G_q : \mathcal{V} \to \mathcal{Y}$, with G_q^{-1} a continuously differentiable left inverse of G_q, $\mathbf{J}_q(r) \triangleq \frac{\partial G_q(\mathbf{r})}{\partial \mathbf{r}}$, and $\hat{\eta}(\mathbf{y}, \mathtt{Z}) \triangleq \frac{v_{\max}}{\max\left\{\left\| \left([\mathbf{Q} \circ G_q^{-1}](\mathbf{y})\right) \cdot \left([\mathbf{J}_q \circ G_q^{-1}](\mathbf{y})\right)' \cdot \bar{\mathbf{y}} \right\|, v_{\max}\right\}}$, the feedback law

$$\mathbf{u} = -\hat{\eta}(\mathbf{y}, \mathtt{Z}) \cdot \left(\left[\mathbf{Q} \circ G_q^{-1} \right](\mathbf{y}) \right) \cdot \left(\left[\mathbf{J}_q \circ G_q^{-1} \right](\mathbf{y}) \right)' \cdot \bar{\mathbf{y}} \tag{26}$$

would still achieve precise positioning exponentially provided that G_q were a good enough approximate model of G.

4 Stability Analysis

4.1 Camera Exactly Calibrated

The feedback law (25), in which the camera model is assumed known exactly, is employed to the system described by Eqs. (1), (2), and (10). Exponential convergence and precise positioning can be guaranteed. Although one would never expect to know precisely what G is, the results presented in this subsection can be further used in the stability analysis for the case when cameras are only imprecisely calibrated.

The algebraic equality stated in Theorem 4.1, whose proof is given in the Appendix, enables us to explore the *binocular visual constraint* addressed in Lemma 4.1.

Theorem 4.1 *Let the positions of the robot and the target in the image plane I_n of camera n be defined respectively as follows, $n = 1, 2$.*

$$\mathbf{y}_n = \begin{bmatrix} y_{ni} \\ y_{nj} \end{bmatrix}, \quad \mathbf{y}_n^* = \begin{bmatrix} y_{ni}^* \\ y_{nj}^* \end{bmatrix} \tag{27}$$

Denote the visual error in the image plane I_n of camera n by

$$\bar{\mathbf{y}}_n = \mathbf{y}_n - \mathbf{y}_n^* = \begin{bmatrix} \bar{y}_{ni} \\ \bar{y}_{nj} \end{bmatrix}, n = 1, 2. \tag{28}$$

In the light of Eqs. (23) and (24), the encoded error for the positioning task can be written as

$$\bar{\mathbf{y}} = \mathbf{y} - \mathbf{y}^* = \begin{bmatrix} \bar{\mathbf{y}}_1 \\ \bar{\mathbf{y}}_2 \end{bmatrix}. \tag{29}$$

The following algebraic equality can be established.

$$\mathbf{J}'(\mathbf{r})\bar{\mathbf{y}} = \begin{bmatrix} \frac{1}{\mathbf{k}_1'\mathbf{r}} (-\alpha_1 \mathbf{k}_1 + f_1 \mathbf{w}_1) & \frac{1}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})} (-\alpha_2 \mathbf{k}_2 + f_2 \mathbf{w}_2) \end{bmatrix} \begin{bmatrix} \|\bar{\mathbf{y}}_1\| \\ \|\bar{\mathbf{y}}_2\| \end{bmatrix} \tag{30}$$

where

$$\mathbf{w}_1 \triangleq \frac{\bar{y}_{1i}\mathbf{i}_1 + \bar{y}_{1j}\mathbf{j}_1}{\|\bar{y}_{1i}\mathbf{i}_1 + \bar{y}_{1j}\mathbf{j}_1\|}, \quad \mathbf{w}_2 \triangleq \frac{\bar{y}_{2i}\mathbf{i}_2 + \bar{y}_{2j}\mathbf{j}_2}{\|\bar{y}_{2i}\mathbf{i}_2 + \bar{y}_{2j}\mathbf{j}_2\|}, \tag{31}$$

and

$$\alpha_n \triangleq \|\bar{\mathbf{y}}_n\| + \|\mathbf{y}_n^*\| \cos\theta_n, \ \theta_n \triangleq \angle(\overrightarrow{\mathbf{y}_n^*\mathbf{y}_n}, \overrightarrow{\mathbf{o}_n\mathbf{y}_n^*}), \ n = 1, 2. \tag{32}$$

Lemma 4.1 states an important algebraic constraint which is used in the proof of Theorem 4.2 to show that the only equilibrium state is the state corresponding to zero positioning error.

Lemma 4.1 (Binocular visual constraint) *Let $\mathcal{A}$ be any set of points in $\mathbb{R}^3$ on which G is bounded and injective. Let $\mathbf{r}^* \in \mathcal{A}$. Then, for every $\mathbf{r} \in \mathcal{A}$*

$$\mathbf{J}'(\mathbf{r})\big(G(\mathbf{r}) - G(\mathbf{r}^*)\big) = 0 \tag{33}$$

implies

$$\mathbf{r} = \mathbf{r}^*. \tag{34}$$

Proof By the injectivity of G, $\mathbf{r} \neq \mathbf{r}^*$ implies $G(\mathbf{r}) \neq G(\mathbf{r}^*)$. We will show that $G(\mathbf{r}) \neq G(\mathbf{r}^*)$ implies $\mathbf{J}'(\mathbf{r})\big(G(\mathbf{r}) - G(\mathbf{r}^*)\big) \neq 0$.

