Passivity-based Visual Pose Regulation for a Moving Target Object in Three Dimensions: Structure Design and Convergence Analysis

Tatsuya Ibuki, Takeshi Hatanaka and Masayuki Fujita

Abstract—This paper investigates passivity-based visual feedback pose regulation whose objective is to control a vision camera pose so that it reaches a desired configuration relative to a moving target object. For this purpose, we present a novel visual feedback control structure including a vision-based observer called visual control observer under the assumption that a pattern of the target motion is available for control. We first focus on evolution of the orientation part and the resulting estimation/control error system is proved to be passive from observer/control input to the estimation/control error output. Accordingly, we also prove that the control objective is achieved by just closing the loop. We then prove convergence of the remaining part of the error system associated with position. Finally, the effectiveness of the present control structure is demonstrated through simulations.

I. INTRODUCTION

A large amount of literature has devoted to fusion of control theory and computer vision. Early works as in [1] are mainly motivated by robot control, but the motivating scenarios of the fusion currently spread over the robotic systems into security and surveillance systems [2], medical imaging procedures [3] and even understanding biological perceptual information processing [4]. The history and recent developments are well summarized in [5]-[8].

In this paper, we focus on vision-based estimation/control of a target object moving in three dimensions as in [9]-[11]. The papers [9], [10] address a problem called “structure and motion from motion” via some estimation mechanisms in systems and control theory. [9] employs an $H_{\infty}$ approach and [10] presents an unknown input observer approach to the problem regarding the object velocities as unknown external disturbances. Meanwhile, the paper [11] presents a vision-based 3D pose estimation mechanism for a moving target object by using passivity of rigid-body motion. Moreover, [11] investigates not only estimation but control of the camera so that it tracks to the object and analyzes the tracking performance in the framework of $L_2$-gain, where the object velocities are viewed as unknown external disturbances.

As another option to deal with the target motion, assuming some target object motion patterns, is commonly used in visual servoing [5], [6] and visual tracking [7] in order to cancel the tracking errors. A typical selection of the target motion patterns is a constant velocity model [6], [7], [12] inspired by classical control theory to cancel tracking errors through an integral term. As the other option, a constant acceleration model is employed in [12]. Periodic motion is also useful especially in medical robotics in order to model the heartbeat and breathing [13].

This paper investigates a vision-based full 3D pose regulation problem investigated in [11] and incorporates the above approach [6], [7], [12], [13] into the framework of [11]. In particular, we present a novel visual feedback control structure with a new vision-based observer integrating the target motion model. After presenting the vision-based observer, we formulate a total estimation/control error system. We next consider a part of the error system describing the evolution of the orientation and angular velocity and show that the present structure of connection lines recovers passivity of the subsystem and just closing the loop via a negative feedback regulates the orientation and angular velocity to desirable states. Then, we shift our focus to the remaining part of the error system describing evolution of position and linear velocity errors and prove that they are also regulated to desirable states by the present control structure. Finally, the effectiveness of the present structure is demonstrated through simulations.

II. PROBLEM STATEMENT

A. Rigid Body Motion

In this paper, we consider a visual feedback system illustrated in Fig. 1, where $\Sigma_{wc}$, $\Sigma_c$ and $\Sigma_o$ represent the world frame, the camera frame and the object frame, respectively. The position vector and rotation matrix from $\Sigma_{wc}$ to $\Sigma_c$ are respectively denoted by $p_{wc} \in \mathbb{R}^3$ and $\hat{\xi}_{wc}\theta_{wc} \in SO(3)$. Here, the vector $\xi_{wc} \in \mathbb{R}^3$ ($\|\xi\| = 1$) specifies the rotation axis and $\theta_{wc} \in \mathbb{R}$ is the rotation angle. For simplicity, we use $\xi \theta_{wc}$ to denote $\xi_{wc}\theta_{wc}$. The notation ‘$\wedge$’ is the operator s.t. $\hat{a} = a \times b$, $a, b \in \mathbb{R}^3$ for the vector cross-product $\times$, i.e. $\hat{a} \in so(3)$ is a $3 \times 3$ skew-symmetric matrix.
The pair of the position \( p_{wc} \) and the orientation \( e^\theta_{wc} \) denoted by \( g_{wc} = (p_{wc}, e^\theta_{wc}) \in SE(3) := \mathbb{R}^3 \times SO(3) \) is called pose of the camera relative to \( \Sigma_w \). Similarly, we denote by \( g_{wo} = (p_{wo}, e^\theta_{wo}) \in SE(3) \) the pose of the object relative to \( \Sigma_w \). We also define the body velocity of the camera relative to \( \Sigma_w \) as \( V^b_{wc} = (v_{wc}, \omega_{wc}) \in \mathbb{R}^6 \), where \( v_{wc} \in \mathbb{R}^3 \) and \( \omega_{wc} \in \mathbb{R}^3 \) respectively represent the linear and angular velocities of the origin of \( \Sigma_c \) relative to \( \Sigma_w \) [14]. Similarly, the body velocity of the object relative to \( \Sigma_w \) is denoted by \( V^b_{wo} = (v_{wo}, \omega_{wo}) \in \mathbb{R}^6 \).

