

Range and Motion Estimation of a Monocular Camera Using Static and Moving Objects

Dongkyoung Chwa, Ashwin P. Dani, *Member, IEEE*, and Warren E. Dixon, *Fellow, IEEE*

Abstract—We propose a method of estimating the motion of a monocular camera looking at moving objects and their range. Unlike the previous studies where the camera and object motion should be constrained in estimating structure and motion (SaM) of moving objects, the proposed method do not require those constraints even though only a monocular camera is used. By first arranging the SaM dynamics in terms of the measurable states, we design robust nonlinear observers in a sequential way for both static (stationary) and dynamic (moving) objects. Through the combination of these estimates obtained by nonlinear observers, the reconstruction of the 3-D structure of the dynamic objects can be achieved using just 2-D images of a monocular camera. Simulations are performed in the case of changing camera and object velocities, such that the advantages of the proposed method can be clearly demonstrated.

Index Terms—3-D structure, monocular camera, motion and range estimation, moving camera, moving object.

I. INTRODUCTION

3-D STRUCTURE reconstruction of an object, which can be either static (stationary) or dynamic (moving), using a camera is an important task with many applications [1]–[5]. A 3-D structure estimation requires the information of the range (or depth) of the object from the camera, and therefore, there have been many studies to solve the range estimation problem. The most commonly employed method is the stereo vision method, where the images simultaneously captured by two cameras are compared and their difference can generate the range information via triangulation [6]–[9]. The disadvantage of this method is that accurate camera calibration is required and a small error in the calibration can translate to large errors in the structure estimation.

Manuscript received June 25, 2015; revised October 12, 2015; accepted November 27, 2015. Manuscript received in final form December 4, 2015. Date of publication December 30, 2015; date of current version June 9, 2016. This work was supported in part by the Information and Communication Technologies Research and Development Program within the Ministry of Science, ICT and Future Planning (MSIP)/Institute for Information and Communications Technology Promotion through the Project entitled Resilient Cyber-Physical Systems Research under Grant 14-824-09-013 and in part by the National Research Foundation of Korea within MSIP through the Korean Government under Grant 2014R1A2A1A11053153. Recommended by Associate Editor M. Fujita.

D. Chwa is with the Department of Electrical and Computer Engineering, Ajou University, Suwon 443-749, Korea (e-mail: dkchwa@ajou.ac.kr).

A. P. Dani is with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 0626 USA (e-mail: ashwin.dani@engr.uconn.edu).

W. E. Dixon is with the Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: wdixon@ufl.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCST.2015.2508001

Thus, a range estimation has been much studied based on multiple images captured using a monocular camera. Triangulation-based methods have been developed for a monocular camera in [10]–[13], which, however, restrict the object to move along simple trajectories, such as a straight line or a circle, and are based on batch processing of large amount of image data. The batch processing and the assumptions on the moving object trajectories become restrictive for many object tracking applications where the real-time operation is necessary, e.g., tracking trajectories of moving objects for driverless cars and reactive collision avoidance of robots.

The information of the camera motion has been inevitable in the previous studies on structure from motion, motion from structure, and structure and motion (SaM), as defined in [14]. The problem of SaM simultaneously solves the problem of the 3-D structure estimation (camera pose) and the camera motion estimation (camera velocities). The SaM algorithms are useful in many applications, where the camera motion is not known using external motion sensors, e.g., an inertial measurement unit (IMU). Solutions to SaM based on an adaptive estimator design for nonlinear systems are given in [15] and [16]. In [17], an unknown input observer-based method is presented in order to estimate a moving camera-moving object relative structure. The approach is validated in [18], using experiments conducted by a camera on a PUMA robot and the feature points tracked on moving objects. In [19] and [20], algorithms for an image-based pose estimation are presented in the case of a known distance between two feature points on the object. In [21] and [22], robust observer algorithms that minimize L_2 gain of the estimation error to the moving object velocity are presented in the passivity-based framework of nonlinear systems.