Defining

$$\mathbf{m}_1 \triangleq \frac{1}{\mathbf{k}_1'\mathbf{r}}\left(-\alpha_1\mathbf{k}_1 + f_1\mathbf{w}_1\right), \qquad \mathbf{m}_2 \triangleq \frac{1}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})}\left(-\alpha_2\mathbf{k}_2 + f_2\mathbf{w}_2\right), \tag{35}$$

$$\mathbf{M}_{3\times 2} \triangleq \left[\,\mathbf{m}_1\ \mathbf{m}_2\,\right], \quad \text{and} \quad \mathbf{h}_{2\times 1} \triangleq \left[\,\|\bar{\mathbf{y}}_1\|\|\bar{\mathbf{y}}_2\|\,\right], \tag{36}$$

it follows from Eq. (30) that

$$\mathbf{J}'(\mathbf{r})\bar{\mathbf{y}} = \mathbf{M}\cdot\mathbf{h}. \tag{37}$$

By virtue of Eq. (37) and given that $G(\mathbf{r}) \neq G(\mathbf{r}^*)$, $\mathbf{J}'(\mathbf{r})\big(G(\mathbf{r}) - G(\mathbf{r}^*)\big) = 0$ if and only if

$$\|\bar{\mathbf{y}}_1\|\,\mathbf{m}_1 + \|\bar{\mathbf{y}}_2\|\,\mathbf{m}_2 = 0. \tag{38}$$

Geometrically, condition (38) is equivalent to saying that 3-D vectors $\mathbf{m}_1$ and $\mathbf{m}_2$ are parallel and are pointing along opposite directions.

In the following proof, we will show that Eq. (38) is impossible to hold given that $\bar{\mathbf{y}} \neq 0$. Thus, we have $\bar{\mathbf{y}} \neq 0 \Rightarrow \mathbf{J}'(\mathbf{r})\bar{\mathbf{y}} \neq 0$. The proof below is based on epipolar geometry as illustrated in Fig. 2).

Let $\mathbf{y}_n \triangleq y_{ni}\mathbf{i}_n + y_{nj}\mathbf{j}_n$, a 3-D point in space. According to *epipolar constraint* [27, 28], if $\mathbf{y}_1$ and $\mathbf{y}_2$ correspond to the same physical point in space, $\mathbf{c}_1$, $\mathbf{c}_2$, $\mathbf{y}_1$, and $\mathbf{y}_2$ must

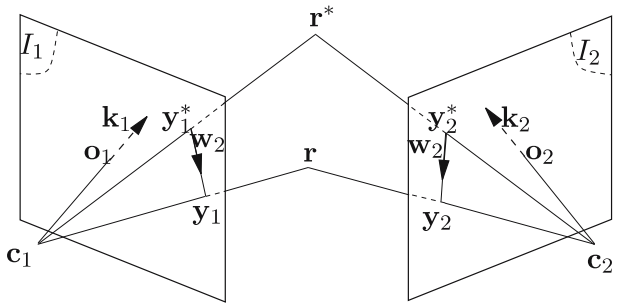

Fig. 2 Epipolar geometry

be in a single plane Π (known as *epipolar plane*). Therefore, $\overrightarrow{\mathbf{c}_1\mathbf{y}_1}$, $\overrightarrow{\mathbf{c}_2\mathbf{y}_2}$, and $\overline{\mathbf{c}_1\mathbf{c}_2}$ are in the same plane.

Consider the rays $\overrightarrow{\mathbf{c}_n\mathbf{y}_n}(n = 1, 2)$ which can be defined by the optical center $\mathbf{c}_n$ and the vectors $f_n\mathbf{k}_n + \alpha_n\mathbf{w}_n + (\|\mathbf{y}_n\|^2 - \alpha_n^2)^{\frac{1}{2}}\mathbf{v}_n$, where $\mathbf{v}_n \triangleq \mathbf{w}_n \times \mathbf{k}_n$.

Since

$$\left(f_n\mathbf{k}_n + \alpha_n\mathbf{w}_n + (\|\mathbf{y}_n\|^2 - \alpha_n^2)^{\frac{1}{2}}\mathbf{v}_n\right)'(-\alpha_n\mathbf{k}_n + f_n\mathbf{w}_n) = 0\ ,\ n = 1, 2 \tag{39}$$

we see that $\overrightarrow{\mathbf{c}_n\mathbf{y}_n} \perp \mathbf{m}_n$.

From Eq. (38), we further see that $\overrightarrow{\mathbf{c}_1\mathbf{y}_1} \perp \mathbf{m}_2$ and $\overrightarrow{\mathbf{c}_2\mathbf{y}_2} \perp \mathbf{m}_1$. That is,

$$\begin{cases} \left(f_1\mathbf{k}_1 + \alpha_1\mathbf{w}_1 + (\|\mathbf{y}_1\|^2 - \alpha_1^2)^{\frac{1}{2}}\mathbf{v}_1\right)'(-\alpha_2\mathbf{k}_2 + f_2\mathbf{w}_2) = 0 \\ \left(f_2\mathbf{k}_2 + \alpha_2\mathbf{w}_2 + (\|\mathbf{y}_2\|^2 - y_2^2)^{\frac{1}{2}}\mathbf{v}_2\right)'(-\alpha_1\mathbf{k}_1 + f_1\mathbf{w}_1) = 0 \end{cases}. \tag{40}$$

Meanwhile,

$$\begin{cases} (-\alpha_1\mathbf{k}_1 + f_1\mathbf{w}_1)'\mathbf{w}_1 = f_1 > 0 \\ (-\alpha_1\mathbf{k}_1 + f_1\mathbf{w}_1)'\mathbf{w}_2 < 0^{\dagger} \end{cases}. \tag{41}$$

From Eqs. (39), (40), and the definition of Π, Π must contain the baseline ($\overline{\mathbf{c}_1\mathbf{c}_2}$) and is normal to the vector $-\alpha_1\mathbf{k}_1 + f_1\mathbf{w}_1$ and $-\alpha_2\mathbf{k}_2 + f_2\mathbf{w}_2$. Thus, we have

$$(-\alpha_n\mathbf{k}_n + f_n\mathbf{w}_n)'\mathbf{l} = 0\ ,\ i = 1, 2. \tag{42}$$

Similarly, the rays $\overrightarrow{\mathbf{c}_n\mathbf{y}_n^*}(n = 1, 2)$ can be defined by the optical center c_n and the vectors $f_n\mathbf{k}_n + \alpha_n\mathbf{w}_n + (\|\mathbf{y}_n\|^2 - \alpha_n^2)^{\frac{1}{2}}\mathbf{v}_n - \|\bar{\mathbf{y}}_n\|\mathbf{w}_n$.