In this paper, we use the following homogeneous representation of \( g = (p, e^\theta) \in SE(3) \) and \( V^b = (v, \omega) \in \mathbb{R}^6 \).

\[
g = \begin{bmatrix} e^\theta & p \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{4 \times 4}, \quad V^b = \begin{bmatrix} \omega \\ v \end{bmatrix} \in \mathbb{R}^{4 \times 4}.
\]

Then, the body velocities \( V^b_{wc} \) and \( V^b_{wo} \) are simply given by \( \dot{V}^b_{wc} = g^{-1}_{wc} \dot{V}^b_{wc} \) and \( \dot{V}^b_{wo} = g^{-1}_{wo} \dot{V}^b_{wo} \), respectively. Additionally, the adjoint transformation [14] associated with \( g \) is denoted by \( Ad(g) \in \mathbb{R}^{6 \times 6} \) which satisfies \( V^t = Ad(g)V \) if \( V^t = gVg^{-1} \).

Let \( g_{co} = (p_{co}, e^\theta_{co}) \in SE(3) \) be the pose of \( \Sigma_c \) relative to \( \Sigma_e \). Then, it is known that \( g_{co} \) can be represented as \( g_{co} = g_{wc}g_{wo} \). By using the body velocities \( V^b_{wc} \) and \( V^b_{wo} \), the motion of the relative pose \( g_{co} \), called relative rigid body motion, is written as follows [14].

\[
\dot{g}_{co} = -\dot{V}^b_{wc}g_{co} + g_{co}\dot{V}^b_{wo}.
\]

**B. Visual Measurement**

In this subsection, we visualize the visual measurement of the camera which is available for estimation/control. Throughout this paper, we use the pinhole camera model with perspective projection [14] (Fig. 2). Note however that the subsequent discussions are applicable to panoramic cameras through the appropriate modifications as in [15].

We assume that the target object has \( m (m \geq 4) \) feature points. Let \( p_{oi} \in \mathbb{R}^3 \) and \( p_{ci} \in \mathbb{R}^3 \) be the position vectors of the target object’s \( i \)-th feature point relative to \( \Sigma_c \) and \( \Sigma_e \), respectively. Using a transformation of the coordinates, we have \( p_{ci} = g_{co}p_{oi} \) where \( p_{ci} \) and \( p_{oi} \) should be respectively regarded, with a slight abuse of notation, as \( [p_{ci}^T 1]^T \) and \( [p_{oi}^T 1]^T \). Let the \( m \) feature points of the object on the image plane coordinate \( f := [f_1^T \cdots f_m^T]^T \in \mathbb{R}^{2m} \) be the visual measurement of the vision camera. Then, it is well known [14] that \( f_i \in \mathbb{R}^2 \) is given by the perspective projection:

\[
f_i = \frac{\lambda}{z_{ci}} \begin{bmatrix} x_{ci} \\ y_{ci} \end{bmatrix}, \quad p_{ci} = \begin{bmatrix} x_{ci} \\ y_{ci} \end{bmatrix} z_{ci}^T,
\]

where \( \lambda \) is a focal length of the vision camera (Fig. 2).

