

Received November 5, 2019, accepted November 19, 2019, date of publication November 29, 2019, date of current version January 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2956777

# Integral-Sliding-Mode-Observer-Based Structure and Motion Estimation of a Single Object in General Motion Using a Monocular Dynamic Camera

DONGKYOUNG CHWA<sup>1</sup>

Department of Electrical and Computer Engineering, Ajou University, Suwon 443-749, South Korea

e-mail: dkchwa@ajou.ac.kr

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT, and Future Planning under Grant 2017R1A2B4009486, and in part by the Korea Electric Power Corporation under Grant R19X001-21.

**ABSTRACT** An integral sliding mode observer (ISMO)-based method for estimating the structure and motion (SaM) of an object in general motion is proposed using a monocular dynamic (moving) camera. As the unknown range and object velocity can be considered as disturbances and should be quickly estimated, it is necessary to maintain the robustness against these disturbances from the start by eliminating the reaching mode. Therefore, the ISMO-based method is proposed on the basis of a relative camera-object motion model. By formulating the relative motion model with three-dimensional measurable state variables, three unknown components among the range, unknown object velocity components, and unknown camera velocity components can be estimated by the proposed method in the following cases: i) the camera is in dynamic motion and the object is in semigeneral motion [i.e., initially static (stationary) and then in general (static or dynamic) motion], ii) the camera is in dynamic motion and the object is in general motion within a constrained space, and iii) both the object and camera are in general motion within a less constrained space when the range information is available. Simulation and experimental results demonstrate that the range and object velocity can be estimated with a satisfactory transient response using a monocular dynamic camera.

**INDEX TERMS** Camera-object relative motion dynamics, integral sliding mode observer, monocular dynamic camera, object in general motion, structure and motion estimation.

## I. INTRODUCTION

Three-dimensional (3D) estimation of the *structure* of an object (i.e., the relative Euclidean distances in the  $X$ ,  $Y$ , and  $Z$  directions of an object with respect to the camera) requires the range information between an object and a camera to be obtained using a camera. In the stereo vision method using two cameras to obtain the range by triangulation [1]–[6], inaccurate camera calibration results in the large estimation errors of the structure. Therefore, triangulation-based range estimation methods using a monocular camera were developed [7]–[10]; however, most of their application for the real-time object tracking is limited in that the object should move along the restricted trajectories and the required batch processing involves a large amount of image data.

The associate editor coordinating the review of this manuscript and approving it for publication was Shihong Ding<sup>2</sup>.

Whereas 3D reconstruction methods and structure from motion (SfM) estimation methods using a monocular camera have been much studied in [7]–[30], the *motion* of an object (i.e., the velocity of an object) along with the structure of an object need to be effectively used in practical applications.

Accordingly, the structure and motion (SaM) estimation problem has been extensively studied in the field of vision-based control [31], [32]. Structure and/or motion estimation can be readily employed for the visual servoing and tracking control of robots based on a vision system [33]–[36]. In particular, in the research on Euclidean 3D reconstruction using a monocular camera for different camera and object motions [12], [19], [37]–[40], [42]–[49], either the range and object velocity were not simultaneously estimated or the camera and object motions were constrained. Therefore, an estimation algorithm for the SaM of an object should be developed for more general object motion

than those in existing studies [19], [37]–[49] for practical applications.

Although the algorithm for the range and motion estimation for a camera using two objects in [50] can be developed for the SaM estimation of an object using a monocular camera, it cannot consider the transient response. As more general camera and object motions come into existence, the existing nonlinear observers for SaM estimation in [19], [37]–[48], [50] cannot be effectively utilized against a fast object motion in that the SaM estimation of an object can give impractical initial estimates to be used for actual applications. Accordingly, the proposed method has been motivated by the fact that i) an improved method for the SaM estimation of an object in general motion should be developed by designing an observer that achieves a satisfactory transient response for the estimation error; ii) the restrictions on the motions of the camera and object and the required velocity information should be reduced and clarified in the SaM estimation method.

An integral sliding mode observer (ISMO) based on integral sliding mode control (ISMC) [51]–[60] is beneficial in that the reaching mode is eliminated in an ISMO and the sliding mode is enforced from the beginning of the estimation. Thus, by considering the unknown range and object velocity as disturbances, the robustness against disturbances throughout the entire estimation response can result in a much improved transient response. These objectives of the ISMO are different from those of sliding mode control methods [61]–[64]. However, ISMO has been relatively less studied [65]–[68] even with the advantages of the ISMC in [51]–[60]. When these advantages of the ISMC can be introduced in the ISMO-based SaM estimation method, the performance of the ISMO-based SaM estimation of an object in general motion using a monocular camera can be significantly improved in such a way that the range and velocity of the object in general motion can be estimated with a rapid and much improved transient response. This is another motivation of the proposed ISMO-based method. In particular, the ISMO-based SaM estimation method has not been studied before for the object in general motion using a monocular camera. Further development of various ISMOs by incorporating the ISMC methods in [51]–[60] is interesting, but this is not within the scope of this paper.

In this paper, an ISMO-based SaM estimation method for a single object in general motion using a monocular camera is proposed by considering the fast object motion and unknown information as disturbances so that the estimation response initially starts from the sliding surface. Owing to the performance of the proposed ISMO, the range and object velocity can be quickly estimated with a significantly improved transient response even in the presence of the unknown components of the camera velocity and the constraints on the camera and object motions unlike the existing algorithms [1]–[19], [19]–[50].

The contributions of the proposed method can be described in the follow way.

- The relative motion dynamics between the camera and the object are introduced in different forms depending on the camera and object motions. Considering that the measurable state vector is three-dimensional in nature, three unknown components among the range and object velocity components are considered as disturbances. Thus, the SaM can be estimated by designing an SaM observer even when the object is in more general motion. The insight for the derivation of the SaM dynamic equations for different scenarios can be explained as follows. When the number of the available state variables in the SaM equations are equal to or more than the number of the information to be obtained, the considered SaM estimation problem can be solved. Therefore, depending on the conditions and the number of available components of the object and camera velocity vectors, the SaM equations has been reformulated for each scenario in this paper such that SaM estimation problem can be solved by the proposed method for the scenarios that have not been studied before in the existing literature [19], [37]–[50]. The dynamic systems such as SaM equations were shown to be useful in modeling the time series data such as the human pose and motion [69]. Similarly, the SaM equations can be effectively used for the SaM estimation of the object, as will be described in the following sections. The dynamic systems have been studied in the numerous computer vision applications including data visualization, dynamic texture categorization, identification of crowd behaviors in visual scenes, and video-based inference [70]–[74].
- An ISMO-based SaM estimation method is proposed for an object in general motion. Robustness against the disturbances is maintained from the start by eliminating the reaching mode such that the transient response of SaM estimation can be greatly improved unlike the existing methods [1]–[19], [19]–[50]. Among these studies, there has been much work on the non-rigid structure from motion (SfM) which studied both camera motion and structural deformation [22]–[30]. Whereas the proposed method does not handle the structural deformation, it studies the motion estimation such as the unknown components of the camera velocity and object velocity. In this sense, the proposed method achieves the objective that is not the same as the objective of the non-rigid SfM. It should be also noted that the performance of the SaM estimation using the proposed ISMO does not depend on the initial conditions of the object position, object velocity, camera position, and camera velocity, which is verified in the proofs of the theorem and corollaries in Sections IV–VI using the Lyapunov stability analysis [75]. In addition, the proposed method does not assume the constant velocities of the camera and object. That is, the proposed ISMO can estimate the range and the unknown components of the camera and object velocities irrespective of whether they are time-varying or time-invariant. These are also

the advantages of the use of the proposed SaM dynamic equations and ISMO designed on the basis of the SaM equations. As will be shown in the figures in Section VII, the proposed observer maintains the improved estimation performance compared with the previous observer in [50]. Although the ISMOs were studied in [65]–[68], the ISMO in [64]–[67] did not involve the integral sliding surface and the ISMO in [68] was designed only for the linear system instead of the nonlinear system unlike the proposed ISMO.

- The restriction on the motions of the camera and object and the requirement of the velocity information in the SaM is clarified by introducing several scenarios. The SaM of an object using a dynamic (moving) camera can be estimated via the ISMO based on the relative motion model even for more general object motion and less camera velocity information than existing studies in [19], [37]–[50]. The following situations have been studied. The first case is when an object is in semigeneral motion (i.e., it initially remains static (stationary) for a certain period of time and is then in general (static or dynamic) motion) and the camera is in dynamic motion. The second case is when the object is in general motion within a constrained space and the camera is in dynamic motion. The third case is when both the object and camera are in general motions assuming that the range is available. The formulation of the SaM equations and the consideration of the several scenarios in the proposed ISMO-based method can obtain the closed-form solution of the SaM estimation problem, which has not been considered in [19], [37]–[50].
- Simulation and experimental results with a camera in dynamic motion and an object in semigeneral and general motions are provided to verify the proposed method. These results show that both the range and object velocity can be estimated well using the proposed method even when both the camera and object are in motion.

**II. RELATIVE CAMERA AND OBJECT MOTION DYNAMICS**

The motion of the feature points between consecutive camera images can be described in terms of two orthogonal coordinates attached to a camera,  $F^*$  and  $F_C$ . Although  $F^*$  is attached to the camera at the location corresponding to the initial point at initial time  $t_0$ ,  $F_C$  is changed from  $F^*$  after the initial time through the rotation matrix  $\bar{R} \in SO(3)$  and translation vector  $\bar{x}_f \in \mathbb{R}^3$ . Unlike the Euclidean coordinates of a feature point from the camera  $\bar{m} := [x_1, x_2, x_3]^T$  in the camera frame  $F_C$ , the normalized coordinates  $m := [x_1/x_3, x_2/x_3, 1]^T$  are available as they are related to the pixel coordinates  $p$  as

$$p = A_c m, \tag{1}$$

where  $p = [p_u, p_v, 1]^T \in \mathbb{R}^3$  and  $A_c \in \mathbb{R}^{3 \times 3}$  is a known intrinsic camera calibration matrix [32].

As the camera moves, the feature point  $q$  in  $F_C$  satisfies

$$\bar{m} = \bar{x}_f + \bar{R}x_{Oq}, \tag{2}$$

where the vector  $x_{Oq}$  defined from the origin of  $F^*$  to  $q$ . From (2), the relative motion model can be obtained as [32], [76]

$$\dot{\bar{m}} = [\omega]_X \bar{m} + v_r, \tag{3}$$

where  $[\omega]_X \in \mathbb{R}^{3 \times 3}$  is a skew-symmetric matrix defined as

$$[\omega]_X := \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix} \tag{4}$$

for the camera’s angular velocity  $\omega := [\omega_1, \omega_2, \omega_3]^T$  in  $F_C$ . Further,  $v_r$  is the camera’s linear velocity relative to  $q$  in  $F_C$ , defined as  $v_r := v_p - v_c$ , where  $v_p := [v_{px}, v_{py}, v_{pz}]^T$  is the object’s linear velocity and  $v_c := [v_{cx}, v_{cy}, v_{cz}]^T$  is the camera’s linear velocity.

Considering the situation in which i) the camera and object are in stable motion for SaM estimation and ii) the measurement or estimation of the angular velocity of the camera using two consecutive 2D images is carried out using epipolar geometry, the following assumption can be made.

*Assumption 1:* The camera velocity  $v_c$ , object velocity  $v_p$ , and their time derivatives are bounded owing to the stable relative camera and object motions. In addition,  $\omega$  is available.