In [10], a batch algorithm for the recovery of the moving object structure using a moving camera is applied to points moving in straight lines and conic trajectories based on five and nine views, respectively. The batch algorithm in [10] is extended to the case of more general object motions represented using curves in [11]. In [12], a matrix factorization-based algorithm is proposed for objects moving with constant speed along a straight line. In [13], a batch algorithm is developed to estimate the SaM of objects on the condition that one of the feature points of the moving object lies on the static background. In [23], the structure estimation algorithm requires an approximation of the trajectories of a moving object in terms of a linear combination of discrete cosine transform basis vectors. In [24], an optimization framework for the image-based pose estimation is presented that optimally

fuses multiple rotation matrices and translation vectors. In [25] and [26], a visual odometry (VO) algorithm is presented for ground robots to estimate the camera motion based on video inputs. The difference between the methods presented in this paper and VO is that we are interested in estimating camera pose in the case of a moving object, whereas VO is interested in estimating the camera trajectory via the video input of the stereo camera pair. In [27], an extension of VO to the omnidirectional camera is presented.

In this paper, estimators are developed that take 2-D camera images as input and estimate the camera velocities based on stationary objects. The camera velocities are then used to estimate the structure of the moving object. In particular, the motion of the camera and the object is considered to be fairly general. In order to estimate both velocities of the camera and range of the moving object, it is necessary to employ the dynamics of the partially measurable state as in [15] and [28]. However, the proposed method involves the recasting of these dynamics into those of the measurable states in such a way that a sequential design and application of the robust integral signed error (RISE)-based robust nonlinear observers can be achieved as the following two steps. First, up-to-a-scale estimates of the camera velocities are obtained by tracking the stationary objects in images via an RISE-based nonlinear observer [15], [29]. This implies that the unknown constant scale factor should be further estimated to estimate the exact camera velocities. Next, the unknown scale factor is estimated by designing another nonlinear observer that uses the velocities of the dynamic objects and estimates the unknown scale of camera velocities. Based on these results, we can further estimate the range between the camera and the object. In [30], a recursive least square (RLS) algorithm [31] is designed for the estimation of velocities of a static object, and the robustness issue, such as the noise in the measurement of image pixel and the velocities, is not considered. In this paper, an RISE-based robust nonlinear observer is designed to provide better performance, compared with the algorithm presented in [30].

Since the advantages of the proposed method lies in the fact that both objects and camera can move in all directions with nonzero velocities, simulations have been performed under those conditions to clearly verify the proposed method. Simulation results show that through the employment of nonlinear observers, which are known to be robust against even fast time-varying uncertainties, for both static and dynamic objects, the proposed method even in the case of relatively fast time-varying camera velocities can estimate well both camera velocities and range although the information of these is not included in the 2-D camera images.

The remainder of this paper is organized as follows. The measurement model and the camera-object relative motion model are presented for the development of the proposed methods in Sections II and III, respectively. In Sections IV and V, the nonlinear observers for the estimation of the camera motion and range estimation are presented for both static and dynamic objects, and the stability of the resulting overall system is analyzed. Simulation results for the evaluation of the proposed control law are

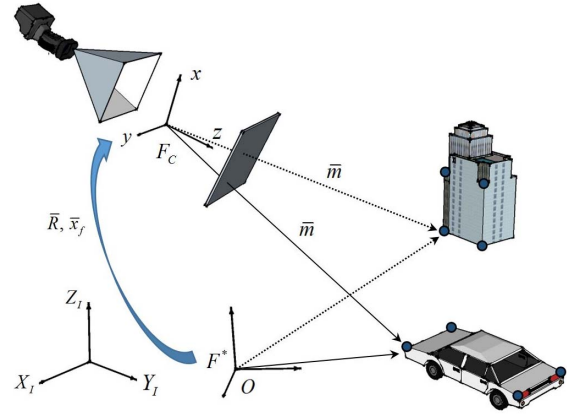


Fig. 1. Moving camera looking at static and dynamic objects.

provided in Section VI. Finally, the conclusion is drawn in Section VII.

II. MEASUREMENT MODEL

The point correspondences between consecutive images of a monocular camera needs to be described through the measurement model in both Euclidean and image spaces, which can be computed using the feature tracking techniques [32]–[34]. In the case of a moving camera, the feature points of both static and dynamic objects appear to be moving in the camera reference frame; thus, the relationship between the measurement of Euclidean space and that of image space needs to be described.