We assume that the feature points \( p_{oi} \) are known a priori. Then, the visual measurement vector \( f(g_{co}) \) depends only on the relative pose \( g_{co} \). Fig. 3 illustrates the block diagram of the relative rigid body motion with the camera model.

**C. Visual Pose Regulation**

In this subsection, we formulate the control objective of this paper. Before mentioning it, we assume that the target object body velocity \( V^b_{wo} \) is (approximately) given in the form of a finite Fourier series expansion:

\[
V^b_{wo} = c + \sum_{i=1}^n a_i \eta_i(t),
\]

where \( a_i = (a_{v,i}, a_{\omega,i}) \in \mathbb{R}^6, c = (c_v, c_\omega) \in \mathbb{R}^6 \) and \( \eta_i(t) := \sin w_i t \in \mathbb{R} \) are the Fourier bases whose frequencies \( w_i > 0, \ i \in \{1, \ldots, n\} \) are known a priori.

Under the assumption of (3), the objective of this paper is to design the camera velocity input \( V^b_{wc} \) so as to achieve the following pose regulation conditions

\[
\lim_{t \to \infty} (V^b_{wc} - Ad(g_{co})V^b_{wo}) = 0, \quad \lim_{t \to \infty} \sum_{i=1}^n a_i \hat{\eta}_i(t) = 0,
\]

where \( e_R(\xi \theta) := \sk(\xi \theta) \in \mathbb{R}^3, \sk(\xi \theta) := (1/2)(\xi \theta - \xi \theta^T) \in \mathbb{R}^{3 \times 3} \) and \( g_d = (p_d, e^\theta_{do}) \in SE(3) \) specifies a fixed desirable configuration of the camera relative to the object.

Throughout this paper, the problem is called visual pose regulation.

We also provide some properties of (3) necessary for the subsequent discussions. Let us now define

\[
x_{z,v,0} = c_v, \quad x_{z,o,0} = a_{v,0} \eta_1, \quad x_{z,v,0} = c_{\omega}, \quad x_{z,o,0} = a_{\omega,0} \eta_1 \in \mathbb{R}^3,
\]

\[
x = (x_{z,v}, x_{z,o}) := (x_{z,v,0}, x_{z,o,0}, \dot{x}_{z,v}, \dot{x}_{z,o}) \in \mathbb{R}^{12n+6}.
\]
Then, it is straightforward to see that the time evolution of $V_{wo}^b$ is represented by the linear time invariant system:

$$\dot{x} = Ax, \quad A := \begin{bmatrix} A_v & 0 \\ 0 & A_\omega \end{bmatrix} \in \mathbb{R}^{(12n+6) \times (12n+6)},$$

(6a)

$$V_{wo}^b = Cx, \quad C := \begin{bmatrix} C_v & 0 \\ 0 & C_\omega \end{bmatrix} \in \mathbb{R}^{6 \times (12n+6)},$$

(6b)

$$A_v = A_\omega := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -\text{diag}(w_2^2, \ldots, w_n^2) & I \end{bmatrix},$$

$$C_v = C_\omega := \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \cdot I_3 \in \mathbb{R}^{n \times 3}.$$

We also have the following lemma.

Lemma 1: Let $B_v = C_v^T$. Then, the linear system $(A_\omega, B_v, C_\omega, 0)$ with state $x_c$ is passive with respect to the storage function $S(x_c) := (1/2)x_c^T P x_c$ with

$$P := \begin{bmatrix} I_{3(n+1)} & 0 \\ 0 & \text{diag}(1/w_1^2, \ldots, 1/w_n^2) \otimes I_3 \end{bmatrix}. $$

Proof: This lemma can be proved by the direct calculation of $\dot{S}$.

Remark 1: An important example of (3) is a constant velocity $V_{wo}^b = c$ [5, 6, 7]. The model is in practice useful not only in the case where the velocity is really constant since any signal is approximated by a piecewise step function. Similarly, (3) is useful even if $V_{wo}^b$ is not really periodic, since a future profile of $V_{wo}^b$ over a finite interval is approximated as (3). Namely, we can regard the estimation process over the infinite time interval as repeats of the estimation over a finite time interval. A variety of real periodic motion is also approximately described in the form.