Assume that the target position $\mathbf{r}^*$ is well defined and its image coordinates are $\mathbf{y}^*$. That is, the rays $\overrightarrow{\mathbf{c}_1\mathbf{y}_1^*}$ and $\overrightarrow{\mathbf{c}_2\mathbf{y}_2^*}$ intersect at a unique 3-D point. Thus, there exist $k_1, k_2 \in \mathbb{R}^+$ such that

$$\begin{aligned} c_1 + k_1(f_1\mathbf{k}_1 + \alpha_1\mathbf{w}_1 + (\|\mathbf{y}_1\|^2 - \alpha_1^2)^{\frac{1}{2}}\mathbf{v}_1 - \|\bar{\mathbf{y}}_1\|\mathbf{w}_1) \\ = c_2 + k_2(f_2\mathbf{k}_2 + \alpha_2\mathbf{w}_2 + (\|\mathbf{y}_2\|^2 - \alpha_2^2)^{\frac{1}{2}}\mathbf{v}_2 - \|\bar{\mathbf{y}}_2\|\mathbf{w}_2). \end{aligned} \tag{43}$$

Multiplying Eq. (43) by $(-\alpha_1\mathbf{k}_1 + f_1\mathbf{w}_1)'$ and using Eqs. (39), (40), and (42), we have

$$k_1\|\bar{\mathbf{y}}_1\|(-\alpha_1\mathbf{k}_1 + f_1\mathbf{w}_1)'\mathbf{w}_1 = k_2\|\bar{\mathbf{y}}_2\|(-\alpha_1\mathbf{k}_1 + f_1\mathbf{w}_1)'\mathbf{w}_2. \tag{44}$$

From Eq. (41), we know that Eq. (44) cannot be true. In other words, a unique 3-D point cannot be defined. Thus, condition (38) cannot hold given that $\bar{\mathbf{y}} \neq 0$. This finishes the proof by contradiction. □

Lemma 4.2 states an important property which will be used in the proof of Theorem 4.2 to show exponential convergence.

[†]Note that $(-\alpha_2\mathbf{k}_2 + f_2\mathbf{w}_2)'\mathbf{w}_2 = f_2 > 0$. $\mathbf{k}_1'\mathbf{r} > 0, \mathbf{k}_2'(\mathbf{r} + \mathbf{l}) > 0$ since the robot is assumed to be in front of the cameras and is in eye-to-hand configuration.

Lemma 4.2 *There exists a positive number μ which depends only on $\mathcal{S}$ such that for every $\mathbf{r}, \mathbf{r}^* \in \mathcal{S}$*

$$||\mathbf{J}'(\mathbf{r})\big(G(\mathbf{r}) - G(\mathbf{r}^*)\big)|| \geq \mu ||G(\mathbf{r}) - G(\mathbf{r}^*)||. \tag{45}$$

Proof As shown in the proof of Lemma 4.1, it is impossible for $\mathbf{m}_1$ and $\mathbf{m}_2$ to be parallel and pointing along opposite directions (i.e., $\mathbf{m}_2 = \gamma\, \mathbf{m}_1$, for some $\gamma < 0$). Hence, we know that $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2]$ can only loose rank when $\mathbf{m}_2 = \gamma\, \mathbf{m}_1$, for some $\gamma > 0$. We thus consider two cases.

– When $\mathbf{m}_1' \mathbf{m}_2 > 0$:
In this case, $\mathbf{M}$ may loose rank or not. Let $\mathbf{m}_2 = \gamma_1\, \mathbf{m}_1 + \mathbf{m}_1^\perp$, where $\gamma_1 = \frac{\mathbf{m}_2' \mathbf{m}_1}{||\mathbf{m}_1||^2} > 0$ and $\mathbf{m}_1^\perp = \mathbf{m}_2 - \gamma_1\, \mathbf{m}_1$. We have

$$\begin{aligned} ||\mathbf{J}'(\mathbf{r})\bar{\mathbf{y}}||^2 &= \left\| (||\bar{\mathbf{y}}_1|| + \gamma_1\, ||\bar{\mathbf{y}}_2||)\mathbf{m}_1 + ||\bar{\mathbf{y}}_2||\, \mathbf{m}_1^\perp \right\|^2 \\ &= (||\bar{\mathbf{y}}_1|| + \gamma_1\, ||\bar{\mathbf{y}}_2||)^2 ||\mathbf{m}_1||^2 + ||\bar{\mathbf{y}}_2||^2\, ||\mathbf{m}_1^\perp||^2 \\ &\geq ||\bar{\mathbf{y}}_1||^2\, ||\mathbf{m}_1||^2 \\ &= \frac{\alpha_1^2 + f_1^2}{(\mathbf{k}_1' \mathbf{r})^2}\, ||\bar{\mathbf{y}}_1||^2. \end{aligned} \tag{46}$$

Similarly let $\mathbf{m}_1 = \gamma_2\, \mathbf{m}_2 + \mathbf{m}_2^\perp$, where $\gamma_2 = \frac{\mathbf{m}_1' \mathbf{m}_2}{||\mathbf{m}_2||^2} > 0$ and $\mathbf{m}_2^\perp = \mathbf{m}_1 - \gamma_2\, \mathbf{m}_2$. We have

$$\begin{aligned} ||\mathbf{J}'(\mathbf{r})\bar{\mathbf{y}}||^2 &= \left\| (\gamma_2\, ||\bar{\mathbf{y}}_1|| + ||\bar{\mathbf{y}}_2||)\mathbf{m}_2 + ||\bar{\mathbf{y}}_1||\, \mathbf{m}_2^\perp \right\|^2 \\ &= (\gamma_2\, ||\bar{\mathbf{y}}_1|| + ||\bar{\mathbf{y}}_2||)^2 ||\mathbf{m}_2||^2 + ||\bar{\mathbf{y}}_1||^2\, ||\mathbf{m}_2^\perp||^2 \\ &\geq ||\bar{\mathbf{y}}_2||^2\, ||\mathbf{m}_2||^2 \\ &= \frac{\alpha_2^2 + f_2^2}{(\mathbf{k}_2'(\mathbf{r} + \mathbf{l}))^2}\, ||\bar{\mathbf{y}}_2||^2. \end{aligned} \tag{47}$$