On the basis of  $y := [y_1, y_2, y_3]^T = [x_1/x_3, x_2/x_3, 1/x_3]^T$  and (3), the relative motion dynamics can be obtained as

$$\dot{y} = \begin{pmatrix} -y_3 & 0 & y_1 y_3 & -y_1 y_2 & 1 + y_1^2 & -y_2 y_3 & 0 & -y_1 y_3 \\ 0 & -y_3 & y_2 y_3 & -(1 + y_2^2) & y_1 y_2 & y_1 & 0 & y_3 - y_2 y_3 \\ 0 & 0 & y_3^2 & -y_2 y_3 & y_1 y_3 & 0 & 0 & -y_3^2 \end{pmatrix} \cdot \begin{pmatrix} v_c \\ \omega \\ v_p \end{pmatrix} \tag{5}$$

where  $y_1$  and  $y_2$  are measurable from (1) and  $y_3$  should be estimated. The available variable  $\bar{y}_3$  satisfying

$$\bar{y}_3 = y_3/d_s \tag{6}$$

is then introduced as in [50] for an unknown positive constant  $d_s$  such that  $d_s$  needs to be estimated instead of the time-varying  $y_3$  for range estimation. The measurable vector  $\bar{y} := [y_1, y_2, \bar{y}_3]^T$  in [50] can be used to obtain from (5) the following relative motion dynamics:

$$\dot{\bar{y}} = \Omega + g, \tag{7}$$

where  $\Omega + g = \Omega_1 v_c d_s + \Omega_2 + \Omega_3 v_p d_s$ ,  $\Omega_i := [\Omega_{i1} \ \Omega_{i2} \ \Omega_{i3}]$ ,

$$\begin{aligned} \Omega_1 &= \begin{pmatrix} -\bar{y}_3 & 0 & y_1 \bar{y}_3 \\ 0 & -\bar{y}_3 & y_2 \bar{y}_3 \\ 0 & 0 & \bar{y}_3^2 \end{pmatrix}, \\ \Omega_2 &= \begin{pmatrix} -y_1 y_2 & 1 + y_1^2 & -y_2 \\ -(1 + y_2^2) & y_1 y_2 & -y_2 \\ -y_2 \bar{y}_3 & y_1 \bar{y}_3 & 0 \end{pmatrix} \omega, \\ \Omega_3 &= \begin{pmatrix} \bar{y}_3 & 0 & -y_1 \bar{y}_3 \\ 0 & \bar{y}_3 & -y_2 \bar{y}_3 \\ 0 & 0 & -\bar{y}_3^2 \end{pmatrix}. \end{aligned}$$

Here,  $\Omega_i$  for  $i = 1, 2, 3$  is a measurable nonsingular matrix and  $g$  is an unmeasurable vector, all of which will be appropriately defined later in Sections IV and V depending on the conditions of the camera and object. Since  $\bar{y}$  consists of three measurable variables, three variables among  $y_3(t)$  and the unknown components of  $v_p$  and  $v_c$  can be estimated by using an ISMO that will be designed based on the relative motion dynamics in (7) in the next section.

SaM estimation of an object can be still very difficult or even impossible to achieve unless the motion of either a dynamic camera or an object is restricted or at least some information of the camera velocity is given. Therefore, the following situations are considered as in Fig. 1: i) an object is in semigeneral motion and a camera is in dynamic motion (Fig. 1(a)); ii) an object is in general motion within a constrained space and a camera is in dynamic motion (Fig. 1(b)); and iii) both the object and camera are in general motions assuming the known range information (Fig. 1(c)). These cases are described in more detail in Sections IV–VI as in Table 1.

### III. INTEGRAL SLIDING MODE OBSERVER BASED ON A RELATIVE CAMERA–OBJECT MOTION MODEL

Since  $g$  is bounded from Assumption 1 as  $0 \leq \|g\|_\infty < \gamma$  for a constant  $\gamma$ , an ISMO is designed on the basis of (7) as

$$\dot{\hat{y}} = \Omega + \hat{g}, \tag{8}$$

where  $\hat{y}$  and  $\hat{g}$  are estimates of  $\bar{y}$  and  $g$ , respectively. The estimation error  $e_s := \bar{y} - \hat{y}$  and integral sliding surface are introduced as

$$s_e = e_s + e_{sI} \tag{9}$$

where  $e_{sI}$  is obtained from

$$\dot{e}_{sI} = \gamma \tanh\{(a + bt)e_s\} \tag{10}$$

with the initial condition  $e_{sI}(0) = -e_s(0)$ . Owing to this choice of  $e_{sI}(0)$ ,  $s_e(0)$  becomes zero from (9) and  $s_e$  will be maintained to zero thereafter by using the ISMO method described below so as to eliminate the reaching mode. Since  $e_s$  is the difference of  $\bar{y}$  and  $\hat{y}$  and both  $\bar{y}$  and  $\hat{y}$  are available,  $e_s(0)$  is available. Then, the estimate  $\hat{g}$  can be obtained as

$$\hat{g} = \bar{k}_1 s_e - v + \gamma \tanh\{(a + bt)e_s\}, \tag{11}$$

where the pseudoinput  $v$  and its derivative  $u$  are obtained from

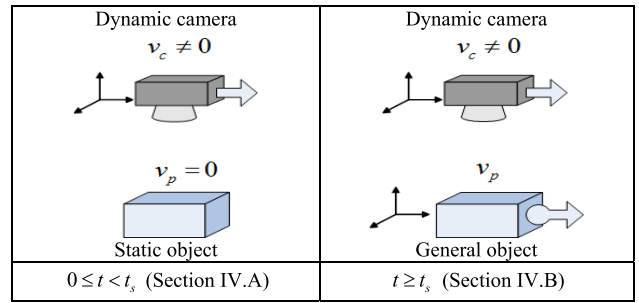
$$\dot{v} = u \tag{12}$$

$$u = \gamma \sec h^2\{(a + bt)s_e\} \cdot \{(a + bt)[\hat{g} - \gamma \tanh\{(a + bt)e_s\}] - bs_e\} - \gamma^2(a + bt) \sec h^2\{(a + bt)s_e\} \text{sgn}(s) - \bar{k}_2 \text{sgn}(s) - s_e, \tag{13}$$

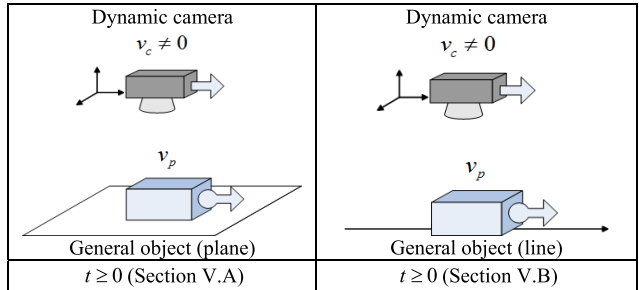
$\bar{k}_1, \bar{k}_2 > 0$  are constants, and  $s$  is another sliding surface given by

$$s := v + \gamma \tanh\{(a + bt)s_e\} \tag{14}$$

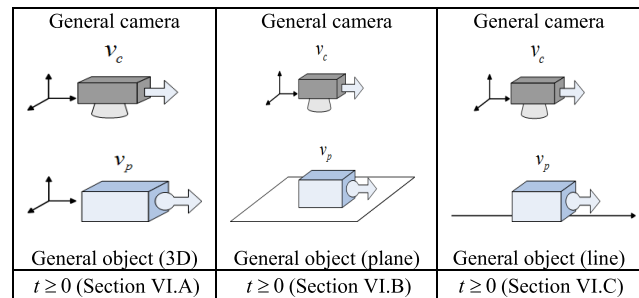
for positive constants  $a$  and  $b$  and  $v(0) = -\gamma \tanh\{as_e(0)\}$ , satisfying  $s(0) = 0$ . Here,  $\tanh(\cdot)$ ,  $\text{sech}(\cdot)$ , and  $|\cdot|$  in



(a) An object is in semigeneral motion in a general space and a dynamic camera moves with a partially known constant velocity.



(b) An object is in general motion within a constrained space and a dynamic camera has a non-zero velocity component orthogonal to a plane or a line.



(c) An object is in general motion and a general camera moves with either a completely known or partially known velocity when the range is available.

FIGURE 1. Scenarios of the SaM estimation of an object using a monocular dynamic camera.

TABLE 1. Conditions on object and camera motions for SaM estimation in sections IV–VI.

| camera \ object | object      |            |
|-----------------|-------------|------------|
|                 | semigeneral | general    |
| dynamic         | Section IV  | Section V  |
| general         | .           | Section VI |

(10)–(19) are assumed to be functions applied to each element of the argument for simplicity. Then,  $s_e$  and  $s$  can remain at zero, eliminating the reaching mode.

The stability and performance of the proposed ISMO designed on the basis of (7) can be analyzed as the following theorem.

*Theorem 1 (ISMO for the Relative Motion Model in (7)):* The state estimation errors between the actual states of the relative camera–object motion model in (7) and their estimates obtained by the ISMO in (8)–(14) under Assumption 1 are



stable in the sense that i)  $s \in L_\infty$ ; ii)  $e_s \in L_2 \cap L_\infty$ ; iii)  $s$ ,  $e_s$ , and  $\dot{e}_s$  converge to zero asymptotically; and iv)  $\hat{g}$  converges to  $g$  asymptotically.

*Proof:* The state estimation error dynamics can be expressed using (7)–(11) and (14) as

$$\begin{aligned}\dot{s}_e &= g - \hat{g} + \gamma \tanh\{(a + bt)e_s\} \\ &= g - \bar{k}_1 s_e + s - \gamma \tanh\{(a + bt)s_e\}.\end{aligned}\quad (15)$$

Owing to (11), the time derivative of  $V_1 = s_e^T s_e / 2$  can satisfy

$$\dot{V}_1 \leq -\bar{k}_1 s_e^T s_e + s_e^T s - \gamma \left| s_e^T \right| \left[ \tanh\{(a + bt)|s_e|\} - \alpha \right] \quad (16)$$

where  $0 \leq \alpha := \|g\|_\infty / \gamma < 1$ . For the design of the ISMO, the time derivative of  $s$  in (14) is arranged as

$$\begin{aligned}\dot{s} &= u + \gamma \sec h^2\{(a + bt)s_e\} \cdot \left\{ [(a + bt)[g - \hat{g} \right. \\ &\quad \left. + \gamma \tanh\{(a + bt)s_e\}] + bs_e \right\}.\end{aligned}\quad (17)$$

Substitution of (13) into (17) then gives

$$\begin{aligned}\dot{s} &= \gamma \sec h^2\{(a + bt)s_e\} \cdot (a + bt)g - \gamma^2(a + bt) \\ &\quad \cdot \sec h^2\{(a + bt)s_e\} \operatorname{sgn}(s) - \bar{k}_2 \operatorname{sgn}(s) - s_e.\end{aligned}\quad (18)$$

The time derivative of  $V_2 = s^T s / 2$  is obtained as

$$\begin{aligned}\dot{V}_2 &= \gamma s^T \sec h^2\{(a + bt)s_e\} (a + bt)g \\ &\quad - \gamma^2(a + bt) \sec h^2\{(a + bt)s_e\} |s| - \bar{k}_2 |s| - s^T s_e \\ &\leq -\bar{k}_2 |s| - s^T s_e.\end{aligned}\quad (19)$$

By combining (16) and (19), the time derivative of the Lyapunov function  $V := V_1 + V_2$  becomes

$$\dot{V} \leq -\bar{k}_1 s_e^2 - \bar{k}_2 |s| - \gamma \left| s_e^T \right| \left[ \tanh\{(a + bt)|s_e|\} - \alpha \right]. \quad (20)$$

If  $|s_e| \geq \bar{e}_s$  where  $\bar{e}_s := \ln\{(1 + \alpha)/(1 - \alpha)\} / \{2(a + bt)\}$  is bounded, then  $\dot{V}$  in (20) satisfies

$$\dot{V} \leq -\bar{k}_1 s_e^2 - \bar{k}_2 |s|. \quad (21)$$

Therefore,  $s_e$  and  $s$  are ultimately bounded, and their ultimate bounds dependent on  $\bar{e}_s$  decrease to zero as  $\bar{e}_s$  decreases to zero. Thus,  $|s_e| \geq \bar{e}_s$  and (21) hold as time goes on. From (21),  $s_e, s \in L_\infty$  and  $s_e \in L_2$  hold. Moreover,  $\dot{s}_e \in L_\infty$  holds from (15), implying the uniform continuity of  $s_e$ . Combining this with the  $L_2$ -property of  $s_e$ , Barbalat's lemma [75] can be used to show that  $s_e$  converges to zero asymptotically. The boundedness of  $\dot{s}_e$  can be also derived to show the asymptotic convergence of  $\dot{s}_e$  and  $e_s$  to zero from Assumption 1 and Barbalat's lemma, which leads to the convergence of  $\hat{g}$  to  $g$ . (Q.E.D.)