III. STRUCTURE DESIGN FOR VISUAL POSE REGULATION

A. Full Information Feedback

In this subsection, we present a visual feedback system achieving pose regulation assuming that the relative pose $g_{co}$ and the object body velocity $V_{wo}^b$ are available for control.

Let us define the control error $\dot{g}_{ce}^f := \tilde{g}_{d}^{-1}g_{co} \in SE(3)$ and $V_{wc}^b := \text{Ad}_{(g_{ce})^{-1}}V_{wo}^b \in \mathbb{R}^b$. Then, we obtain

$$\dot{g}_{ce}^f = -\tilde{V}_{wc}^b g_{ce}^f + \tilde{g}_{ce}^f \dot{V}_{wo}^b. $$

(7)

We now fix the form of the velocity input $V_{wc}^b$ as

$$V_{wc}^b = -u_c + \text{Ad}_{(g_{ce})}V_{wo}^b,$$

(8)

where $u_c = (u_{exp}, u_{cr}) \in \mathbb{R}^6$ is the new input to be determined in order to drive $g_{co}$ to the desirable $g_d$. Substituting (8) into (7) cancels the second term $g_{ce}^f \dot{V}_{wo}^b$ of (7) and hence we have

$$\dot{g}_{ce}^f = \tilde{u}_c g_{ce}^f. $$

(9)

Let $u_c$ and $\nu^f := E_R(g_{ce}^f) \in \mathbb{R}^b$ be respectively the input and output vectors of system (9), whose block diagram is illustrated in Fig. 4. Then, we have the following lemma.

Fact 1: [11] The system (9) is passive from $u_c$ to $\nu^f$ with the storage function

$$\psi(g_{ce}^f) = \frac{1}{2} ||p_{ce}||^2 + \phi(e^{\xi \theta_{ce}}), \quad \phi(e^{\xi \theta_{ce}}) := \frac{1}{2} \text{tr}(I_3 - e^{\xi \theta_{ce}}).$$

Fact 1 implies that closing the loop with

$$u_c = -k_c \nu^f = -k_c E_R(g_{ce}^f), \quad k_c > 0$$

(10)

assures asymptotic stability of the equilibrium $\nu^f = 0$. We see from the definition of $\nu^f$ that $\nu^f = 0$ is equivalent to $E_R(g_{ce}^f) = 0$ and hence (5). In addition, in case of $E_R(g_{ce}^f) = 0$, we obtain $u_c = 0$ and $V_{wc}^b = V_{wo}^b$ from (8), which implies (4). Namely, the pose regulation is achieved by (8) and (10). The resulting control input $V_{wc}^b$ is given by

$$V_{wc}^b = k_c \text{Ad}_{(g_{ce})}E_R(g_{d}^{-1}g_{co}) + \text{Ad}_{(g_{co})}V_{wo}^b $$

(11)

from (8) and (10).

We now see that (11) consists of the state feedback term $-k_c \text{Ad}_{(g_{ce})} \nu^f$ and the disturbance feedforward term $\text{Ad}_{(g_{co})}V_{wo}^b$. However, the velocity $V_{woo}^b$ is in practice not available for control and it is not straightforward to extract $g_{co}$ from the visual measurement $f(g_{co})$. Therefore, we will present a vision-based observer estimating these parameters only from $f(g_{co})$ in the next subsection.

B. Observer Design

In this subsection, we consider estimation of the relative pose $g_{co}$ and the object body velocity $V_{wo}^b$ from the visual measurement $f(g_{co})$ in case that $V_{wo}^b$ is given as (3).

Similarly to [11], we first prepare a model of the 3D object motion (1) and (6) as

$$\dot{\tilde{g}}_{co} = -\tilde{V}_{wc}^b \tilde{g}_{co} - \tilde{g}_{co} \tilde{u}_c + \tilde{g}_{co} \tilde{V}_{wo}^b, $$