Therefore, from Eqs. (46) and (47), we see that

$$||\mathbf{J}'(\mathbf{r})\bar{\mathbf{y}}||^2 \geq \mu_1\, ||\bar{\mathbf{y}}||^2 \tag{48}$$

where $\mu_1 = \frac{1}{2} \min\limits_{\mathbf{r} \in \mathcal{S}} \left\{ \frac{\alpha_1^2 + f_1^2}{(\mathbf{k}_1' \mathbf{r})^2}, \frac{\alpha_2^2 + f_2^2}{(\mathbf{k}_2'(\mathbf{r} + \mathbf{l}))^2} \right\} > 0.$

– When $\mathbf{m}_1' \mathbf{m}_2 \leq 0$:
In this case, $\mathbf{M}$ is of full rank. Therefore,

$$||\mathbf{J}'(\mathbf{r})\bar{\mathbf{y}}||^2 = ||\mathbf{M}\, \mathbf{h}||^2 \geq \sigma_{\mathbf{M}}^2\, ||\bar{\mathbf{y}}||^2 \tag{49}$$

where $\sigma_{\mathbf{M}}$ is the least singular value of $\mathbf{M}$ when $\mathbf{m}_1' \mathbf{m}_2 \leq 0$. That is, $\sigma_{\mathbf{M}} > 0$.

Therefore, from Eqs. (48) and (49), we have

$$||\mathbf{J}'(\mathbf{r})\bar{\mathbf{y}}||^2 \geq \mu\, ||\bar{\mathbf{y}}||^2 \tag{50}$$

where $\mu = \min\left\{\mu_1, \sigma_{\mathbf{M}}^2\right\} > 0$. □

By virtue of Lemmas 4.1 and 4.2, the exponential stability of the closed-loop system using exactly calibrated cameras can be established as follows.

Theorem 4.2 (Exponential stability) *Let* $\mathcal{B} \triangleq \{\mathbf{y} \in G(\mathcal{X}) \mid \|\mathbf{y} - \mathbf{y}^*\| \leq \rho\}$ *such that* $\mathcal{S} \triangleq \{\mathbf{r} \in \mathcal{X} \mid G(\mathbf{r}) \in \mathcal{B}\}$ *is bounded and* G *is injective on* $\mathcal{S}$. *For the system defined by Eqs. (1), (2), (10), and (25), there exist positive numbers* a, λ *which depend only on* $\mathcal{S}$ *such that if* $\mathbf{r}(0), \mathbf{r}^* \in \mathcal{S}$,

$$\|\mathbf{r}(t) - \mathbf{r}^*\| \leq a \cdot e^{-\lambda t} \cdot \|\mathbf{r}(0) - \mathbf{r}^*\|, \ \forall t \geq 0.$$

Proof Define a Lyapunov function

$$V = ||\bar{\mathbf{y}}||^2 = \bar{\mathbf{y}}'\bar{\mathbf{y}}. \tag{51}$$

In the light of Eqs. (21) and (25), the differentiation of V with respect to time can be seen as follow.

$$\dot{V} = -2\eta(\mathbf{y}, \mathbb{Z}) \cdot ||\mathbf{Q}^{\frac{1}{2}}(\mathbf{r})\, \mathbf{J}'(\mathbf{r})\, \bar{\mathbf{y}}||^2 \leq 0. \tag{52}$$

From Eq. (52) we see that V is a nonincreasing function. This, together with Eq. (51), implies that

$$\|\mathbf{y}(t) - \mathbf{y}^*\| \leq \|\mathbf{y}(0) - \mathbf{y}^*\|, \forall t \geq 0. \tag{53}$$

This is equivalent to saying that $\mathbf{y}$ stays inside the ball $\mathcal{B}$. In other words, solution $\mathbf{r}$ never leaves $\mathcal{S}$. Thus, $\mathcal{S}$ is an bounded invariant set. Hence, $\mathbf{r}$ is bounded. Meanwhile, for all $t \geq 0$

$$\int_0^t |\dot{V}(\tau)| d\tau = -\int_0^t \dot{V}(\tau) d\tau = V(0) - V(t) \leq V(0). \tag{54}$$

We see that $\dot{V} \in \mathcal{L}_1$. Furthermore, one can check that $\ddot{V}$ is piecewise continuous and bounded. Therefore, we conclude that

$$\dot{V} \longrightarrow 0 \tag{55}$$

and thus

$$\mathbf{J}'(\mathbf{r})\, \bar{\mathbf{y}} \longrightarrow 0 \tag{56}$$

since $\mathbf{Q}(\mathbf{r})$ is a symmetric positive definite matrix. From Lemma 4.1 we know that if $\bar{\mathbf{y}}$ is in the null space of $\mathbf{J}'(\mathbf{r})$ (i.e., Eq. (33) holds), it must be true that $\mathbf{r} = \mathbf{r}^*$ for all $\mathbf{r}, \mathbf{r}^* \in \mathcal{S}$. Therefore, we see that

$$\mathbf{r} \longrightarrow \mathbf{r}^*. \tag{57}$$

Moreover, from Eq. (52) and Lemma 4.2, we have

$$\dot{V} \leq -2\eta(\mathbf{y}, \mathbb{Z})\sigma^2 ||\mathbf{J}'(\mathbf{r})\, \bar{\mathbf{y}}||^2 \leq -2\eta(\mathbf{y}, \mathbb{Z})\sigma^2\, \mu\, V \tag{58}$$

where $\sigma > 0$ is the least singular value of $\mathbf{Q}^{\frac{1}{2}}(\mathbf{r}), \mathbf{r} \in \mathcal{S}$.