To this end, as shown in Fig. 1, we need to introduce the orthogonal coordinate systems attached to a monocular camera denoted by F^* and F_C , each of which is, respectively, placed at the initial location with the initial time t_0 and the moved location via a rotation matrix $\bar{R}(t) \in SO(3)$ and a translational vector $\bar{x}_f(t) \in \mathbb{R}^{3 \times 1}$ from the initial location. Then, the Euclidean coordinates of a feature point observed by a moving camera should be introduced as $\bar{m}(t) := [x_1(t), x_2(t), x_3(t)]^T \in \mathbb{R}^{3 \times 1}$ in the camera frame F_C . Instead of using the vector $\bar{m}(t)$ with three unknown components, it is beneficial to use $m(t)$ with two unknown components obtained by normalizing the Euclidean coordinates as $m(t) := [x_1(t)/x_3(t), x_2(t)/x_3(t), 1]^T$ in F_C . Still, for easier development of the proposed method, a state vector $y(t) := [y_1(t), y_2(t), y_3(t)]^T$ is defined instead of $m(t)$ as

$$y = \begin{bmatrix} \frac{x_1}{x_3}, \frac{x_2}{x_3}, \frac{1}{x_3} \end{bmatrix}^T. \quad (1)$$

The rationale behind this choice of $y(t)$ can be explained in the following way. First, the first two components of $y(t)$ are set to be the same as those of $m(t)$, since they can be obtained from the pixel coordinates $p(t) := [u, v, 1]^T$ in the image space through the relationship

$$p = A_c m \quad (2)$$

for a known constant invertible camera calibration matrix $A_c \in \mathbb{R}^{3 \times 3}$ [14]. Second, the last component of $y(t)$ is an inverse of unmeasurable state $x_3(t)$, which is the range

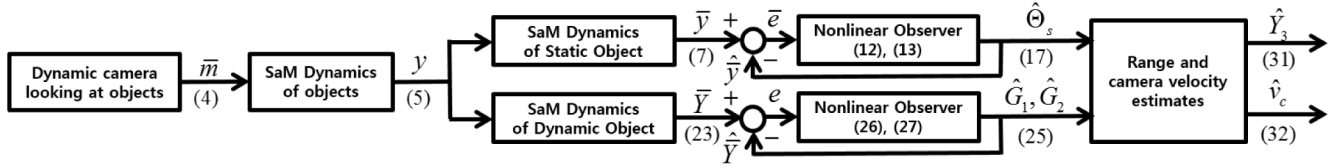


Fig. 2. Structure of the proposed range and motion estimation method of a monocular camera using static and moving objects.

between the camera and the feature point, and thus, the estimation of $y(t)$ can yield the range estimation.

Since the dynamic object moving away from the camera is hard to be estimated in practice, we are not considering that situation. Therefore, we need to introduce the following assumption.

Assumption 1: The vectors $\bar{m}(t)$, $m(t)$, and $y(t)$ are bounded.

The boundedness of $\bar{m}(t)$ implies the boundedness of $m(t)$ and $p(t)$. From the relationship between p and A_c , the boundedness of the first two components of $y(t)$ can be related to the finite size of the images, which is a natural assumption in the application of the proposed method.

III. STRUCTURE AND MOTION DYNAMICS OF OBJECTS

Fig. 1 describes the moving camera looking at the feature point q in the object. Since the camera has moved from the initial location, the Euclidean coordinates of q can be represented in F_C as $\bar{m}(t) = \bar{x}_f + \bar{R}x_{Oq}$ for a vector x_{Oq} , which starts from the origin of the coordinate system F^* to q . Its time derivative results in the following relative motion dynamics between the camera and object in the camera coordinate system [14], [35]:

$$\dot{\bar{m}}(t) = [\omega]_X \bar{m} + v_r. \quad (3)$$

Here, $\omega(t) := [\omega_1(t), \omega_2(t), \omega_3(t)]^T$ is a camera angular velocity in F_C , $[\omega]_X \in \mathbb{R}^{3 \times 3}$ is a skew-symmetric matrix defined as