*Remark 1:* The basic idea of the ISMO is based on the choice of the integral sliding surface and the Lyapunov candidate function, both of which are defined in terms of the estimation error and are made to converge to zero such that unmeasurable  $g$  can be asymptotically estimated well. The details on the stability analysis used in the proof of Theorem 1 can be referred to [75]. It should be noted that whereas the integral sliding surface  $s_e$  in (9) is introduced to eliminate

the reaching mode, the sliding surface  $s$  in (14) is introduced to inhibit the chattering phenomenon in the presence of the signum function in (13) by using the idea of backstepping method [75].

On the basis of the ISMO in this section, an ISMO-based SaM estimation method will be designed in Sections IV, V, and VI.

#### IV. SAM ESTIMATION OF AN OBJECT IN SEMIGENERAL MOTION USING A PARTIALLY KNOWN CONSTANT DYNAMIC CAMERA

In this section, the case where a dynamic camera moves with a partially known nonzero constant velocity is considered. To estimate the range and object velocity, the following assumption for the object's motion is introduced.

*Assumption 2:* The object is in semigeneral motion; that is, it initially undergoes static motion followed by general motion.

Two ISMOs should be designed through the following two steps. First, a scale factor of the depth information (i.e.,  $d_s$ ) and the unknown components of the camera velocity can be obtained using the first ISMO in Section IV.A. The scale factor estimate is then used to estimate the range and object velocity using the second ISMO in Section IV.B. The structure of the SaM estimation method in this section is described in Fig. 2.

##### A. ESTIMATION OF THE RANGE AND UNKNOWN CONSTANT COMPONENTS OF THE CAMERA VELOCITY WHEN AN OBJECT IS IN STATIC MOTION

When an object remains static for a certain amount of time (i.e.,  $v_p = 0$  for  $0 \leq t < t_s$ ),  $\Omega$  and  $g$  in (7) are given by

$$\Omega = \Omega_2, \quad g = \Omega_1 v_c d_s. \quad (22)$$

To estimate  $y_3$ , the following assumption is introduced.

*Assumption 3:* One component of the camera velocity vector  $v_c$  (i.e., its  $i$ -th component  $v_{ci}$ ) is nonzero and known. The remaining components are unknown constants.

The range and unknown constant components of the camera velocity can be estimated using the ISMO in (8)–(14) as in the following corollary to Theorem 1.

*Corollary 1 (SaM Estimation of a Static Object):* Suppose that Assumptions 2 and 3 are satisfied and that the ISMO in (8)–(14) for the SaM equations in (7) and (22) is used. On the basis of (22), an estimate of  $d_s$  (i.e.,  $\hat{d}_s$ ) can then be obtained using the information of the known component  $v_{ci}$  as

$$\hat{d}_s = (e_i^T v_c)^{-1} e_i^T \Omega_1^{-1} \hat{g} = v_{ci}^{-1} e_i^T \Omega_1^{-1} \hat{g}. \quad (23)$$

Here,  $e_i$  is a column vector with the appropriate dimensions whose  $i$ -th element is 1 and every other element is zero. In addition, a range estimate can be obtained from (6) and (23) as

$$\hat{y}_3 = \hat{d}_s \bar{y}_3, \quad (24)$$

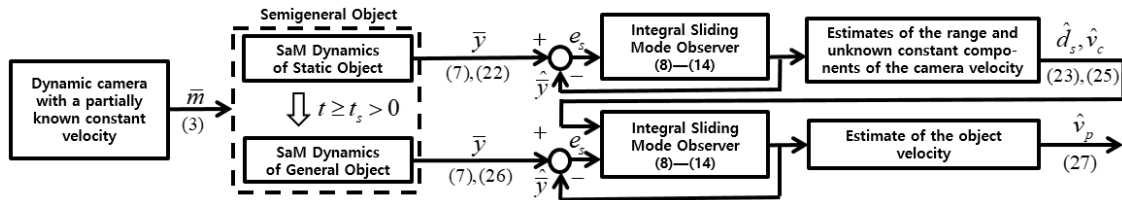


FIGURE 2. Structure of the SaM estimation in Section IV for a dynamic camera looking at a semi-general object.

and the estimate of  $v_c$  can be obtained from (22) as

$$\hat{v}_c = \Omega_1^{-1} \hat{g} / \hat{d}_s. \quad (25)$$

*Proof:*  $v_c d_s$  is the same as  $\Omega_1^{-1} g$  from (22) owing to the nonsingularity of  $\Omega_1$  and all components of  $\Omega_1^{-1} g$  are known from their estimates  $\Omega_1^{-1} \hat{g}$  via the ISMO as in Theorem 1. Since one component of  $v_c d_s$  is available,  $\hat{d}_s$  and  $\hat{v}_c$  can be obtained from (23) and (25). In addition, the range information  $y_3$  can be obtained from (6) as (24) using  $\hat{d}_s$  in (23). (Q.E.D.)

### B. ESTIMATION OF THE OBJECT VELOCITY WHEN AN OBJECT IS IN GENERAL MOTION

When a static object undergoes general motion (i.e.,  $t \geq t_s$ ), both the range and object velocity should be estimated. As the constant scaling factor  $d_s$  and range information are respectively available from (23) and (24), these estimates will be used to estimate the velocity of an object in general motion.

Using the available components of the camera velocity and the estimates of the unknown constant components of the camera velocity in (25),  $\Omega$  and  $g$  in (7) can be given by

$$\Omega = \Omega_1 v_c d_s + \Omega_2, \quad g = \Omega_3 v_p d_s. \quad (26)$$

The estimate of  $v_p$  then becomes

$$\hat{v}_p = \Omega_3^{-1} \hat{g} \hat{d}_s^{-1}. \quad (27)$$

Owing to the aforementioned arguments, the range and object velocity can then be estimated using the ISMO in (8)–(14) as in Corollary 2.

*Corollary 2 (SaM Estimation of an Object in General Motion):* Suppose that Assumptions 2 and 3 are satisfied and that the ISMO in (8)–(14) for the SaM equations in (7) and (26) is used. The object velocity can then be estimated as  $\hat{v}_p$  in (27).

*Proof:* Owing to the estimation performance of  $\hat{g}$  in Theorem 1 and the aforementioned arguments,  $v_p$  can be estimated by using (27). (Q.E.D.)

### V. SAM ESTIMATION OF AN OBJECT IN GENERAL MOTION IN A CONSTRAINED SPACE USING A DYNAMIC CAMERA

In this section, the following case is considered. The object is in general motion within a constrained space and the camera moves with a nonzero velocity component orthogonal to the object motion space (i.e., a space in which the object exists

such that the object motion space and camera motion space are not parallel to each other. Depending on whether a camera velocity is either completely or partially known, the object motion should be constrained on either a plane or a line for its SaM estimation. One of these situations would be the one where a camera in the air looks down on a vehicle moving either on flat ground or along a line on the ground. One of these situations is experimented in Section VII. The structure of the SaM estimation method in this section is described in Fig. 3.

### A. SAM ESTIMATION OF AN OBJECT IN GENERAL MOTION ON A PLANE USING A COMPLETELY KNOWN DYNAMIC CAMERA

In this case, an object moves along a plane or remains on it and a camera is in dynamic motion with a nonzero velocity component orthogonal to the object motion space. Thus, the following assumptions are made for an object in general motion and a dynamic camera.

*Assumption 4:* In the object velocity vector  $v_p$ , two components (i.e., its  $i$ -th and  $j$ -th components,  $v_{pi}$  and  $v_{pj}$ ) are unknown and the remaining component is zero (i.e.,  $v_{pk} = 0$ ). Thus, the object moves along a plane or remains on it.

*Assumption 5:* A camera velocity vector  $v_c$  is known and its component orthogonal to the unknown components of  $v_p$  satisfying Assumption 4 is nonzero (i.e.,  $v_{ck}$  orthogonal to  $v_{pi}$  and  $v_{pj}$  is nonzero, where  $k$  is different from  $i$  and  $j$  such that  $[\Omega_{3i}, \Omega_{3j}, \Omega_{1k}]$  becomes nonsingular).

Considering the situation where Assumptions 4 and 5 hold,  $\Omega$  and  $g$  in (7) are arranged as

$$\Omega = \Omega_2, \quad g = \bar{\Omega}_g \cdot [d_s, v_{pi} d_s, v_{pj} d_s]^T, \quad (28)$$

where  $\bar{\Omega}_g := [\Omega_1 v_c, \Omega_{3i}, \Omega_{3j}]$  is always nonsingular owing to the fact that  $v_{ck} \neq 0$  in Assumption 5 and thus the nonsingularity of  $[\Omega_{1k} v_{ck}, \Omega_{3i}, \Omega_{3j}]$  hold. On the basis of (28), the estimates of  $d_s$ ,  $v_{pi}$ , and  $v_{pj}$  can all be obtained from

$$[\hat{d}_s, \hat{v}_{pi} \hat{d}_s, \hat{v}_{pj} \hat{d}_s]^T = \bar{\Omega}_g^{-1} \hat{g}. \quad (29)$$

Therefore, both the range and object velocity can be estimated as in Corollary 3.

*Corollary 3 (SaM Estimation of an Object in General Motion on a Plane):* Suppose that Assumptions 4 and 5 are satisfied and that the ISMO in (8)–(14) for the SaM equations in (7) and (28) is used. Then,  $d_s$  (i.e.,  $y_3$ ) and  $v_p$  can be estimated using (29).

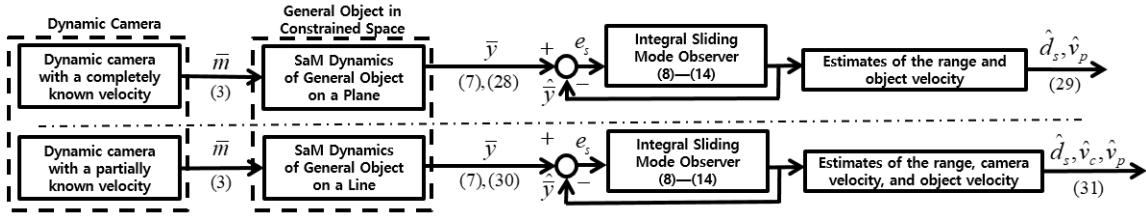


FIGURE 3. Structure of the SaM estimation in Section V for a general object in a constrained space using a dynamic camera.

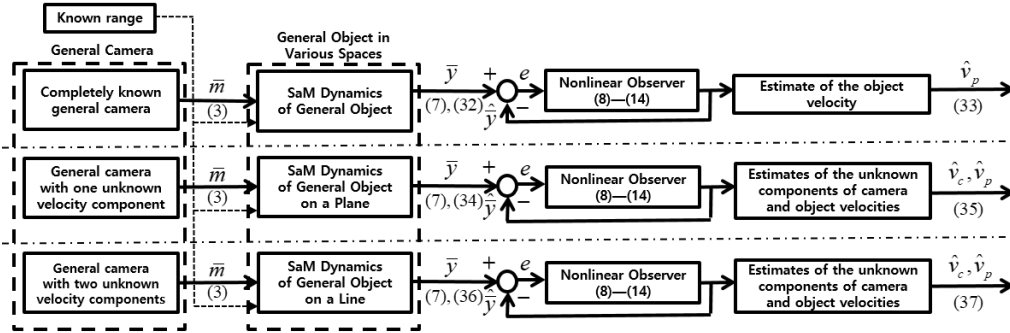


FIGURE 4. Structure of the motion estimation in Section VI for a general camera looking at a general object when the range is available.

*Proof:* Under Assumptions 4 and 5,  $d_s$  and the components of  $v_p$  (i.e.,  $v_{pi}$  and  $v_{pj}$ ) can be estimated using (29). This yields the estimate of  $y_3$  as (24). (Q.E.D.)