Therefore, $||\bar{\mathbf{y}}||$ goes to zero exponentially fast. By injectivity of G on $\mathcal{S}$, one thus conclude that there exist positive numbers a, λ which depend only on $\mathcal{S}$ such that if $\mathbf{r}(0), \mathbf{r}^* \in \mathcal{S}$,

$$\|\mathbf{r}(t) - \mathbf{r}^*\| \leq a \cdot e^{-\lambda t} \cdot \|\mathbf{r}(0) - \mathbf{r}^*\|, \ \forall t \geq 0. \tag{59}$$

□

4.2 Camera Almost Exactly Calibrated

The feedback law (26), in which the camera model is assumed known approximately, is employed to te system described by Eqs. (1), (2), and (10). Theorem 4.3 states that if the approximate camera model is close enough to the true model and the initial positioning error is small enough, the introduced image-based control law can still guarantee exponential convergence and precise positioning.

Theorem 4.3 (Robustness) *Let $\tilde{\eta}$ be a positive number defined as*

$$\tilde{\eta} \triangleq \min_{\mathbf{y}\in\mathcal{B}} \left\{\hat{\eta}(\mathbf{y}, \mathtt{Z})\right\}. \tag{60}$$

Suppose that there exists a positive number $\nu < \tilde{\eta}\sigma^2\mu$ and such that for every $\mathbf{y} \in \mathcal{B}$

$$\left\|\boldsymbol{\Delta}(\mathbf{y})\right\| \leq \nu \tag{61}$$

where

$$\boldsymbol{\Delta}(\mathbf{y}) \triangleq \Big(\big[\mathbf{J}\circ G^{-1}\big](\mathbf{y})\Big)\cdot\Big(\Big(\big[\mathbf{Q}\circ G^{-1}\big](\mathbf{y})\Big)\cdot\Big(\big[\mathbf{J}\circ G^{-1}(\mathbf{y})\big)' - \Big(\Big[\mathbf{Q}\circ G_q^{-1}\Big](\mathbf{y})\Big)\cdot\Big(\Big[\mathbf{J}_q\circ G_q^{-1}\Big](\mathbf{y})\Big)'\Big). \tag{62}$$

The solution to the system defined by Eqs. (1), (2), (10), and (26) exists globally and $\mathbf{r} \longrightarrow \mathbf{r}^*$ exponentially if $\|\bar{\mathbf{y}}(0)\| \leq h\,\rho$ for every $h \in [0, 1)$.

Proof From Eqs. (21), (26), and (62), we have

$$\dot{\bar{\mathbf{y}}} = \hat{\eta}(\mathbf{y}, \mathtt{Z})\cdot\mathbf{J}(\mathbf{r})\cdot\left(\left[\mathbf{Q}\circ G^{-1}\right](\mathbf{y})\right)\cdot\left(\left[\mathbf{J}\circ G^{-1}\right](\mathbf{y})\right)'\bar{\mathbf{y}} + \hat{\eta}(\mathbf{y}, \mathtt{Z})\cdot\boldsymbol{\Delta}(\mathbf{y})\cdot\bar{\mathbf{y}}. \tag{63}$$

In the light of Eq. (63), the differentiation of V with respect to time can be seen as follow.

$$\dot{V} = -2\hat{\eta}(\mathbf{y}, \mathtt{Z})\cdot\|\mathbf{Q}^{\frac{1}{2}}(\mathbf{r})\,\mathbf{J}'(\mathbf{r})\,\bar{\mathbf{y}}\|^2 + 2\hat{\eta}(\mathbf{y}, \mathtt{Z})\bar{\mathbf{y}}'\boldsymbol{\Delta}(\mathbf{y})\bar{\mathbf{y}}. \tag{64}$$

By differentiating Eq. (51) with respect to time, we can also see that

$$\dot{V} = 2\|\bar{\mathbf{y}}\|\frac{d\|\bar{\mathbf{y}}\|}{dt}. \tag{65}$$

Therefore, from Eq. (60) and the facts that for every $\mathbf{y} \in \mathcal{B}$

$$\|\mathbf{Q}^{\frac{1}{2}}(\mathbf{r})\,\mathbf{J}'(\mathbf{r})\,\bar{\mathbf{y}}\|^2 \geq \sigma^2\mu\|\bar{\mathbf{y}}\|^2 \tag{66}$$

and

$$\hat{\eta}(\mathbf{y}, \mathtt{Z}) \in (0, 1], \tag{67}$$

it follows from Eqs. (64) and (65) that

$$\frac{d\|\bar{\mathbf{y}}(t)\|}{dt} \leq -\tilde{\eta}\sigma^2\mu\|\bar{\mathbf{y}}(t)\| + \left\|\boldsymbol{\Delta}\big(\mathbf{y}(t)\big)\right\|\cdot\|\bar{\mathbf{y}}(t)\|. \tag{68}$$

Hence, by using the variation of constant formula one concludes that

$$\|\bar{\mathbf{y}}(t)\| \leq \|\bar{\mathbf{y}}(0)\|e^{-\tilde{\eta}\sigma^2\mu t} + \int_0^t e^{-\tilde{\eta}\sigma^2\mu(t-\tau)}\left\|\boldsymbol{\Delta}\big(\mathbf{y}(\tau)\big)\right\|\cdot\|\bar{\mathbf{y}}(\tau)\|d\tau. \tag{69}$$

Multiplying both sides by $e^{\tilde{\eta}\sigma^2\mu t}$, we have

$$e^{\tilde{\eta}\sigma^2\mu t}\|\bar{\mathbf{y}}(t)\| \leq \|\bar{\mathbf{y}}(0)\| + \int_0^t e^{\tilde{\eta}\sigma^2\mu\tau}\left\|\boldsymbol{\Delta}\big(\mathbf{y}(\tau)\big)\right\| \cdot \|\bar{\mathbf{y}}(\tau)\| d\tau \tag{70}$$