(12a)

$$\dot{x} = Ax - Bu_c, \quad B := C \tilde{F},$$

(12b)

$$\tilde{V}_{wo}^b = C \tilde{x},$$

(12c)
where \( \bar{g}_{co} = (\bar{p}_{co}, \hat{\bar{e}}\hat{\theta}_{co}) \in SE(3) \), \( \bar{x} = (\bar{x}_v, \bar{x}_\omega) \in \mathbb{R}^{12n+6} \) and \( \bar{V}_{wo}^b = (\bar{v}_{wo}, \bar{\omega}_{wo}) \in \mathbb{R}^6 \) are the estimates of \( g_{co}, \bar{x} \) and \( V_{wo}^b \), respectively. The inputs \( u_c \in \mathbb{R}^6 \) should be designed so that these estimates asymptotically converge to their actual values. Note that the paper [11] builds a model only of (1) with \( V_{wo}^b = 0 \) since [11] assumes no prior information on the target object motion.

Let us now define the estimation errors:
\[
g_{ce} = (g_{ce}, e^\hat{\theta}_{ce}) \in SE^e(3), \quad x = (x_v, x_\omega) \in \mathbb{R}^{12n+6}, \quad V_{wo} = (v_{wo}, \omega_{wo}) \in \mathbb{R}^6
\]
Then, we have the following estimation error system
\[
\hat{g}_{ce} = \hat{u}_c g_{ce} - \hat{V}_{wo}^b g_{ce} + g_{ce} \hat{V}_{wo}^b, \quad \hat{x} = A \hat{x} + B u_v, \quad V_c = C x
\]
from (1), (6) and (12). It is also shown in [11] that, as long as \( m \geq 4 \), the estimation error vector \( e_c \) is approximately reconstructed from \( \hat{f} := f - \hat{f} \) as
\[
e_c = J^\dagger(\hat{g}_{co}) f_e, \quad \hat{f} \text{ is computed by (2) using } \hat{g}_{co} \text{ instead of } g_{co} \text{ and } J^\dagger(\hat{g}_{co}) \text{ is the pseudo-inverse of the image Jacobian [14].}
\]

We next present the following input to the observer.
\[
u_v = -k_v e_v, \quad u_e = -K_e e_v, \quad K_e := \begin{bmatrix} k_e I_3 & 0 \\ 0 & k_e I_3 \end{bmatrix}
\]
which is called visual motion observer throughout this paper. The proof is straightforward from the results in the subsequent section.

The total estimation mechanism is formulated as
\[
\begin{align*}
\dot{x} &= Ax + Bu_v, \quad \dot{V}_{wo}^b = Cx, \quad \dot{\hat{g}}_{co} = -\hat{V}_{wo}^b \hat{g}_{co} + \hat{g}_{co} \hat{V}_{wo}^b - \hat{u}_c, \\
e_c &= J^\dagger(\hat{g}_{co}) f_e, \\
\dot{u}_v &= -k_v e_v, \quad \dot{u}_e = -K_e e_v,
\end{align*}
\]
which is called visual motion observer throughout this paper. The block diagram of the resulting estimation mechanism (16) together with the relative rigid body motion (1) and the velocity generator block (3) is illustrated in Fig. 5.

**C. Total Estimation/Control Error System**

In this subsection, we present a novel visual feedback structure achieving visual pose regulation based on the contents in Subsections III-A and III-B.

Similarly to Subsection III-B, we build the 3D object motion model (12), formulate the estimation error system (13) and close the loop of input \( u_v \) by
\[
\dot{u}_v = -k_v e_v.
\]

We next formulate the control error system. Basically, we try to imitate the structure of the full information feedback case in Subsection III-A. However, neither \( V_{wo}^b \) nor \( g_{co} \) are available for control. We thus replace \( \hat{V}_{wo}^b \) and \( \hat{g}_{co} \) by their estimates \( \hat{V}_{wo}^b \) and \( \hat{g}_{ce} \) and fix the form of \( V_{wo}^b \) as
\[
V_{wo} := Ad(\hat{g}_{co}^{-1}) V_{wo}^b, \quad V_{wo}' := -u_c + Ad(g_{ce}) \hat{V}_{wo}^b
\]
in imitation of (8). Here, \( g_{ce} := g_{ce}^{-1} \in SE^e(3) \) is the control error and we also define \( e_v := \hat{E}_R(g_{ce}) \). The time evolution of the control error \( g_{ce} \) is then described by
\[
e_v = \hat{u}_c g_{ce} - g_{ce} \hat{u}_e
\]
from (12a) and (18).