**B. SAM ESTIMATION OF AN OBJECT IN GENERAL MOTION ON A LINE USING A PARTIALLY KNOWN DYNAMIC CAMERA**

In this case, an object moves along a line or remains on it and a camera is in dynamic motion. The following assumptions are made for the object and camera.

*Assumption 6:* In the object velocity vector  $v_p$ , one component (i.e.,  $v_{pi}$ ) is unknown and the remaining two components are zero (i.e.,  $v_{pj} = v_{pk} = 0$ ), implying that the object must move along a line or remain on it.

*Assumption 7:* For an object with motion satisfying Assumption 6, a component of a camera velocity vector  $v_c$  parallel to  $v_{pi}$  (i.e.,  $v_{ci}$ ) is known and one of the two components orthogonal to  $v_{pi}$  (i.e.,  $v_{cj}$  among the components  $v_{cj}$  and  $v_{ck}$  orthogonal to  $v_{pi}$  where  $i, j$ , and  $k$  are different) is nonzero and known.

Considering the situation in which Assumptions 6 and 7 are satisfied,  $\Omega$  and  $g$  in (7) are given by

$$\Omega = \Omega_2, \quad g = \bar{\Omega}_g \cdot [d_s, v_{ck}d_s, v_{pi}d_s]^T, \quad (30)$$

where  $\bar{\Omega}_g := [\Omega_{1i}v_{ci} + \Omega_{1j}v_{cj}, \Omega_{1k}, \Omega_{3i}]$  is nonsingular owing to  $v_{cj} \neq 0$  in Assumption 7; thus,  $[\Omega_{1j}v_{cj}, \Omega_{1k}, \Omega_{3i}]$  is nonsingular. On the basis of (30), the estimates of  $d_s$ ,  $v_{ck}$ , and  $v_{pi}$  can then be obtained from

$$[\hat{d}_s, \hat{v}_{ck}\hat{d}_s, \hat{v}_{pi}\hat{d}_s]^T = \bar{\Omega}_g^{-1}\hat{g}. \quad (31)$$

For an object in general motion on a line as well, the range and object velocity can then be estimated by Corollary 4.

*Corollary 4 (SaM Estimation of an Object in General Motion on a Line):* Suppose that Assumptions 6 and 7 are satisfied and that the ISMO in (8)–(14) for the SaM equations in (7) and (30) is used. Then,  $d_s$  (i.e.,  $y_3$ ), an unknown component of  $v_c$  (i.e.,  $v_{ck}$ ), and  $v_p$  (i.e.,  $v_{pi}$ ) can all be estimated using (31).

*Proof:* Corollary 3 and the aforementioned arguments can be used to obtain the results in this corollary. Since only  $v_{pi}$  among the components of  $v_p$  needs to be estimated from Assumption 6,  $d_s$  and another unknown component  $v_{ck}$  can be estimated by using (31) since non-zero  $v_{cj}$  is known from Assumption 7. (Q.E.D.)

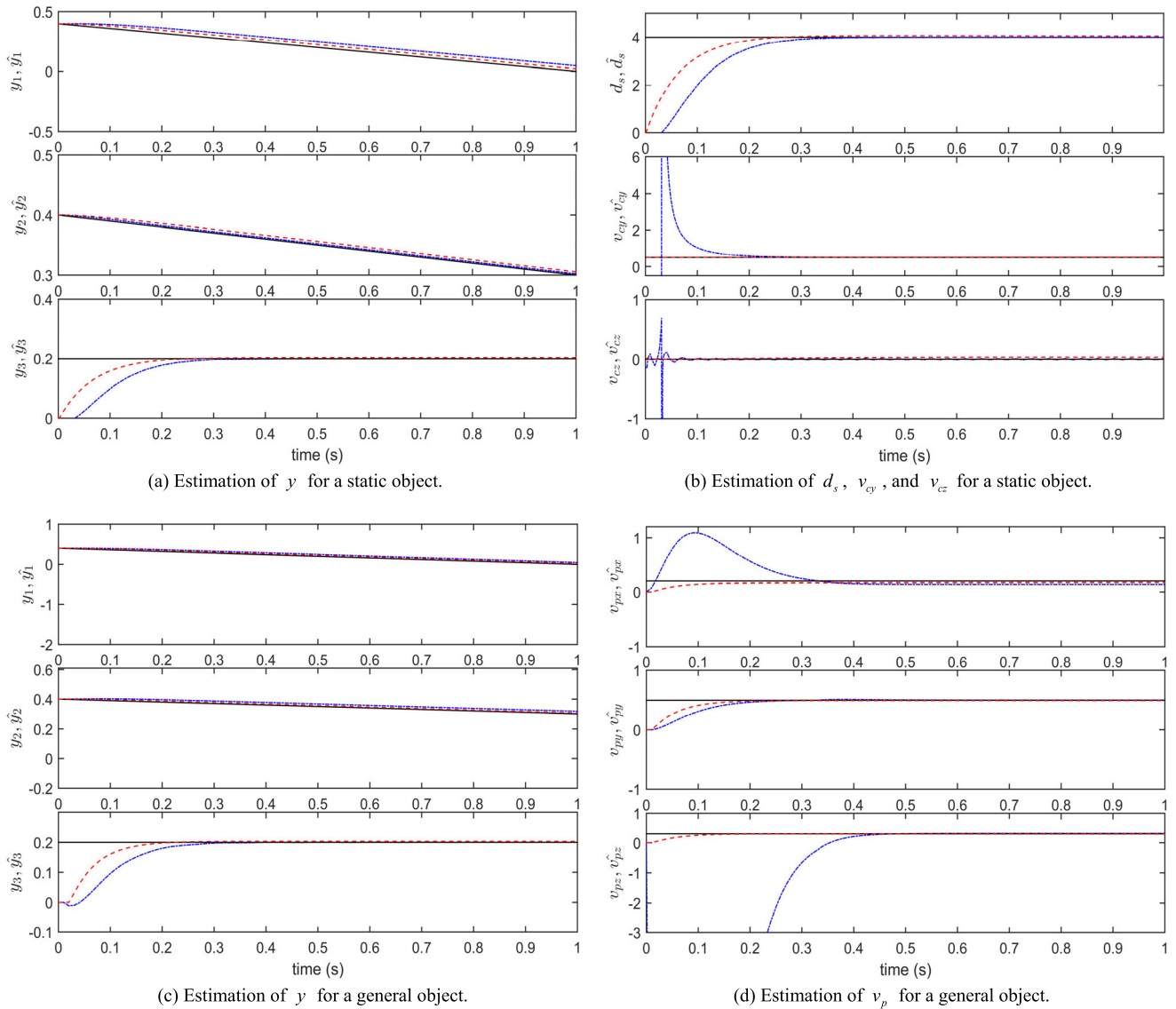
**VI. MOTION ESTIMATION OF A GENERAL OBJECT USING A GENERAL CAMERA WHEN THE RANGE IS AVAILABLE**

In this section, the motions of the object and camera are estimated when the range is available as in Assumption 8.

*Assumption 8:* The range information  $y_3$  (i.e.,  $d_s$ ) is available.

In this case, the motions of the object and camera can be more general as shown in Fig. 1(c) and the requirement regarding the available information of the camera in the previous sections can be relieved.

Depending on the number of available components of the camera velocity, the motion of the object should be constrained as in the following subsections. The structure of the motion estimation method in this section is described in Fig. 4.



**FIGURE 5.** Simulation results using the proposed method for an object in semigeneral motion and a camera in dynamic motion in Section IV (black solid: actual; blue dash-dotted: estimate by the SaM method using [50]; red dashed: estimate by the proposed method).

**A. MOTION ESTIMATION OF A GENERAL OBJECT USING A GENERAL CAMERA WITH COMPLETELY KNOWN VELOCITY**

In this case, the camera velocity is completely known and the object can have a general motion. Thus, the following assumption is introduced instead of Assumptions 3, 5, and 7.

*Assumption 9:* The camera velocity vector  $v_c$  is completely known.

Considering the situation where Assumptions 8 and 9 hold,  $\Omega$  and  $g$  in (7) are arranged as

$$\Omega = \Omega_1 v_c d_s + \Omega_2, \quad g = \Omega_3 d_s v_p. \quad (32)$$

On the basis of (32), the estimate of  $v_p$  can be obtained as

$$\hat{v}_p = \Omega_3^{-1} \hat{g} d_s^{-1}. \quad (33)$$

Therefore, the object velocity can be estimated as in Corollaries 5, 6, and 7, the proofs of which are straightforward and thus will be omitted.

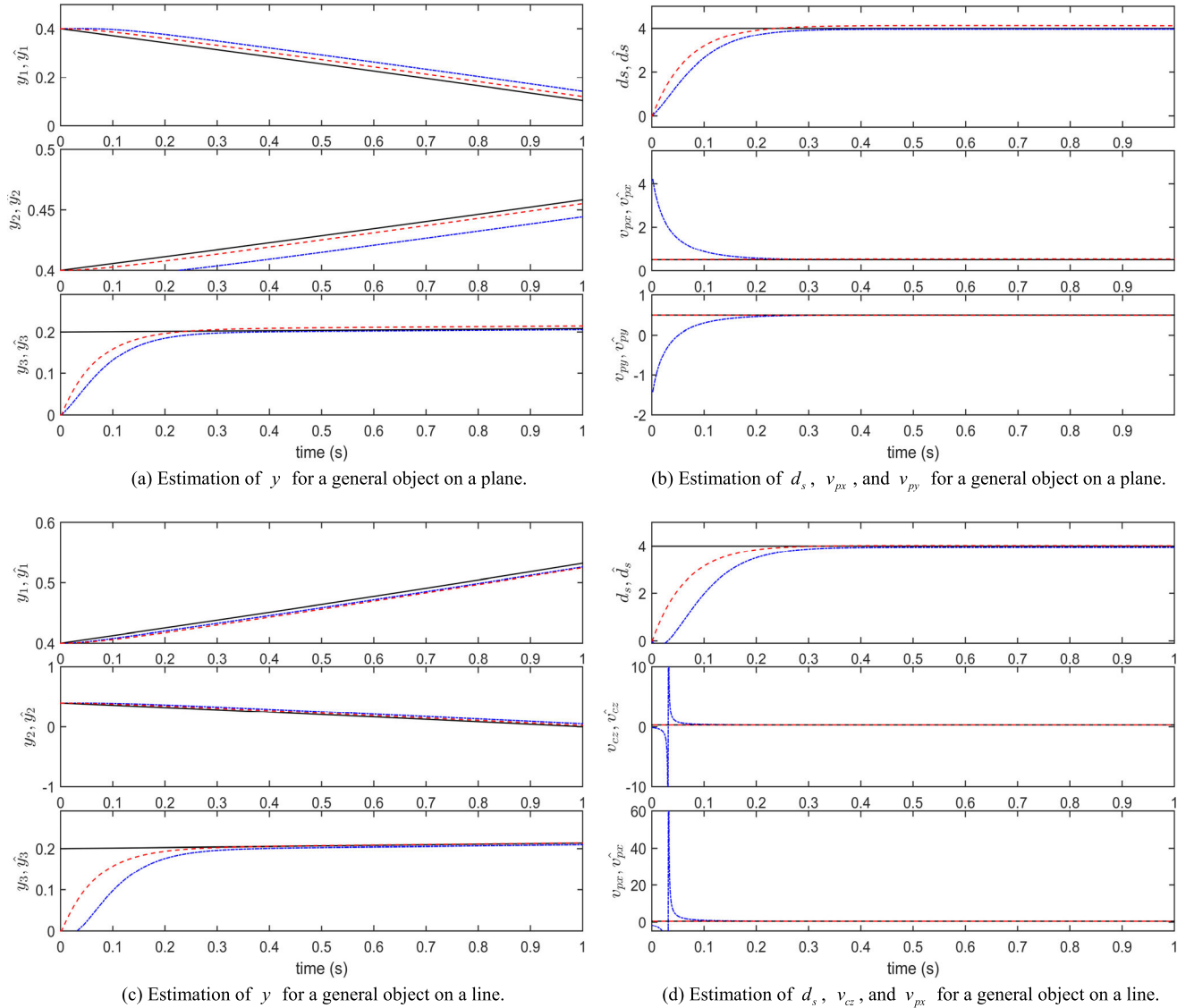
*Corollary 5 (Motion Estimation of an Object in General Motion Using the Known Range):* Suppose that Assumptions 8 and 9 are satisfied and that the ISMO in (8)–(14) for the SaM equations in (7) and (32) is used. Then,  $v_p$  can be estimated using (33).