Applying the Bellman–Gronwall Lemma [29] that is listed in the Appendix, one further concludes that

$$e^{\tilde{\eta}\sigma^2\mu t}\|\bar{\mathbf{y}}(t)\| \leq \|\bar{\mathbf{y}}(0)\| e^{\int_0^t \|\boldsymbol{\Delta}(\mathbf{y}(\tau))\| d\tau}. \tag{71}$$

Therefore,

$$\|\bar{\mathbf{y}}(t)\| \leq \|\bar{\mathbf{y}}(0)\| e^{-\int_0^t (\tilde{\eta}\sigma^2\mu - \|\boldsymbol{\Delta}(\mathbf{y}(\tau))\|) d\tau}, \ \forall\, t \geq 0. \tag{72}$$

Henceforth, we will show that there exists a neighborhood of the origin such that if $\bar{\mathbf{y}}(0)$ starts in that neighborhood, $\bar{\mathbf{y}}(t) \to 0$ as $t \to \infty$. We claim that if $\|\bar{\mathbf{y}}(0)\| \leq h\rho$, for every $h \in [0, 1)$, $\bar{\mathbf{y}} \to 0$ as $t \to \infty$.

We first show that if $\bar{\mathbf{y}}(0) \leq h\rho$, $\|\bar{\mathbf{y}}(t)\| < \rho$ for all $t \geq 0$. By contradiction assume that there exists some time for which $\|\bar{\mathbf{y}}(t)\|$ gets to be larger or equal to ρ. Let t^* be the first instant for which $\|\bar{\mathbf{y}}(t)\|$ crosses ρ, i.e.

$$\|\bar{\mathbf{y}}(t^*)\| = \rho \quad \text{and} \quad \|\bar{\mathbf{y}}(t)\| < \rho, \ \forall\, t \in [0, t^*). \tag{73}$$

Therefore, one can see that

$$\left\|\boldsymbol{\Delta}\big(\mathbf{y}(\tau)\big)\right\| \leq \nu, \ \forall\, t \in [0, t^*). \tag{74}$$

Using this fact and Eq. (72), one concludes that

$$\|\bar{\mathbf{y}}(t^*)\| \leq \|\bar{\mathbf{y}}(0)\| e^{-\int_0^{t^*} (\tilde{\eta}\sigma^2\mu - \nu) d\tau} \leq \|\bar{\mathbf{y}}(0)\| \leq h\rho < \rho \tag{75}$$

which contradicts the definition of t^*.

Having proved that $\|\bar{\mathbf{y}}(t)\| < \rho$ for all $t \geq 0$, we will show that actually $\bar{\mathbf{y}} \to 0$ as $t \to \infty$. To do this, note once again that

$$\|\bar{\mathbf{y}}(t)\| < \rho, \forall\, t \geq 0 \quad \Longrightarrow \quad \left\|\boldsymbol{\Delta}\big(\mathbf{y}(\tau)\big)\right\| \leq \nu, \ \forall\, t \geq 0. \tag{76}$$

Using this together with Eq. (72) and the fact that $\nu < \tilde{\eta}\sigma^2\mu$, one concludes that

$$\|\bar{\mathbf{y}}(t)\| \leq \|\bar{\mathbf{y}}(0)\| e^{-\int_0^t (\tilde{\eta}\sigma^2\mu - \nu) d\tau} \to 0 \ \text{ as } t \to \infty. \tag{77}$$

□

Theorem 4.3 has verified the fact that exponential stability of an image-based visual feedback system in eye-to-hand configuration can be assured provided that the binocular model imprecision is small. Moreover, three-degree-of-freedom positioning tasks, if satisfying the invariance property (13), can be accomplished with precision using the proposed image-based control law even if the camera calibration is not precisely known.

5 Conclusion

Numerous vision-based robotic positioning systems have been successfully implemented and validated by supporting experimental results. Nevertheless, the aim of

the research is to provide stability analysis for a class of robotic set-point control systems employing image-based feedback laws. In particular, as long as invariant target maps in image space can be determined for the considered control tasks, exponential stability of the feedback systems employing the image-based control law can be assured.

In this paper, we have proposed a class of image-based control laws for the visual servoing of a rigid robotic manipulator in eye-to-hand configuration. The set-point in the system we consider is a visually determined target position. The proposed control law is robust with respect to the camera modelling error and can exponentially stabilize the set-point control system. Furthermore, precise positioning is guaranteed if the desired positioning task can be encoded using the proposed image-based task encoding approaches.

Similar stability analysis can be applied to vision-based robotic systems in eye-in-hand configuration provided that both the end-effector and the target can be observed by the vision system during the maneuvering of the robotic arm.

Acknowledgements The author was supported by Grants NSC-89-2218-E-259-004 and NSC-94-2213-E-027-006 from the National Science Council of Taiwan, Republic of China.

Appendix

Proof of Theorem 4.1

Proof In the light of Eqs. (22), (28), and (29), one can write $\mathbf{J}'(\mathbf{r})\bar{\mathbf{y}}$ as follows.

$$\begin{aligned}\mathbf{J}'(\mathbf{r})\bar{\mathbf{y}} = {} & \bar{y}_{1i}\frac{f_1}{\mathbf{k}_1'\mathbf{r}}\left(\mathbf{i}_1 - \frac{\mathbf{i}_1'\mathbf{r}}{\mathbf{k}_1'\mathbf{r}}\mathbf{k}_1\right) + \bar{y}_{1j}\frac{f_1}{\mathbf{k}_1'\mathbf{r}}\left(\mathbf{j}_1 - \frac{\mathbf{j}_1'\mathbf{r}}{\mathbf{k}_1'\mathbf{r}}\mathbf{k}_1\right) \\ & + \bar{y}_{2i}\frac{f_2}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})}\left(\mathbf{i}_2 - \frac{\mathbf{i}_2'(\mathbf{r}+\mathbf{l})}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})}\mathbf{k}_2\right) + \bar{y}_{2j}\frac{f_2}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})}\left(\mathbf{j}_2 - \frac{\mathbf{j}_2'(\mathbf{r}+\mathbf{l})}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})}\mathbf{k}_2\right) \quad (78)\end{aligned}$$