From (6), (13), (17) and (19), the total estimation/control error system is formulated by
\[
\begin{align*}
\dot{x} &= Ax, \quad \dot{V}_{wo}^b = Cx, \\
\dot{x}_e &= A x_e - k_v B e_v, \quad \dot{V}_{wo} = V_{wo}^b - \hat{V}_{wo}^b = C x_e, \\
\dot{g}_{ce} &= \hat{u}_c g_{ce} - g_{ce} \hat{u}_e, \\
\dot{g}_{ce} &= \hat{u}_c g_{ce} - \hat{V}_{wo}^b g_{ce} + g_{ce} \hat{V}_{wo}^b.
\end{align*}
\]
It should be noted that the time evolution of the orientation part \( (x_\omega, x_{e_\omega}, \hat{e}_\omega, \hat{e}^\theta_{ce}, \hat{e}^\theta_{ee}) \) is independent of that of position part \( (x_v, x_{e_v}, p_{ce}, p_{ee}) \) while evolution of \( (x_v, x_{e_v}, p_{ce}, p_{ee}) \) depends on that of \( (x_v, x_{e_v}, \hat{e}_\omega, \hat{e}^\theta_{ce}, \hat{e}^\theta_{ee}) \).

**IV. MAIN RESULT**

In this section, we close the loops of \( u_c = (u_{cp}, u_{cR}) \) and \( u_e = (u_{ep}, u_{eR}) \) and prove the visual pose regulation.

**A. Visual Feedback System: Orientation Part**

Extracting the orientation/angular part from (20) yields
\[
\begin{align*}
\dot{x}_\omega &= A_\omega x_\omega, \quad \dot{\omega}_{wo} = C_\omega x_\omega, \\
\dot{x}_{e_\omega} &= A_\omega x_{e_\omega} + k_v B_\omega e_v, \quad \dot{\omega}_{ee} = C_\omega x_{e_\omega}, \\
\dot{\hat{e}}^{\theta}_{ce} &= \hat{e}_R \hat{e}^{\theta}_{ce} - \hat{e}^{\theta}_{ee} \hat{u}_e, \\
\dot{\hat{e}}^{\theta}_{ee} &= (\hat{u}_R - \hat{\omega}_{wo}) \hat{e}^{\theta}_{ee} + \hat{e}^{\theta}_{ee} \hat{\omega}_{wo}.
\end{align*}
\]

Let the controlled output of the system (21) be
\[
\nu_R = N \begin{bmatrix} \hat{E}_R(\hat{e}^{\theta}_{ce}) \\ \hat{E}_R(\hat{e}^{\theta}_{ee}) \end{bmatrix}, \quad N := \begin{bmatrix} I_3 & 0 \\ -I_3 & I_3 \end{bmatrix}.
\]
Then, the block diagram from \((u_{cR}, u_{eR})\) to \(\nu_R\) is illustrated in Fig. 6 and we have the following lemma.

**Lemma 2:** The total error system (21) is passive from \((u_{cR}, u_{eR})\) to \(\nu_R\) with the storage function \(\dot{U}_R := \phi(e^{\theta_{ee}}) + \phi(e^{\theta_{ee}}) + (1/k_e)S(x_{e,\omega})\), where \(S(x_{e,\omega}) = (1/2)x_{e,\omega}^TPx_{e,\omega}\).

**Proof:** This lemma can be proved by the direct calculation of \(\dot{U}_R\).

From Lemma 2, we present the input
\[
\begin{bmatrix}
  u_{cR} \\
  u_{eR}
\end{bmatrix} = -k_c\nu_R = k_c
\begin{bmatrix}
  -I_3 & 0 \\
  I_3 & -I_3
\end{bmatrix}
\begin{bmatrix}
  e_{R}(\theta_{ee}) \\
  e_{R}(\theta_{ee})
\end{bmatrix}.
\tag{22}
\]

Then, we have the following lemma.