**B. MOTION ESTIMATION OF A GENERAL OBJECT ON A PLANE USING A GENERAL CAMERA WITH ONE UNKNOWN VELOCITY COMPONENT**

In this case, one component of the camera velocity is unknown and the object should move or remain along a plane. Thus, instead of Assumption 5, the following assumption is introduced.

*Assumption 10:* In the non-zero camera velocity vector  $v_c$ , one component orthogonal to the unknown components of





**FIGURE 6.** Simulation results using the proposed method for an object in general motion in a constrained space and a camera in dynamic motion in Section V (black solid: actual; blue dash-dotted: estimate by the SaM method using [50]; red dashed: estimate by the proposed method).

$v_p$  satisfying Assumption 4 (i.e.,  $v_{ck}$ , where  $k$  is different from  $i$  and  $j$  such that  $[\Omega_{3i}, \Omega_{3j}, \Omega_{3k}]$  becomes nonsingular) is unknown and the remaining two components (i.e.,  $v_{ci}$  and  $v_{cj}$ ) are known.

Considering the situation where Assumptions 4, 8, and 10 hold,  $\Omega$  and  $g$  in (7) are arranged as

$$\Omega = \Omega_{1i}v_{ci}d_s + \Omega_{1j}v_{cj}d_s + \Omega_2, \quad g = \bar{\Omega}_g \cdot [v_{ck}, v_{pi}, v_{pj}]^T \quad (34)$$

where  $\bar{\Omega}_g := [\Omega_{1k}d_s, \Omega_{3i}d_s, \Omega_{3j}d_s]$  is always nonsingular. In this case, the estimates of  $v_{ck}$ ,  $v_{pi}$ , and  $v_{pj}$  can be obtained from

$$[\hat{v}_{ck}, \hat{v}_{pi}, \hat{v}_{pj}]^T = \bar{\Omega}_g^{-1} \hat{g}. \quad (35)$$

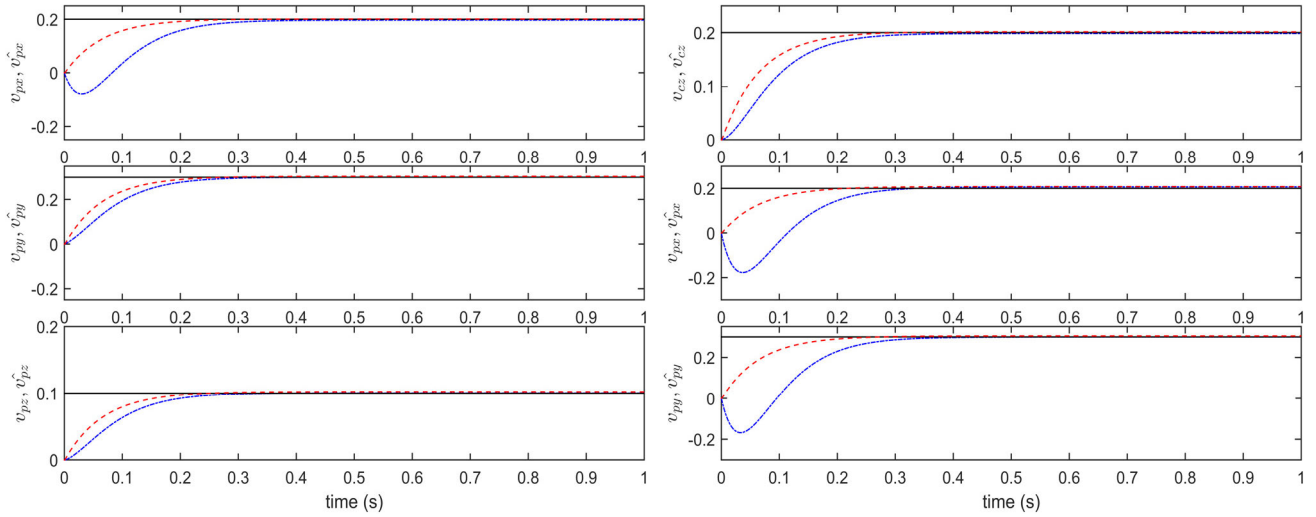
Therefore, the unknown components of the camera and object velocities can be estimated as in Corollary 6.

*Corollary 6 (Motion Estimation of a General Object on a Plane Using the Known Range):* Suppose that Assumptions 4, 8, and 10 are satisfied and that the ISMO in (8)–(14) for the SaM equations in (7) and (34) is used. Then, unknown components of  $v_c$  (i.e.,  $v_{ck}$ ) and  $v_p$  (i.e.,  $v_{pi}$  and  $v_{pj}$ ) can be estimated using (35).

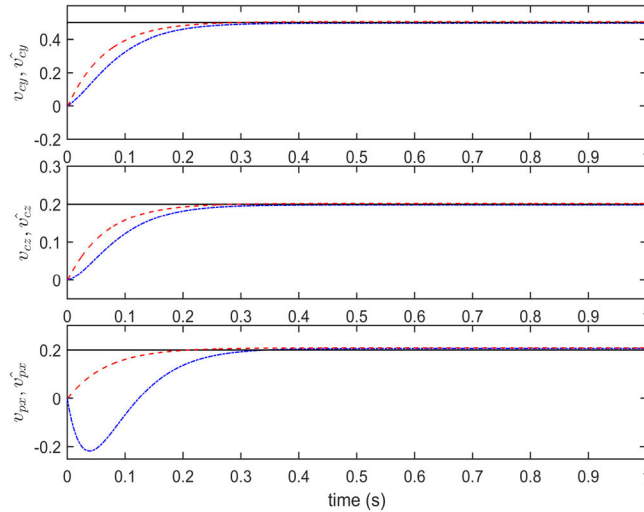
### C. MOTION ESTIMATION OF A GENERAL OBJECT ON A LINE USING A GENERAL CAMERA WITH TWO UNKNOWN VELOCITY COMPONENTS

In this case, two components of the camera velocity are unknown and the object should move or remain along a line. Thus, instead of Assumption 7, the following assumption is introduced.

*Assumption 11:* In the non-zero camera velocity vector  $v_c$ , two components orthogonal to the unknown component of  $v_p$  satisfying Assumption 6 (i.e.,  $v_{ci}$  and  $v_{ck}$ , where  $j$



(a) Estimation of  $v_p$  for a completely known general camera and a general object. (b) Estimation of  $v_{cz}, v_{px}, v_{py}$  for a partially known general camera and a general object on a plane.



(c) Estimation of  $v_{cy}, v_{cz}, v_{px}$  for a partially known general camera and a general object on a line.

**FIGURE 7.** Simulation results using the proposed method for a general object and a general camera when the range is available in Section VI (black solid: actual; blue dash-dotted: estimate by the SaM method using [50]; red dashed: estimate by the proposed method).

and  $k$  are different from  $i$  such that  $[\Omega_{3i}, \Omega_{3j}, \Omega_{3k}]$  becomes nonsingular) are unknown and the remaining one component (i.e.,  $v_{ci}$ ) is known.

Considering the situation where Assumptions 6, 8, and 11 are satisfied,  $\Omega$  and  $g$  in (7) are arranged as

$$\Omega = \Omega_{1i}v_{ci}d_s + \Omega_2, \quad g = \bar{\Omega}_g \cdot [v_{cj}, v_{ck}, v_{pi}]^T \quad (36)$$

where  $\bar{\Omega}_g := [\Omega_{1j}d_s, \Omega_{1k}d_s, \Omega_{3i}d_s]$  is nonsingular. On the basis of (36), the estimates of  $v_{cj}, v_{ck}$ , and  $v_{pi}$  can be obtained from

$$[\hat{v}_{cj}, \hat{v}_{ck}, \hat{v}_{pi}]^T = \bar{\Omega}_g^{-1} \hat{g}. \quad (37)$$

Therefore, the unknown components of the camera and object velocities can be estimated as in Corollary 7.

*Corollary 7 (Motion Estimation of a General Object on a Line Using the Known Range):* Suppose that Assumptions 6, 8, and 11 are satisfied and that the ISMO in (8)–(14) for the SaM equations in (7) and (36) is used. Then, unknown components of  $v_c$  (i.e.,  $v_{cj}$  and  $v_{ck}$ ) and  $v_p$  (i.e.,  $v_{pi}$ ) can be estimated using (37).

*Remark 2:* When the range information is available as Assumption 8, the dimension of the object motion space (i.e., the space where the object can move or remain) increases owing to the known range (the dimension of which is 1). Therefore, whereas the object should be constrained on a line or a plane depending on the conditions of the camera and object as described in Section V, it can be less constrained on a plane or even unconstrained in a 3D space in this section for the same conditions of the camera and object.

## VII. SIMULATION AND EXPERIMENTAL RESULTS

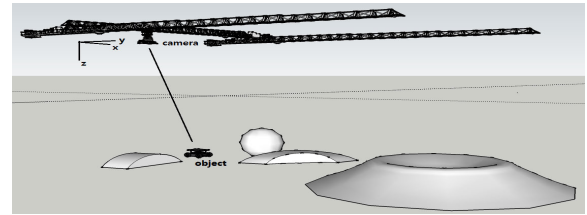
Simulation and experimental results will demonstrate the performance of the proposed SaM estimation method. The subscripts  $i, j$ , and  $k$  in the components of  $v_c$  and  $v_p$  are set to be the same as  $x, y$ , and  $z$  without loss of generality.

### A. SIMULATION RESULTS

The ISMO gains in (8)–(14) are chosen as  $\bar{k}_1 = \gamma = 15$ ,  $a = \bar{k}_2 = 1$ , and  $b = 0.1$ . The design parameters including  $\gamma$  in the integral sliding surface (14) need to be chosen through the trial and error by observing the estimation performance, as in the case of nonlinear control methods. The initial conditions for several estimates are chosen as  $\hat{g}(0) = \hat{y}(0) = [0, 0, 0]^T$  and the initial target location is selected with respect to the initial camera frame as  $y(0) = [0.4, 0.4, 0.2]^T$  (m). In addition, camera calibration matrix  $A_c := [A_{c1}, A_{c2}, A_{c3}]$  in (1) is set to be  $A_{c1} = [720, 0, 0]^T$ ,  $A_{c2} = [0, 720, 0]^T$ , and  $A_{c3} = [320, 240, 1]^T$ . The unknown constant  $d_s$  is chosen to be 4 such that  $\bar{y}(0) = [0.4, 0.4, 0.05]^T$  (m). The angular velocity of the camera is simply set to zero (i.e.,  $\omega = [0, 0, 0]^T$  (rad/s)) and the linear velocities of the camera and the object have different values in each scenario. To validate the robustness of the proposed method, white Gaussian noise with a signal-to-noise ratio of 100 dB is added to the image pixels. For comparison, the method in [50], which was originally developed for the SaM estimation of a monocular camera using two objects, has been formulated into a method for the SaM estimation of a single object using a monocular camera, which will be referred to as “SaM method using [50]” for brevity. The simulation results of the proposed method described in Sections IV, V, and VI are shown in Figs. 5, 6, and 7.