By virtue of Eqs. (20) and (27), Eq. (78) can be rewritten as

$$\begin{aligned}\mathbf{J}'(\mathbf{r})\bar{\mathbf{y}} = {} & \frac{\bar{y}_{1i}}{\mathbf{k}_1'\mathbf{r}}\left(f_1\,\mathbf{i}_1 - (\bar{y}_{1i} + y_{1i}^*)\mathbf{k}_1\right) + \frac{\bar{y}_{1j}}{\mathbf{k}_1'\mathbf{r}}\left(f_1\,\mathbf{j}_1 - (\bar{y}_{1j} + y_{1j}^*)\mathbf{k}_1\right) \\ & + \frac{\bar{y}_{2i}}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})}\left(f_2\,\mathbf{i}_2 - (\bar{y}_{2i} + y_{2i}^*)\mathbf{k}_2\right) + \frac{\bar{y}_{2j}}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})}\left(f_2\,\mathbf{j}_2 - (\bar{y}_{2j} + y_{2j}^*)\mathbf{k}_2\right). \quad (79)\end{aligned}$$

Using this together with the fact that $\|\bar{\mathbf{y}}_n\|^2 = \|\bar{y}_{ni}\|^2 + \|\bar{y}_{nj}\|^2$, $n = 1, 2$, one can see that

$$\begin{aligned}\mathbf{J}'(\mathbf{r})\bar{\mathbf{y}} = {} & -\frac{1}{\mathbf{k}_1'\mathbf{r}}\left(\|\bar{\mathbf{y}}_1\|^2 + \bar{y}_{1i}\,y_{1i}^* + \bar{y}_{1j}\,y_{1j}^*\right)\mathbf{k}_1 + \frac{f_1}{\mathbf{k}_1'r}(\bar{y}_{1i}\,\mathbf{i}_1 + \bar{y}_{1j}\mathbf{j}_1) \\ & - \frac{1}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})}\left(\|\bar{\mathbf{y}}_2\|^2 + \bar{y}_{2i}\,y_{2i}^* + \bar{y}_{2j}\,y_{2j}^*\right)\mathbf{k}_2 + \frac{f_2}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})}(\bar{y}_{2i}\,\mathbf{i}_2 + \bar{y}_{2j}\mathbf{j}_2). \quad (80)\end{aligned}$$

In the light of Eqs. (27), (28), and (31), we have

$$\mathbf{J}'(\mathbf{r})\bar{\mathbf{y}} = -\frac{1}{\mathbf{k}_1'\mathbf{r}}\left(\|\bar{\mathbf{y}}_1\|^2 + \bar{\mathbf{y}}_1'\mathbf{y}_1^*\right)\mathbf{k}_1 - \frac{1}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})}\left(\|\bar{\mathbf{y}}_2\|^2 + \bar{\mathbf{y}}_2'\mathbf{y}_2^*\right)\mathbf{k}_2 + \frac{f_1}{\mathbf{k}_1'\mathbf{r}}\|\bar{\mathbf{y}}_1\|\,\mathbf{w}_1 + \frac{f_2}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})}\|\bar{\mathbf{y}}_2\|\,\mathbf{w}_2. \tag{81}$$

where $\mathbf{w}_1$ is the unit visual error vector in image plane I_1 and is thus orthogonal to $\mathbf{z}_1$. (cf. Fig. 2). Similarly, $\mathbf{w}_2$ is the unit visual error vector in image plane I_2 and is thus orthogonal to $\mathbf{z}_2$.

From the definition of $\theta_n, n = 1, 2$ in Eq. (32), Eq. (81) can be further written as

$$\mathbf{J}'(\mathbf{r})\bar{\mathbf{y}} = -\frac{1}{\mathbf{k}_1'\mathbf{r}}\left(\|\bar{\mathbf{y}}_1\| + \|\mathbf{y}_1^*\|\cos\theta_1\right)\|\bar{\mathbf{y}}_1\|\mathbf{k}_1 - \frac{1}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})}\left(\|\bar{\mathbf{y}}_2\| + \|\mathbf{y}_2^*\|\cos\theta_2\right)\|\bar{\mathbf{y}}_2\|\mathbf{k}_2 + \frac{f_1}{\mathbf{k}_1'\mathbf{r}}\|\bar{\mathbf{y}}_1\|\,\mathbf{w}_1 + \frac{f_2}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})}\|\bar{\mathbf{y}}_2\|\,\mathbf{w}_2. \tag{82}$$

Therefore, from the definition of $\alpha_n, n = 1, 2$ in Eq. (32), we conclude that

$$\mathbf{J}'(\mathbf{r})\bar{\mathbf{y}} = \begin{bmatrix} \frac{1}{\mathbf{k}_1'\mathbf{r}}\left(-\alpha_1\mathbf{k}_1 + f_1\mathbf{w}_1\right) & \frac{1}{\mathbf{k}_2'(\mathbf{r}+\mathbf{l})}\left(-\alpha_2\mathbf{k}_2 + f_2\mathbf{w}_2\right) \end{bmatrix} \cdot \begin{bmatrix} \|\bar{\mathbf{y}}_1\| \\ \|\bar{\mathbf{y}}_2\| \end{bmatrix}. \tag{83}$$