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

and $v_r(t)$ is a camera linear velocity relative to q , such as $v_r = v_c - v_p$, where $v_c := [v_{cx}, v_{cy}, v_{cz}]^T$ is a camera velocity, $v_p := [v_{px}, v_{py}, v_{pz}]^T$ is a velocity of q in F_C obtained as $v_p(t) = \bar{R}\bar{v}_p$, \bar{R} is a rotational matrix between the camera frame F_C and inertial reference frame, and \bar{v}_p is a velocity of q in the inertial reference frame. For the development of the proposed method provided later in Theorems 1 and 2, we also introduce the following assumptions.

Assumption 2: The linear camera velocity vector $v_c(t)$ and the object velocity $v_p(t)$ are bounded and continuously differentiable.

Assumption 3: The velocities $v_p(t)$, $\bar{v}_p(t)$, and $\omega(t)$ can be measured.

Remark 1: The boundedness of the camera velocities and also the continuous differentiability of the linear velocities in Assumption 2 can be made without loss of generality considering the physical properties of actual camera motion.

Assumption 3 can be introduced in the following sense. If the velocity of the object in the inertial reference frame is controlled and known as in Assumption 3, then it can be obtained in the camera reference frame by using the rotational matrix between the camera and inertial reference frames, which can be obtained by the integration of angular velocities. Next, angular velocities can be assumed to be available, since they can be measured using IMU or estimated using the homography matrix decomposition between consecutive camera frames [15], [29].

The measurement model of the feature point with respect to a moving camera can be derived by substituting $\omega(t)$ and $v_r(t)$ into (3) as follows:

$$\dot{\bar{m}}(t) = \begin{bmatrix} 1 & 0 & 0 & 0 & x_3 & -x_2 \\ 0 & 1 & 0 & -x_3 & 0 & x_1 \\ 0 & 0 & 1 & x_2 & -x_1 & 0 \end{bmatrix} \begin{bmatrix} v_r \\ \omega \end{bmatrix}. \quad (4)$$

By combining (1) and (4), the SaM dynamics of the objects given by

$$\begin{cases} \dot{y}_1 = (v_{cx} - y_1 v_{cz})y_3 - y_1 y_2 \omega_1 + (1 + y_1^2)\omega_2 - y_2 \omega_3 \\ \quad - (v_{px} - y_1 v_{pz})y_3 \\ \dot{y}_2 = (v_{cy} - y_2 v_{cz})y_3 - (1 + y_2^2)\omega_1 + y_1 y_2 \omega_2 + y_1 \omega_3 \\ \quad - (v_{py} - y_2 v_{pz})y_3 \\ \dot{y}_3 = -y_3^2 v_{cz} - y_2 y_3 \omega_1 + y_1 y_3 \omega_2 + y_3^2 v_{pz} \end{cases} \quad (5)$$

can be derived in terms of $y_1(t)$, $y_2(t)$, which are measurable as in (2), and $y_3(t)$, which is dependent on the unmeasurable $x_3(t)$ and, thus, should be estimated. From the fact that $x_3(t)$ is naturally lower bounded by the camera focal length λ_m , $y_3(t)$ can be assumed to be both upper and lower bounded [36]. Since these SaM dynamics contain the dynamic equation of an unmeasurable state, these should be further arranged as the SaM dynamics in terms of only measurable states for the design of robust nonlinear observers as provided in Section IV.

The overall structure of the proposed method is shown in Fig. 2. The proposed method for the camera velocity and range estimation will be achieved as the following two steps. First, using a fact that the value of the depth or the inverse depth up-to-a-scale factor can be obtained using a monocular camera, the camera velocity is estimated up-to-a-scale factor based on the static object in Section IV. Next, the exact camera velocity information is estimated based on the dynamic object by estimating the unknown scale factor, and then, the range is also estimated in Section V.

Remark 2: Our approach takes multiple images and uses the information from feature point tracking in a sequential manner of a filtering framework, as in MonoSLAM [37]. In addition, if we know the linear velocity of the object and the angular velocity of the camera, Chwa *et al.* [30] showed that we can

obtain the inverse depth of a feature point with the scaling information.