**Lemma 3:** Suppose that the input (22) is applied to the total error system (21). Then, all the trajectories of its state \(x_R := (x_{e,\omega},x_{e,\omega},e^{\theta_{ee}},e^{\theta_{ee}})\) satisfy
\[
\lim_{t \to \infty} (\omega_e, e_R(\theta_{ee}), e_R(\theta_{ee})) = 0.
\tag{23}
\]

**Proof:** From Lemma 2, we immediately obtain
\[
\dot{U}_R \leq -k_c\|\nu_R\|^2 \leq 0.
\]

Since the matrix \(N\) is nonsingular, \(\dot{U}_R = 0\) holds iff \(e_R(\theta_{ee}) = e_R(\theta_{ee}) = 0\). Then, this lemma can be proved by LaSalle’s Invariance Principle [16].

The proof of Lemma 3 relies on the passivity in Lemma 2. Note that the inner loops (17) and (18) closed before determining \((u_{cR}, u_{eR})\) allow the system to be passive. Namely, the operations in (17) and (18) are regarded as a kind of passivation of the estimation/control error system.

**B. Visual Feedback System: Position Part**

In this subsection, we close the loop of \((u_{cp}, u_{ep})\) and prove the visual pose regulation.

The time evolution of \((x,v,\nu_{pce},\nu_{pee})\) in the total error system (20) with (22) is described by
\[
\begin{align*}
\dot{x}_v &= A_vx_v, \quad v_{wo} = C_vx_v, \tag{24a} \\
\dot{x}_{v,e} &= A_vx_{v,e} - k_vB_pp_{pee}, \quad v_e = C_vx_{v,e}, \tag{24b} \\
p_{pee} &= u_{cp} - u_{ep} + \delta_{ee}, \tag{24c} \\
p_{ce} &= u_{cp} + v_e + \tilde{p}_{cee}\omega_{wo} + \delta_{ee}, \tag{24d}
\end{align*}
\]

where
\[
\delta_{ee} := k_p\tilde{p}_{ce}e_R(\theta_{ee}) + (I_3 - e^{\theta_{ee}})u_{ep},
\]
\[
\delta_{ee} := (e^{\theta_{ee}} - I_3)v_{wo} + k_p\tilde{p}_{ce}(e_R(\theta_{ee}) - e_R(\theta_{ee})).
\]

Note that \(\lim_{t \to \infty} \delta_{ee} = \lim_{t \to \infty} \delta_{ee} = 0\) for bounded \(u_{ep}\) from Lemma 3.

We next close the loop of \((u_{cp}, u_{ep})\) as
\[
\begin{bmatrix}
  u_{cp} \\
  u_{ep}
\end{bmatrix} =
\begin{bmatrix}
  -k_cI_3 & \tilde{\omega}_{wo} \\
  0 & -k_cI_3 + \tilde{\omega}_{wo}
\end{bmatrix}
\begin{bmatrix}
  p_{ce} \\
  p_{pee}
\end{bmatrix}.
\tag{25}
\]

Then, substituting (25) into (24) yields
\[
\begin{align*}
\dot{p}_{ce} &= -k_c\tilde{p}_{ce} + k_p\tilde{p}_{pee} + \delta_{ee}, \\
\dot{p}_{pee} &= -k_c\tilde{p}_{pee} + v_e + \delta_{ee}.
\end{align*}
\]

Let us now define \(x_p = (x_{e,v}, \nu_{pce}, \nu_{pee})\). Then, we obtain
\[
\dot{x}_p = \Phi x_p + \begin{bmatrix}
  0 \\
  \delta_{ee}
\end{bmatrix}, \quad \Phi :=
\begin{bmatrix}
  A_v & -k_vB_v & 0 \\
  C_v & -k_vI_3 & 0 \\
  0 & k_vI_3 & -k_vI_3
\end{bmatrix}
\tag{26}
\]

and the following theorem holds.

**Theorem 2:** All the trajectories of the total error system (24) with (25) satisfy
\[
\lim_{t \to \infty} (v_e, \nu_{pee}, \nu_{pee}) = 0.
\tag{27}
\]

**Proof:** Let us view the system (26) as a linear time invariant system \(\dot{x}_p = \Phi x_p\) with perturbations \(\delta_{ee}\) and \(\delta_{ee}\). Remark that \(\delta_{ee}\) and \(\delta_{ee}\) are vanishing as time goes to infinity from their definitions and Lemma 3. Also, it can be proved that \(\Phi\) is a stable matrix from the definitions of \(A_v, B_v\) and \(C_v\), which means that the origin of the nominal system \(\dot{x}_p = \Phi x_p\) is exponentially stable. It is thus immediately proved from stability theory of perturbed systems ([16], Lemma 9.6) that \(\lim_{t \to \infty} x_p = 0\).