Fig. 5 shows the performance of SaM estimation for a dynamic camera looking at an object in semigeneral motion. First, in the case of static object motion, the ISMO for the estimate of  $\bar{y}$  in (8)–(14) is used such that  $\hat{y}$  can asymptotically converge to  $y$  in Fig. 5(a). As the variable  $y$  is determined by the position variable  $x$  through  $y = [x_1/x_3, x_2/x_3, 1/x_3]^T$ , Figs. 5(a), 6(a), and 7(a) indirectly show the relative position between the camera and object is changing with time. Using the estimate of  $g$  in (11),  $\hat{d}_s$  and the estimates of the unknown components of  $v_c$  (i.e.,  $\hat{v}_{cy}$  and  $\hat{v}_{cz}$ ) can be obtained using (23) and (25) in Fig. 5(b), which indicates that their estimates converge to the actual values within a short period of time. On the contrary,  $\hat{v}_{cy}$  and  $\hat{v}_{cz}$  obtained using the SaM method using [50] initially become too large to be used in practice, although they eventually converge to the actual values. Once  $d_s$  and  $v_c$  are available from their estimates based on a static object,  $\bar{y}$  and  $v_p$  in the case of an object in general motion can be estimated using (6), (23), (24), and (27) in Figs. 5(c) and 5(d). In this case as well,  $\hat{v}_{pz}$  by the SaM method using [50] is initially too large.

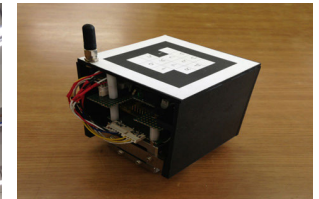
Fig. 6 shows the performance of the SaM estimation of an object in general motion within a constrained space using a dynamic camera. The motion of this object is constrained along a plane (or a line). The ISMO in (8)–(14) is used



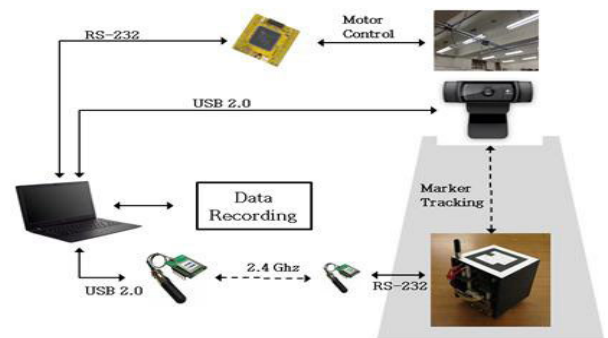
(a) Structure of SaM estimation system with a camera and an object.



(b) Dynamic camera on the ceiling.



(c) Mobile robot object.

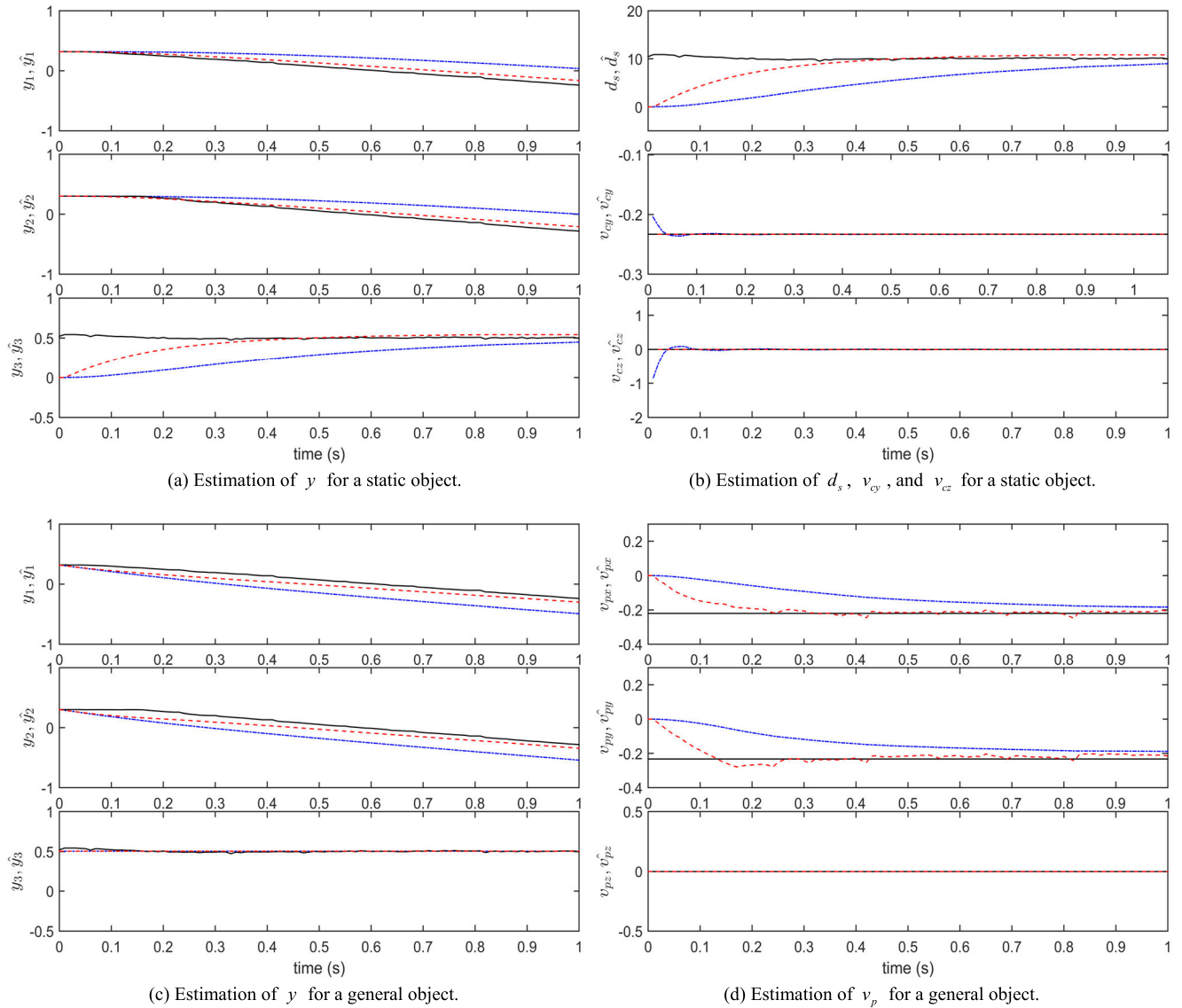


(d) Experimental structure of the SaM estimation system.

**FIGURE 8.** Experimental environment for SaM estimation.

such that  $\hat{y}$  can asymptotically converge to  $y$  in Fig. 6(a) (or Fig. 6(c)). Furthermore, the estimate of  $g$  in (11) can be used to obtain  $\hat{d}_s$ ,  $\hat{v}_p$ , and/or an unknown component of  $\hat{v}_c$  in Fig. 6(b) (or Fig. 6(d)). Although  $\hat{v}_{px}$  and  $\hat{v}_{py}$  immediately converge to their actual values by using the proposed method in Fig. 6(b), their transient responses of the SaM method using [50] are much worse. In Fig. 6(d),  $\hat{v}_{cz}$  and  $\hat{v}_{px}$  by the SaM method using [50] show greatly degraded transient responses. On the other hand, the proposed method achieves fast estimation performance for  $\hat{v}_{cz}$  and  $\hat{v}_{px}$  without almost any transient estimation errors.

Fig. 7 shows the performance of the motion estimation of a general camera and a general object within various spaces when the range is available. When all (or two or one) of the components of the velocity of the general camera are known and the motion of the general object is unconstrained (or constrained along a plane or constrained along a line), the ISMO in (8)–(14) is used such that  $\hat{d}_s$ , unknown components of  $\hat{v}_c$ , and all (or two or one) unknown components of  $\hat{v}_p$  are estimated in Fig. 7(a) (or Fig. 7(b) or Fig. 7(c)). Since the range is assumed to be known,  $\hat{y}$  does not need to be provided in the present case. All of the simulation results in Figs. 5–7 indicate that the estimates using the proposed method quickly converge to their actual values well.



**FIGURE 9.** Experimental results using the proposed method for an object in semigeneral motion in Section IV (black solid: actual; black dotted: estimate by the SaM method using [50]; red dashed: estimate by the proposed method).

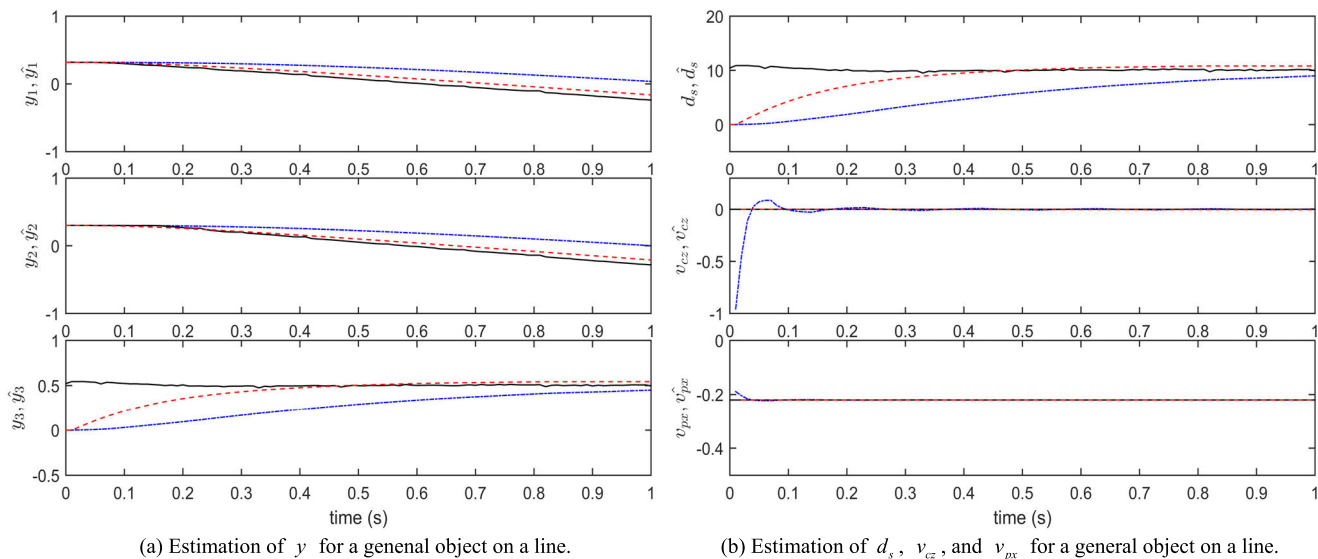
**B. EXPERIMENTAL ENVIRONMENT AND RESULTS**

The experimental environment is shown in Fig. 8. Fig. 8(a) shows the configuration of SaM estimation system, by which the performance of the proposed ISMO-based SaM estimation method in Sections IV.A, IV.B, V.B, VI.B, and VI.C is experimentally shown in Figs. 9, 10, and 11. A camera moving on the ceiling in Fig. 8(b) is used for observing from above a mobile robot object in Fig. 8(c) moving on the floor. Moreover, the SaM reference information is obtained by transferring the object information visualized by the camera to the external computer in Fig. 8(d). Although the feature points required for the implementation of the proposed method has been obtained by using the in-lab setting in Fig. 8, the state variables of the proposed ISMO are dependent on only the feature points and there is no other restriction in obtaining the required state variables. In this sense, it can be

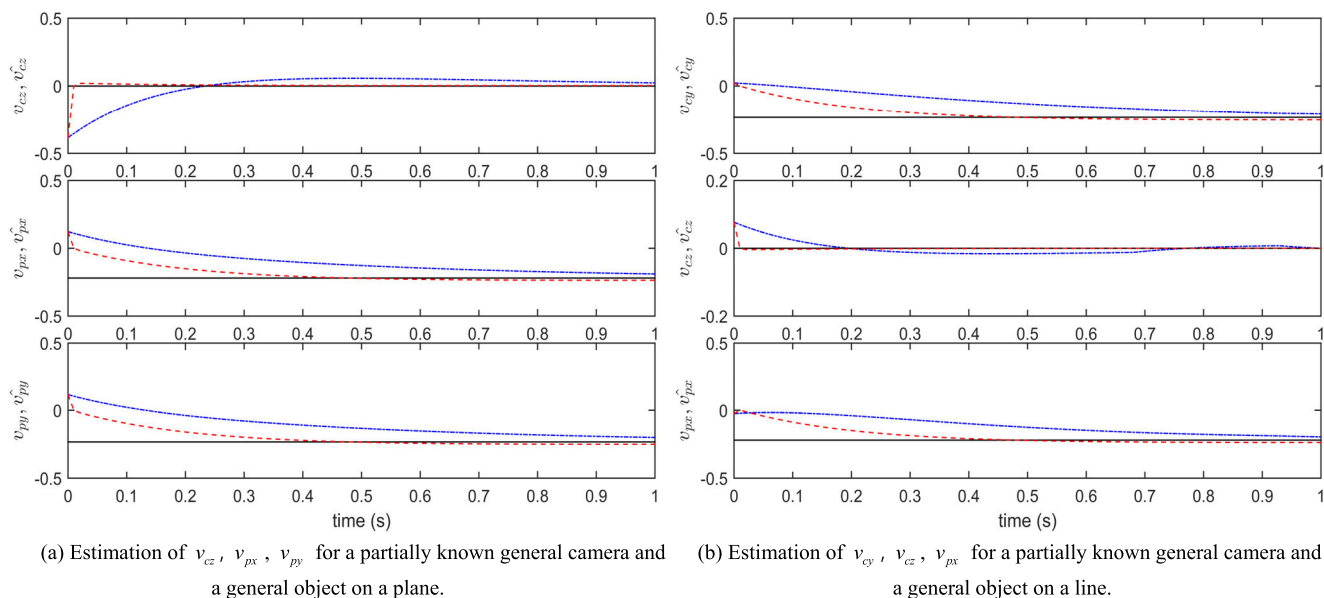
said that the proposed method can be applied to any image sequences as long as the feature points of the object can be readily extracted and the assumptions on the camera and object motions in Sections IV, V, and VI are satisfied. The presented work has focused more on the SaM problem than the feature extraction issues.