Remark 3: In the application of the proposed method, we have considered the situation when there exists an obvious static object, such as the building, house, and so on. When any static object cannot be seen in the camera image, it is hard to apply the proposed method. Thus, the information of what the static object is may be needed beforehand for more practical application. In addition, a moving object can include any vehicle, such as unmanned guided vehicle or unmanned aerial vehicle.

Remark 4: A monocular camera is assumed to be able to see both static and dynamic objects simultaneously. The proposed method can be effectively employed in the case where 2-D images taken by a monocular camera can be reconstructed as 3-D images offline, such that the estimates of two nonlinear observers for static and dynamic objects can be combined in a sequential manner. Although we are not combining the estimates of both observers at the same time, Fig. 4(a) shows that the camera velocity estimation from the static object is sufficiently fast. Therefore, the overall performance of the estimation using the combined observers can be expected to become satisfactory.

Remark 5: Once the unknown variable Θ_s defined right after (9) has been estimated as in (17) during the time when the static object is visible, the disappearance of the static object due to the limited field of view (FOV) of the monocular camera does not matter. In the cases that the static object disappears in such a short time such that estimation has not been completed yet, we need to assume that there are several other static objects, which are known beforehand, such that the information of the static objects can be used in the proposed method even with the limited FOV of the camera.

IV. ESTIMATION OF CAMERA VELOCITY FOR STATIC OBJECT

In this section, SaM dynamics in (5) is changed as those appropriate for the camera velocity estimation based on a static object image. Then, a robust nonlinear observer based on an RISE method is proposed to obtain an up-to-a-scale camera velocity estimate containing an unknown constant scale factor.

A. SaM Dynamic Model for Static Object

Since $y_3(t)$ is unmeasurable, the SaM model in (5) is still not appropriate for the camera velocity and range estimation. In the case of a static object, $y(t)$ corresponds to the static feature points, and in particular, the exact value of y_3 is not available using the vision sensor. It should be noted that the numerous previous estimation methods cannot readily be applied to the SaM model in (5) due to the fact that the number of two measurable outputs is smaller than that of three unknown camera velocities. On the other hand, its value up-to-the-scale factor instead of its exact value can be obtained in the form of

$$y_3 = d_s \bar{y}_3 \quad (6)$$

for a nonzero \bar{y}_3 with an unknown constant d_s . Therefore, we need to introduce the vector consisting of measurable variables, such as $\bar{y} = [y_1, y_2, \bar{y}_3]^T$, in such a way that the dynamic equations in (5) can be arranged in a matrix-vector form as follows:

$$\dot{\bar{y}} = \Omega_{s1} d_s v_c + \Omega_{s2} + \Omega_{s3} \quad (7)$$

where

$$\begin{aligned} \Omega_{s1} &= \begin{pmatrix} \bar{y}_3 & 0 & -y_1 \bar{y}_3 \\ 0 & \bar{y}_3 & -y_2 \bar{y}_3 \\ 0 & 0 & -\bar{y}_3^2 \end{pmatrix} \\ \Omega_{s2} &= \begin{pmatrix} -y_1 y_2 & 1 + y_1^2 & -y_2 \\ -(1 + y_2^2) & y_1 y_2 & y_1 \\ -y_2 \bar{y}_3 & y_1 \bar{y}_3 & 0 \end{pmatrix} \omega \\ \Omega_{s3} &= \begin{pmatrix} -\bar{y}_3 & 0 & y_1 \bar{y}_3 \\ 0 & -\bar{y}_3 & y_2 \bar{y}_3 \\ 0 & 0 & \bar{y}_3^2 \end{pmatrix} d_s v_p. \end{aligned}$$

The idea of introducing \bar{y}_3 and d_s is motivated by the fact that the SaM dynamics in (7) can be separated by the known and unknown parts and the unknown part can be linear in the unknown parameter $d_s v_c$, which leads to the easier estimation of the camera velocity. Here, Ω_{s2} (i.e., ω) can be obtained and Ω_{s3} becomes zero due to $v_p = 0$ in the case of a static object.