The combination of (23) and (27) is equivalent to (4) and (5). Therefore, the control objective is proved to be achieved by the present visual feedback system. The block diagram of the total estimation/control structure is shown in Fig. 7.

We finally give a remark on the present estimation/control structure. The velocity input \(V_{wce}^b\) is given by
\[
V_{wce}^b = k_cA_{d(g_d)}E_R(g_d^{-1}\dot{g}_{ee}) + A_{d(\hat{g}_d)}\hat{w}_{wo}\omega_{pee}
\]

from (18), (22) and (25). We see that the first and second terms of (28) have the same form as the full information feedback (11). However, since the actual values of \(g_{ee}\) and \(V_{wce}^b\) are not available, these variables are replaced by the estimates produced by the observer (12) with inputs
\[
\begin{bmatrix}
  u_v \\
  u_e
\end{bmatrix} =
\begin{bmatrix}
  -k_cI_6 & 0 \\
  -k_cI_6 & 0
\end{bmatrix}
\begin{bmatrix}
  E_R(g_{ee}) \\
  E_R(\hat{g}_{ee})
\end{bmatrix} +
\begin{bmatrix}
  0 \\
  0
\end{bmatrix}
\]

using the technique in (14). Namely, the present control mechanism has the same form as the structure of [17] except for the third term of (28) and the second term of (29). These terms are correction terms of the error between \((g_{ee}^{-1}\hat{g}_{ee})^v\)
and \( V_c = V^b_{wco} - \bar{V}^b_{wco} \) and appears since we impose the assumption of (3) on the body velocity.

V. VERIFICATION

We finally demonstrate the present control structure through simulations. We assume that the target object moves with a constant body velocity \( \nu_{wco} = [0.2 \ 0 \ 0.1]^T \) [m] and \( \omega_{wco} = [0 \ 0 \ 0.1]^T \) [rad/s] from the initial state \( p_{wco}(0) = [0 \ 0 \ 1]^T \) [m] and \( \theta_{wco}(0) = [0 \ 0 \ 0]^T \) [rad]. The objective here is to lead the camera pose with the initial pose \( p_{wco} = [0 \ 0 \ 0]^T \) [m], \( \theta_{wco} = [0 \ 0 \ 0]^T \) [rad] to a desirable configuration \( \theta_d = ([0 \ 0.1 \ 0.1]^T, I_3) \) with respect to the object.

In order to meet the requirement, we run the present method with \( k_v = 4, k_c = 4 \) and \( k_t = 6 \). Figs. 8-13 show the time responses of variables. Figs. 8 and 9 include responses of variables associated with linear and angular velocities, where the solid curves show the estimated object velocities, dashed ones the actual object velocities, respectively. We see from the figures that the observer correctly estimates \( V^b_{wco} \). In addition, in order to confirm (4), the camera body velocity and object velocity transformed to the camera frame are depicted in Figs. 10 and 11. We see that (4) is actually satisfied by the present mechanism.

We next demonstrate (5). Figs. 12 and 13 illustrate the time responses of the relative pose \( \gamma_{co} \) and its estimate \( \hat{\gamma}_{co} \). We see from these figures that both of the actual and estimated relative poses asymptotically converge to the prescribed \( \gamma_d \), which implies (5). In summary, we conclude that the present estimation/control mechanism achieves visual pose regulation (4) and (5).

VI. CONCLUSIONS

This paper has investigated passivity-based visual pose regulation whose objective is to lead a vision camera pose to a desired configuration relative to a moving target object. For this purpose, we have presented a novel visual feedback control structure including a vision-based observer with a 3D target object motion model. We have also proved, based on passivity-based control theory and stability theory of perturbed systems, that the control objective is achieved by using the present control structure. Finally, the effectiveness of the present control structure has been demonstrated through simulations.

REFERENCES