The proposed observer gains in (12) and the initial conditions for several estimates are chosen to be the same as those used to obtain the simulation results. On the other hand, the initial conditions of the several estimates such as  $\hat{v}_c$  and  $\hat{v}_p$  are set to be different depending on the scenarios to show that the performance of the SaM estimation using the proposed observer does not depend on the initial conditions of the object position, object velocity, camera position, and camera velocity. The initial location of the point on the target with respect to the initial camera frame is selected as





**FIGURE 10.** Experimental results using the proposed method for an object in general motion on a line in Section V.B (black solid: actual; blue dash-dotted: estimate by the SaM method using [50]; red dashed: estimate by the proposed method).



**FIGURE 11.** Experimental results using the proposed method for a general object and a single general camera when the range is available in Sections VI.B and VI.C (black solid: actual; blue dash-dotted: estimate by the SaM method using [50]; red dashed: estimate by the proposed method).

$y(0) = [0.32, 0.3, 0.52]^T$  (m) and the unknown constant  $d_s$  is chosen to be 10.4 such that  $\bar{y}(0) = [0.32, 0.3, 0.05]^T$  (m). The angular velocity of the camera is simply set to zero (i.e.,  $\omega = [0, 0, 0]^T$  (rad/s)). The camera's linear velocity  $v_c$  is respectively selected as  $[-0.22, -0.23, 0]^T$  (m/s),  $[-0.44, -0.47, 0]^T$  (m/s),  $[-0.44, -0.23, 0]^T$  (m/s),  $[-0.44, -0.47, 0]^T$  (m/s),  $[-0.44, -0.23, 0]^T$  (m/s) for the cases in Sections IV.A, IV.B, V.B, VI.B, and VI.C. The object's velocity  $v_p$  is respectively selected as  $[0, 0, 0]^T$  (m/s),  $[-0.22, -0.23, 0]^T$  (m/s),  $[-0.22, 0, 0]^T$  (m/s),  $[-0.22, -0.23, 0]^T$  (m/s),  $[-0.22, 0, 0]^T$  (m/s) for the cases

in Sections IV.A, IV.B, V.B, VI.B, and VI.C. Experimental results for Sections V.A and VI.A are not provided owing to the fact that  $v_{cz} = 0$  in the considered experimental environment. As the results of the proposed method is not dependent on any particular component of the coordinates, the validity of the proposed method can be evaluated by the experimental results of Sections IV, V.B, VI.B, and VI.C in Figs. 9, 10, and 11.

Fig. 9 shows the experimental results for the SaM estimation performance for the case in Section IV where a dynamic camera looks at an object in semigeneral motion.

Since  $v_c$  is set to be constant, only one of its nonzero components (i.e.,  $v_{cx}$ ) is assumed to be available, and  $v_{cy}$  and  $v_{cz}$  can be estimated in this case. Figs. 9(a) and 9(b) show that the estimates  $\hat{y}$ ,  $\hat{d}_s$ ,  $\hat{v}_{cy}$ , and  $\hat{v}_{cz}$  converge to their actual values well for a static object motion. Using these estimates,  $\hat{y}$  and  $v_p$  for an object in general motion can also be estimated well in Figs. 9(c) and 9(d).

Fig. 10 shows the experimental results for the case in Section V.B for a dynamic camera on a plane looking at an object in general motion on a line. Here,  $v_{cz}$  is set to zero in the experiments, and its estimate  $\hat{v}_{cz}$  immediately becomes zero when using the proposed method. Moreover, the known nonzero components  $v_{cy}$  is used to obtain  $\hat{y}$ ,  $\hat{d}_s$ ,  $\hat{v}_{cz}$ , and  $\hat{v}_{px}$  of an object moving along a line in Figs. 10(a) and 10(b).

Fig. 11 shows the experimental results for the case in Sections VI.B and VI.C where a general camera looks at a general object on a plane and a line. In all of these results, the unknown components of  $v_c$  and  $v_p$  are shown to be estimated well. The improved performance of the proposed method over the SaM method using [50] can be seen in the experimental results in Figs. 9, 10, and 11.

## VIII. CONCLUSION

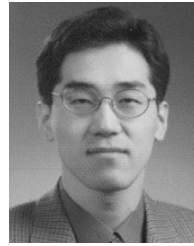
An ISMO-based method for estimating the SaM of a single object was proposed using a monocular camera for various camera and object motions. Unlike the existing methods, our work can be applied even with the constraints on the object and camera motions, such as a dynamic camera and an object in semigeneral and general motions. Studying the situation where a camera freely moves and observes a dynamic object is more interesting. However, without any restriction on the camera and object motions it can be very difficult to estimate altogether the information of range, camera velocity, and object velocity. Therefore, both a camera and an object in general motions are considered by assuming the known range information such that the motions of a general object and a general camera can be estimated. Although the range is time-varying and both the object and camera are in motion, the proposed ISMO can work well such that the transient estimation performance of the range and object velocity becomes satisfactory. Several interesting issues for the future work need to be considered. First, the SaM estimation of an object using a monocular camera without any camera velocity information was not considered in this paper, which can be pursued in a further separate study. Second, the extension of the proposed method to the cases of multiple objects and multiple cameras can be pursued by reformulating the SaM depending on the conditions and the number of objects and cameras. In this case, whether a static object exists and its information is available may be important in handling the cases of multiple objects. Third, the application of the proposed method to various vision-based robot systems with a camera can be studied by considering that the SaM estimation as a typical computer vision problem relies heavily on the camera characteristics and a various object motion such as the rotation or sudden changing motion.

## REFERENCES

- [1] Y. Yakimovsky and R. Cunningham, "A system for extracting three-dimensional measurements from a stereo pair of TV cameras," *Comput. Graph. Image Process.*, vol. 7, no. 2, pp. 195–210, Apr. 1978.
- [2] S. T. Barnard and M. A. Fischler, "Computational stereo," *ACM Comput. Surv.*, vol. 14, no. 4, pp. 553–572, Dec. 1982.
- [3] W. E. L. Grimson, "Computational experiments with a feature based stereo algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-7, no. 1, pp. 17–33, Jan. 1985.
- [4] U. R. Dhond and J. K. Aggarwal, "Structure from stereo—A review," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. SMC-19, no. 6, pp. 1489–1510, Nov./Dec. 1989.
- [5] Q. Zhang and T.-J. Chin, "Coresets for triangulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2095–2108, Sep. 2018.
- [6] M. Alterman, Y. Y. Schechner, and Y. Swirski, "Triangulation in random refractive distortions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 603–616, Mar. 2017.
- [7] S. Avidan and A. Shashua, "Trajectory triangulation: 3D reconstruction of moving points from a monocular image sequence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 4, pp. 348–357, Apr. 2000.
- [8] J. Y. Kaminski and M. Teicher, "A general framework for trajectory triangulation," *J. Math. Imag. Vis.*, vol. 21, no. 1, pp. 27–41, Jul. 2004.
- [9] M. Han and T. Kanade, "Reconstruction of a scene with multiple linearly moving objects," *Int. J. Comput. Vis.*, vol. 59, no. 3, pp. 285–300, Sep. 2004.
- [10] G. Wang, H.-T. Tsui, Z. Hu, and F. Wu, "Camera calibration and 3D reconstruction from a single view based on scene constraints," *Image Vis. Comput.*, vol. 23, no. 3, pp. 311–323, Mar. 2005.
- [11] D. J. Crandall, A. Owens, N. Snavely, and D. P. Huttenlocher, "SfM with MRFs: Discrete-continuous optimization for large-scale structure from motion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2841–2853, Dec. 2013.
- [12] D. Crandall, A. Owens, N. Snavely, and D. Huttenlocher, "Discrete-continuous optimization for large-scale structure from motion," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Colorado Springs, CO, USA, Jun. 2011, pp. 3001–3008.
- [13] S. Cao and N. Snavely, "Minimal scene descriptions from structure from motion models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 461–468.
- [14] R. Vidal and R. Hartley, "Three-view multibody structure from motion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 214–227, Feb. 2008.
- [15] T. Collins and A. Bartoli, "Planar structure-from-motion with affine camera models: Closed-form solutions, ambiguities and degeneracy analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1237–1255, Jun. 2017.
- [16] E. Zheng, D. Ji, E. Dunn, and J.-M. Frahm, "Self-expressive dictionary learning for dynamic 3D reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2223–2237, Sep. 2016.
- [17] T. Simon, J. Valmadre, I. Matthews, and Y. Sheikh, "Kronecker–Markov prior for dynamic 3D reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2201–2214, Nov. 2017.
- [18] H. S. Park and Y. Sheikh, "3D reconstruction of a smooth articulated trajectory from a monocular image sequences," in *Proc. Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 201–208.
- [19] H. S. Park, T. Shiratori, I. Matthews, and Y. Sheikh, "3D reconstruction of a moving point from a series of 2D projections," in *Proc. 11th Eur. Conf. Comput. Vis.*, Crete, Greece, Sep. 2010, pp. 158–171.
- [20] H. S. Park, T. Shiratori, I. Matthews, and Y. Sheikh, "3D trajectory reconstruction under perspective projection," *Int. J. Comput. Vis.*, vol. 115, no. 2, pp. 115–135, Nov. 2015.
- [21] M. Vo, S. G. Narasimhan, and Y. Sheikh, "Spatiotemporal bundle adjustment for dynamic 3D reconstruction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, Jun. 2016, pp. 1710–1718.
- [22] A. Agudo, F. Moreno-Noguer, B. Calvo, and J. M. M. Montiel, "Sequential non-rigid structure from motion using physical priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 5, pp. 979–994, May 2016.
- [23] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade, "Nonrigid structure from motion in trajectory space," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*. Cambridge, MA, USA: MIT Press, 2008, pp. 41–48.