B. Recursive Least Squares Algorithm for Camera Velocity Estimation

A brief description of the RLS algorithm in [30] is provided for the completeness of this paper. Since the RLS algorithm is based on an algebraic equation where the unknown parameter is present to be linear with respect to known functions, it cannot be directly applied to the model in (7). Therefore, it becomes necessary to derive the algebraic equation from the dynamic equation in (7).

By introducing the differential operator $\bar{p} = d/dt$ and the stable filter $H_f(\bar{p}) := 1/(\bar{p} + h)$, (7) can be expressed as

$$\frac{\bar{p}}{\bar{p} + h} \bar{y} = \Omega_{s1f} d_s v_c + \Omega_{s2f} \quad (8)$$

where $\Omega_{s1f} = \Omega_{s1} H_f(p)$ and $\Omega_{s2f} = \Omega_{s2} H_f(p)$. For the estimation of unknown parameters using RLS algorithm, (8) is arranged in the standard linear regression form as

$$y_s = \Phi_s^T \Theta_s \quad (9)$$

where $y_s = (p/(p + h))\bar{y} - \Omega_{s2f}$, $\Theta_s := d_s v_c$ is an uncertain parameter vector to be estimated, and $\Phi_s := \Omega_{s1f}^T$ is a regression variable. Thus, the following RLS algorithm [31] can be used to obtain $\hat{\Theta}_s$, which is the estimate of Θ_s :

$$\begin{cases} \frac{d\hat{\Theta}_s(t)}{dt} = P_s(t) \Phi_s(t) E_s(t) \\ E_s(t) = y_s(t) - \Phi_s^T(t) \hat{\Theta}_s \\ \frac{dP_s(t)}{dt} = \alpha P_s(t) - P_s(t) \Phi_s(t) \Phi_s^T(t) P_s(t) \end{cases} \quad (10)$$

where $E_s(t)$ is a residual variable and $P_s(t)$ is a time-varying matrix. However, an accurate estimate of v_c cannot be obtained just from $\hat{\Theta}_s$; this requires further development of the SaM

dynamic model for a dynamic object and the robust nonlinear observer based on that model, as provided in Section IV-C. In particular, the RLS algorithm has the following several limitations. First, since Θ_s is dependent on v_c , its application can be impractical in the case of the camera motion with time-varying velocities. Although the exponential forgetting factor [38] can be introduced to the RLS algorithm in (10) in this situation, the estimation errors can be guaranteed to be ultimately bounded [39]. Thus, more practical estimation algorithm should be developed even for more general time-varying velocities. Second, even when the camera moves with constant or slowly time-varying velocities, whether the parameter estimation error converges toward zero or not is dependent on the properties of Φ_s (i.e., the persistency of excitation condition [40]). This can lead to the significant degradation of estimation performance. Finally, its robustness against the noise and uncertainties is not guaranteed, and thus, much more robust algorithm should be further developed by considering the actual properties of the camera images. As shown in Section VI, the performance of the RLS algorithm degrades considerably in the presence of pixel noise although it is satisfactory when there is no noise in the image pixel measurement.

C. Nonlinear Observer for Camera Velocity Estimation

As commented in Section IV-B, the camera velocity estimation algorithm, which is robust against noise, should be developed for more practical application.

Unlike the RLS algorithm described in Section IV-B, a robust nonlinear observer can estimate the time-varying velocities well. Here, the nonlinear observer to be designed in this section is based on the RISE term, which suppresses the influence of the disturbance in the measurements, such as image pixel noise, and also the uncertain parameters, such as unknown moving object velocities, thereby guaranteeing the robustness against them [41].

To design a robust nonlinear observer, we first need to define an unmeasurable vector $\bar{g} := \Omega_{s1}\Theta_s$, such that (7) can be rewritten as

$$\dot{\bar{y}} = \bar{g} + \Omega_{s2} + \Omega_{s3}. \quad (11)$$

As in the RLS algorithm, the information of \bar{g} can be used to obtain Θ_s (i.e., $d_s v_c$) from $\Omega_{s1}^{-1}\bar{g}$. The existence of Ω_{s1}^{-1} is reasonable in the sense that \bar{y}_3 is nonzero in practice. For the further development of the nonlinear observer, we introduce the practical assumption as in [42] that $\|\bar{g}(\cdot)\| \leq \bar{\zeta}_1$, $\|\dot{\bar{g}}(\cdot)\| \leq \bar{\zeta}_2$, and $\|\ddot{\bar{g}}(\cdot)\| \leq \bar{\zeta}_3$ for known positive constants $\bar{\zeta}_1$, $\bar{\zeta}_2$, and $\bar{\zeta}_3$.