□

Lemma A.1 (Bellman–Gronwall [29]). *Assume given an interval $\mathcal{I} \subseteq \mathbb{R}$, a constant $c \geq 0$, and two functions*

$$\alpha, \mu : \mathcal{I} \to \mathbb{R}_+ \tag{84}$$

such that α is locally integrable and μ is continuous. Suppose further that for some $\sigma \in \mathcal{I}$ it holds that

$$\mu(t) \leq \nu(t) \triangleq c + \int_\sigma^t \alpha(\tau)\mu(\tau)d\tau \tag{85}$$

for all $t \geq \sigma, t \in \mathcal{I}$. Then, it must hold that

$$\mu(t) \leq ce^{\int_\sigma^t \alpha(\tau)d\tau}. \tag{86}$$

References

1. Shirai, Y., Inoue, H.: Guiding a robot by visual feedback in assembling tasks. Pattern Recogn. **5**, 99–108 (1973)
2. Hill, J., Park, W.T.: Real time control of a robot with a mobile camera. In: Proc. of the 9th ISIR, pp. 233–246, (March 1979)
3. Hutchinson, S.A., Hager, G.D., Corke, P.I.: A tutorial on visual servo control. IEEE Trans. Robot. Autom. **12**(5), 651–670 (October 1996)
4. Hager, G.D., Chang, W.-C., Stephen Morse, A.: Robot hand-eye coordination based on stereo vision. IEEE Control Syst. Mag. **15**(1), 30–39 (February 1995)
5. Colombo, C., Allotta, B.: Image-based robot task planning and control using a compact visual representation. IEEE Trans. Syst. Man Cybern. **29**(1), 92–100 (January 1999)
6. Krupa, A., Gangloff, J., Doignon, C., de Mathelin, M.F., Morel, G., Leroy, J., Soler, L., Marescaux, J.: Autonomous 3-d positioning of surgical instruments in robotized laparoscopic surgery using visual servoing. IEEE Trans. Robot. Autom. **19**(5), 842–853 (October 2003)

7. Backes, P., Diaz-Calderon, A., Robinson, M., Bajracharya, M., Helmick, D.: Automated rover positioning and instrument placement. In: Proc. of the 2005 IEEE Conference on Aerospace, Big Sky, MT, pp. 60–71. IEEE, Los Alamitos, CA (March 2005)
8. Kofman, J., Wu, X., Luu, T.J., Verma, S.: Teleoperation of a robot manipulator using a vision-based human!vrobot interface. IEEE Trans. Ind. Electron. **52**(5), 1206–1219 (October 2005)
9. Chang, W.-C.: Hybrid force and vision-based contour following of planar robots. J. Intell. Robot. Syst. **47**(3), 215–237 (November 2006)
10. Wijesoma, S.W., Wolfe, D.F.H., Richards, R.J.: Eye-to-hand coordination for vision-guided robot control applications. Int. J. Rob. Res. **12**(1), 65–78 (February 1993)
11. Hager, G.D.: Calibration-free visual control using projective invariance. In: Proc. ICCV, pp. 1009–1015 (1995)
12. Kelly, R.: Robust asymptotically stable visual servoing of planar robots. IEEE Trans. Robot. Autom. **12**(5), 759–766 (October 1996)
13. Hager, G.D.: A modular system for robust positioning using feedback from stereo vision. IEEE Trans. Robot. Autom. **13**(4), 582–595 (August 1997)
14. Samson, C., Le Borgne, M., Espiau, B.: Robot control: the task function approach. In: The Oxford Engineering Science Series, no. 22. Clarendon Press, Oxford (1991)
15. Rizzi, A.A.: Dexterous robot manipulation. PhD thesis, Yale University (November 1995)
16. Chang, W.-C.: Vision-based control of uncertain systems. PhD thesis, Yale University, New Haven, CT (December 1997)
17. Chang, W.-C., Hespanha, J.P., Morse, A.S., Hager, G.D.: Task re-encoding in vision-based control systems. In: Proc. of the 1997 Conference on Decision and Control, San Diego, CA, vol. 1, pp. 48–53. IEEE (December 1997)
18. Chang, W.-C., Morse, A.S.: Six degree-of-freedom task encoding in vision-based control systems. In: Proc. of the 14th World Congress of IFAC, Beijing, China, vol. B, pp. 311–316. International Federation of Automatic Control (July 1999)
19. Hespanha, J.P., Dodds, Z., Hager, G.D., Morse, A.S.: What tasks can be performed with an uncalibrated stereo vision system? Int. J. Comput. Vis. **35**(1), 65–85 (1999)
20. Weiss, L.E., Sanderson, A.C., Neuman, C.P.: Dynamic sensor-based control of robots with visual feedback. IEEE J. Robot. Autom. **RA-3**(5), 404–417 (October 1987)
21. Feddema, J.T., Mitchell, O.: Vision-guided servoing with feature-based trajectory generation. IEEE Trans. Robot. Autom. **5**(5), 691–700 (October 1989)
22. Skaar, S.B., Brockman, W.H., Jang, W.S.: Three-dimensional camera space manipulation. Int. J. Rob. Res. **9**(4), 22–39 (August 1990)
23. Hashimoto, K., Kimoto, T., Ebine, T., Kimura, H.: Manipulator control with image-based visual servo. In: Proc. IEEE Int'l Conf. on Robotics and Automation, pp. 2267–2272 (April 1991)
24. Papanikolopoulos, N.P., Khosla, P.K., Kanade, T.: Visual tracking of a moving target by a camera mounted on a robot: a combination of control and vision. IEEE Trans. Robot. Autom. **9**(1), 14–35 (February 1993)
25. Castaño, A., Hutchinson, S.A.: Visual compliance: task-directed visual servo control. IEEE Trans. Robot. Autom. **10**(3), 334–342 (June 1994)
26. Horn, B.K.P.: Robot Vision. The MIT Electrical Engineering and Computer Science Series, 8th edn. McGraw-Hill, New York (1986)
27. Faugeras, O.: Three-Dimensional Computer Vision: A Geometric Viewpoint. MIT Press, Cambridge, MA (1993)
28. Zhang, Z., Luong, Q.-T., Faugeras, O.: Motion of an uncalibrated stereo rig: self-calibration and metric reconstruction. IEEE Trans. Automat. Contr. **12**(1), 103–113 (February 1996)
29. Sontag, E.D.: Mathematical control theory. In: Texts in Applied Mathematics, vol. 6. Springer, Berlin Heidelberg New York (1990)