- [24] L. Wolf and A. Shashua, "On projection matrices  $\mathcal{P}^k \rightarrow \mathcal{P}^2, k=3, \dots, 6$ , and their applications in computer vision," *Int J. Comput. Vis.*, vol. 48, no. 1, pp. 53–67, 2002.
- [25] J. Valmadré and S. Lucey, "General trajectory prior for non-rigid reconstruction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 1394–1401.
- [26] R. Hartley and R. Vidal, "Perspective nonrigid shape and motion recovery," in *Proc. Eur. Conf. Comput. Vis.*, Marseille, France, Oct. 2008, pp. 276–289.
- [27] Y. Zhu, D. Huang, F. De La Torre, and S. Lucey, "Complex non-rigid motion 3D reconstruction by union of subspaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 1542–1549.
- [28] C. Bregler, A. Hertzmann, and H. Biermann, "Recovering non-rigid 3D shape from image streams," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Hilton Head Island, SC, USA, Jun. 2000, pp. 690–696.
- [29] J. Xiao, J.-X. Chai, and T. Kanade, "A closed-form solution to non-rigid shape and motion recovery," in *Proc. Eur. Conf. Comput. Vis.*, Prague, Czech Republic, May 2004, pp. 573–587.
- [30] R. Vidal and D. Abretské, "Nonrigid shape and motion from multiple perspective views," in *Proc. Eur. Conf. Comput. Vis.*, Graz, Austria, May 2006, pp. 205–218.
- [31] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [32] Y. Ma, S. Soatto, J. Kosecká, and S. S. Sastry, *An Invitation to 3-D Vision*. New York, NY, USA: Springer, 2004.
- [33] L. Yang, Z. Liu, X. Wang, and Y. Xu, "An optimized image-based visual servo control for fixed-wing unmanned aerial vehicle target tracking with fixed camera," *IEEE Access*, vol. 7, pp. 68455–68468, 2019.
- [34] J. Gao, G. Zhang, P. Wu, X. Zhao, T. Wang, and W. Yan, "Model predictive visual servoing of fully-actuated underwater vehicles with a sliding mode disturbance observer," *IEEE Access*, vol. 7, pp. 25516–25526, 2019.
- [35] X. Zhong, X. Zhong, H. Hu, and X. Peng, "Adaptive neuro-filtering based visual servo control of a robotic manipulator," *IEEE Access*, vol. 7, pp. 76891–76901, 2019.
- [36] Z. Qiu, S. Hu, and X. Liang, "Model predictive control for uncalibrated and constrained image-based visual servoing without joint velocity measurements," *IEEE Access*, vol. 7, pp. 73540–73554, 2019.
- [37] C. Yuan and G. Medioni, "3D reconstruction of background and objects moving on ground plane viewed from a moving camera," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, New York, NY, USA, vol. 2, Jun. 2006, pp. 2261–2268.
- [38] M. Fujita, H. Kawai, and M. W. Spong, "Passivity-based dynamic visual feedback control for three-dimensional target tracking: Stability and  $L_2$ -gain performance analysis," *IEEE Trans. Control Syst. Technol.*, vol. 15, no. 1, pp. 40–52, Jan. 2007.
- [39] T. Hatanaka and M. Fujita, "Passivity-based visual motion observer integrating three dimensional target motion models," *SICE J. Control, Meas., Syst. Integr.*, vol. 5, no. 5, pp. 276–282, Sep. 2012.
- [40] G. Hu, D. Aiken, S. Gupta, and W. E. Dixon, "Lyapunov-based range identification for paracatadioptric system," *IEEE Trans. Autom. Control*, vol. 53, no. 7, pp. 1775–1781, Aug. 2008.
- [41] N. Gans, G. Hu, and W. E. Dixon, "Image based state estimation," in *Encyclopedia of Complexity and Systems Science*, vol. 5. New York, NY, USA: Springer, 2009, pp. 4751–4776.
- [42] A. P. Dani, N. R. Fischer, Z. Kan, and W. E. Dixon, "Globally exponentially stable observer for vision-based range estimation," *Mechatronics*, vol. 22, no. 4, pp. 381–389, Jun. 2012.
- [43] A. P. Dani, N. Fischer, and W. E. Dixon, "Single camera structure and motion estimation," *IEEE Trans. Autom. Control*, vol. 57, no. 1, pp. 238–243, Jan. 2012.
- [44] V. K. Chitrakaran, D. M. Dawson, W. E. Dixon, and J. Chen, "Identification of a moving object's velocity with a fixed camera," *Automatica*, vol. 41, no. 3, pp. 553–562, Mar. 2005.
- [45] W. E. Dixon, Y. Fang, D. M. Dawson, and T. J. Flynn, "Range identification for perspective vision systems," *IEEE Trans. Autom. Control*, vol. 48, no. 12, pp. 2232–2238, Dec. 2003.
- [46] A. P. Dani, Z. Kan, N. R. Fischer, and W. E. Dixon, "Structure and motion estimation of a moving object using a moving camera," in *Proc. Amer. Control Conf.*, Baltimore, MD, USA, Jun./Jul. 2010, pp. 6962–6967.
- [47] A. P. Dani, Z. Kan, N. R. Fischer, and W. E. Dixon, "Structure estimation of a moving object using a moving camera: An unknown input observer approach," in *Proc. IEEE Conf. Decis. Control Eur. Control Conf.*, Orlando, FL, USA, Dec. 2011, pp. 5005–5010.
- [48] S. Jang, A. P. Dani, C. D. Crane, III, and W. E. Dixon, "Experimental results for moving object structure estimation using an unknown input observer approach," in *Proc. ASME Dyn. Syst. Control Conf.*, Fort Lauderdale, FL, USA, Oct. 2012, pp. 597–606.
- [49] L. Ma, C. Cao, N. Hovakimyan, C. Woolsey, and W. E. Dixon, "Fast estimation for range identification in the presence of unknown motion parameters," *IMA J. Appl. Math.*, vol. 75, no. 2, pp. 165–189, Feb. 2010.
- [50] D. Chwa, A. P. Dani, and W. E. Dixon, "Range and motion estimation of a monocular camera using static and moving objects," *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 4, pp. 1174–1183, Jul. 2016.
- [51] Q. Gao, L. Liu, G. Feng, and Y. Wang, "Universal fuzzy integral sliding-mode controllers for stochastic nonlinear systems," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2658–2669, Dec. 2014.
- [52] J. Li, Q. Zhang, X.-G. Yan, and S. K. Spurgeon, "Robust stabilization of T-S fuzzy stochastic descriptor systems via integral sliding modes," *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2736–2749, Sep. 2018.
- [53] S. Ijaz, L. Yan, M. T. Hamayun, W. M. Baig, and C. Shi, "An adaptive LPV integral sliding mode FTC of dissimilar redundant actuation system for civil aircraft," *IEEE Access*, vol. 6, pp. 65960–65973, 2018.
- [54] S. Ijaz, M. T. Hamayun, L. Yan, H. Ijaz, and C. Shi, "Adaptive fault tolerant control of dissimilar redundant actuation system of civil aircraft based on integral sliding mode control strategy," *Trans. Inst. Meas. Control*, vol. 41, no. 13, pp. 3756–3768, Sep. 2019.
- [55] C. P. Vo, X. D. To, and K. K. Ahn, "A novel adaptive gain integral terminal sliding mode control scheme of a pneumatic artificial muscle system with time-delay estimation," *IEEE Access*, vol. 7, pp. 141133–141143, 2019.
- [56] M. Munir, Q. Khan, and F. Deng, "Integral sliding mode strategy for robust consensus of networked higher order uncertain non linear systems," *IEEE Access*, vol. 7, pp. 85310–85318, Jun. 2019.
- [57] Q. Wu, Z. Liu, F. Liu, and X. Chen, "LPV-based self-adaption integral sliding mode controller with  $L_2$  gain performance for a morphing aircraft," *IEEE Access*, vol. 7, pp. 81515–81531, Jun. 2019.
- [58] R. M. Asl, Y. S. Hagh, R. Palm, and H. Handroos, "Integral non-singular terminal sliding mode controller for nth-order nonlinear systems," *IEEE Access*, vol. 7, pp. 102792–102802, 2019.
- [59] Y. Y. Wang, F. Yan, K. W. Zhu, B. Chen, and H. T. Wu, "Practical adaptive integral terminal sliding mode control for cable-driven manipulators," *IEEE Access*, vol. 6, pp. 78575–78586, 2018.
- [60] M. A. Qureshi, I. Ahmad, and M. F. Munir, "Double integral sliding mode control of continuous gain four quadrant quasi-Z-source converter," *IEEE Access*, vol. 6, pp. 77785–77795, 2018.
- [61] V. I. Utkin and H.-C. Chang, "Sliding mode control on electro-mechanical systems," *Math. Problems Eng.*, vol. 8, nos. 4–5, pp. 451–473, Jan. 2002.
- [62] D. Chwa, "Sliding-mode-control-based robust finite-time antiway tracking control of 3-D overhead cranes," *IEEE Trans. Ind. Electron.*, vol. 64, no. 8, pp. 6775–6784, Aug. 2017.
- [63] S. Baek, J. Baek, and S. Han, "An adaptive sliding mode control with effective switching gain tuning near the sliding surface," *IEEE Access*, vol. 7, pp. 15563–15572, 2019.
- [64] M. S. Park, D. Chwa, and M. Eom, "Adaptive sliding-mode antiway control of uncertain overhead cranes with high-speed hoisting motion," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 5, pp. 1262–1271, Oct. 2014.
- [65] I. Elleuch, A. Khedher, and K. Ben Othmen, "Proportional integral sliding mode observer for uncertain Takagi Sugeno systems with unknown inputs," in *Proc. 7th Int. Conf. Modelling, Identificat. Control*, Sousse, Tunisia, Dec. 2015, pp. 1–5.
- [66] H.-C. Chen, C.-I. Wu, C.-W. Chang, Y.-H. Chang, and H.-W. Lin, "Integral sliding-mode flux observer for sensorless vector-controlled induction motors," in *Proc. Int. Conf. Syst. Sci. Eng.*, Taipei, Taiwan, Jul. 2010, pp. 298–303.
- [67] Y.-S. Lu and C.-W. Chiu, "Experimental assessment of an uncertainty-bounds estimator for an integral sliding disturbance observer," in *Proc. IEEE Int. Conf. Mechatronics Robot.*, Takamatsu, Japan, Aug. 2013, pp. 1010–1014.
- [68] P. V. Surjagade, S. R. Shimjith, and A. P. Tiwari, "Second order integral sliding mode observer and controller for a nuclear reactor," *Nucl. Eng. Technol.*, to be published, doi: 10.1016/j.net.2019.08.013.
- [69] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Gaussian process dynamical models for human motion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 283–298, Feb. 2008.

- [70] A. Ravichandran, R. Chaudhry, and R. Vidal, "Categorizing dynamic textures using a bag of dynamical systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 342–352, Feb. 2013.
- [71] B. Solmaz, B. E. Moore, and M. Shah, "Identifying behaviors in crowd scenes using stability analysis for dynamical systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 2064–2070, Oct. 2012.
- [72] O. Zoeter and T. Heskes, "Hierarchical visualization of time-series data using switching linear dynamical systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1202–1214, Oct. 2003.
- [73] R. Li, T.-P. Tian, and S. Sclaroff, "Divide, conquer and coordinate: Globally coordinated switching linear dynamical system," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 654–669, Apr. 2012.
- [74] V. Venkataraman and P. Turaga, "Shape distributions of nonlinear dynamical systems for video-based inference," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 12, pp. 2531–2543, Dec. 2016.
- [75] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1996.
- [76] S. Hutchinson, G. D. Hager, and P. I. Corke, "A tutorial on visual servo control," *IEEE Trans. Robot. Autom.*, vol. 12, no. 5, pp. 651–670, Oct. 1996.



**DONGKYOUNG CHWA** received the B.S. and M.S. degrees in control and instrumentation engineering and the Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 1995, 1997, and 2001, respectively.

From 2001 to 2003, he was a Postdoctoral Researcher with Seoul National University, where he was also a BK21 Assistant Professor, in 2004.

He was a Visiting Scholar with the University of New South Wales, Australian Defence Force Academy; the University of Melbourne, Melbourne, VIC, Australia, in 2003; and the University of Florida, Gainesville, in 2011. Since 2005, he has been with the Department of Electrical and Computer Engineering, Ajou University, Suwon, South Korea, where he is currently a Professor. His research interests include nonlinear, robust, and adaptive control theories and their applications to robotics; underactuated systems, including wheeled mobile robots, underactuated ships, cranes, and guidance and control of flight systems.

• • •