Since the estimator for \bar{g} should be developed, $\bar{e} := [\bar{e}_1, \bar{e}_2, \bar{e}_3]^T = \bar{y} - \hat{y}$ is introduced considering that \bar{y} is measurable and \hat{y} is its estimate available from the observer to be designed later in (12).

Remark 6: Although the information of the rotation and translation of the camera (i.e., \bar{R} and \bar{x}_f) is used in deriving (5) from (3), only the availability of the variables y_1 and y_2 (not \bar{R} and \bar{x}_f) is necessary, which is possible from (2), and the availability of the pixel coordinates in the image space. Therefore, the error vector \bar{e} used in the observer is available.

In addition, a filtered estimation error $\bar{r} = \dot{\bar{e}} + \bar{\alpha}\bar{e}$ is introduced for a diagonal constant matrix $\bar{\alpha} \in \mathbb{R}^{3 \times 3}$. Then, a nonlinear observer can be designed for (11) as

$$\dot{\hat{y}} = \hat{g} + \Omega_{s2}. \quad (12)$$

Here, $\hat{g} := [\hat{g}_1, \hat{g}_2, \hat{g}_3]^T \in \mathbb{R}^{3 \times 1}$, which is an estimate of \bar{g} , comes from [42], [43]

$$\dot{\hat{g}} = -(\bar{k}_s + \bar{\alpha})\hat{g} + \bar{\gamma} \text{sgn}(\bar{e}) + \bar{\alpha}\bar{k}_s\bar{e} \quad (13)$$

where $\bar{k}_s := \text{diag}(\bar{k}_{s1}, \bar{k}_{s2}, \bar{k}_{s3})$ and $\bar{\gamma} := \text{diag}(\bar{\gamma}_1, \bar{\gamma}_2, \bar{\gamma}_3) \in \mathbb{R}^{3 \times 3}$ are constant positive definite diagonal matrices and $\text{sgn}(\bar{e})$ is a standard signum function applied to each element of the argument. Since $\Omega_{s3} = 0$ in the case of static object, the error dynamics between (11) and (12) can be easily obtained as

$$\dot{\bar{e}} = \bar{g} - \hat{g}. \quad (14)$$

Thus, for $\bar{\eta} := [\bar{\eta}_1, \bar{\eta}_2, \bar{\eta}_3]^T = \dot{\bar{g}} + (\bar{k}_s + \bar{\alpha})\bar{g} \in \mathbb{R}^{3 \times 1}$, the time derivative of the filtered tracking error can be expressed as

$$\begin{aligned} \dot{\bar{r}} &= \dot{\bar{g}} - \dot{\hat{g}} + \bar{\alpha}\dot{\bar{e}} \\ &= \dot{\bar{g}} + (\bar{k}_s + \bar{\alpha})\hat{g} - \{\bar{\gamma} \text{sgn}(\bar{e}) + \bar{\alpha}\bar{k}_s\bar{e}\} + \bar{\alpha}\dot{\bar{e}} \\ &= \bar{\eta} + (\bar{k}_s + \bar{\alpha})(-\bar{e}) - \{\bar{\gamma} \text{sgn}(\bar{e}) + \bar{\alpha}\bar{k}_s\bar{e}\} + \bar{\alpha}\dot{\bar{e}} \\ &= \bar{\eta} - \bar{k}_s\bar{r} - \bar{\gamma} \text{sgn}(\bar{e}) \end{aligned} \quad (15)$$

where the second and third equalities come from (13) and (14), respectively. From the aforementioned boundedness of \bar{g} and $\dot{\bar{g}}$, that of $\bar{\eta}(\cdot)$ and $\dot{\bar{\eta}}(\cdot)$ follows in the form of $\|\bar{\eta}(\cdot)\| \leq \bar{\zeta}_4$ and $\|\dot{\bar{\eta}}(\cdot)\| \leq \bar{\zeta}_5$ for positive constants $\bar{\zeta}_4$ and $\bar{\zeta}_5$. Then, the properties of the robust nonlinear observer in (12) and (13) can be described as the following theorem.

Theorem 1: Suppose that Assumptions 1–3 are satisfied. When each element of the diagonal constant matrix $\bar{\gamma}$ in (13) satisfies

$$\bar{\gamma}_i \geq \bar{\zeta}_4 + \frac{1}{\bar{\alpha}_i}\bar{\zeta}_5 \quad (16)$$

for $i = 1, 2, 3$, \hat{g} asymptotically converges to \bar{g} , in such a way that the estimate of the unknown variable Θ_s (i.e., $d_s v_c$) in SaM dynamics (5) or (7) can be obtained using the nonlinear observer in (12) and (13) as

$$\hat{\Theta}_s := \Omega_{s1}^{-1}\hat{g}. \quad (17)$$

Proof: The time derivative of the Lyapunov function candidate

$$V = \bar{r}^T \bar{r} / 2 \quad (18)$$

can be expressed using (15) as

$$\begin{aligned} \dot{V} &= \bar{r}^T (\bar{\eta} - \bar{k}_s\bar{r} - \bar{\gamma} \text{sgn}(\bar{e})) \\ &= -\bar{r}^T \bar{k}_s\bar{r} + (\dot{\bar{e}} + \bar{\alpha}\bar{e})^T \cdot (\bar{\eta} - \bar{\gamma} \text{sgn}(\bar{e})). \end{aligned} \quad (19)$$

If we introduce $\bar{\gamma}_v = [\bar{\gamma}_1, \bar{\gamma}_2, \bar{\gamma}_3]^T \in \mathbb{R}^{3 \times 1}$, each element of which is the diagonal component of the matrix $\bar{\gamma}$, and $|\cdot|$ which

is an absolute value applied to each element of the argument vector, then the integration of (19) gives

$$\begin{aligned}
V(t) &= V(t_0) - \int_{t_0}^t \bar{r}(\tau)^T \bar{k}_s \bar{r}(\tau) d\tau \\
&\quad + \int_{t_0}^t \dot{\bar{e}}(\tau)^T \cdot \{\bar{\eta}(\tau) - \bar{\gamma} \operatorname{sgn}(\bar{e}(\tau))\} d\tau \\
&\quad + \int_{t_0}^t (\bar{\alpha} \bar{e})^T \cdot \{\bar{\eta}(\tau) - \bar{\gamma} \operatorname{sgn}(\bar{e}(\tau))\} d\tau \\
&\leq V(t_0) - \int_{t_0}^t \bar{r}(\tau)^T \bar{k}_s \bar{r}(\tau) d\tau \\
&\quad + \bar{e}(\tau)^T \{\bar{\eta}(\tau) - \bar{\gamma} \operatorname{sgn}(\bar{e}(\tau))\} \Big|_{t_0}^t \\
&\quad - \int_{t_0}^t \bar{e}(\tau)^T \cdot \{\dot{\bar{\eta}}(\tau) - 2\bar{\gamma} \delta(\bar{e}(\tau)) \dot{\bar{e}}(\tau)\} d\tau \\
&\quad + \int_{t_0}^t |\bar{\alpha} \bar{e}|^T \cdot \{|\bar{\eta}(\tau)| - \bar{\gamma}_v\} d\tau \\
&= V(t_0) - \int_{t_0}^t \bar{r}(\tau)^T \bar{k}_s \bar{r}(\tau) d\tau \\
&\quad + \bar{e}(t)^T \bar{\eta}(t) - \bar{\gamma}_v^T |\bar{e}(t)| - \bar{e}(t_0)^T \bar{\eta}(t_0) + \bar{\gamma}_v^T |\bar{e}(t_0)| \\
&\quad - \int_{t_0}^t \bar{e}(\tau)^T \cdot \{\dot{\bar{\eta}}(\tau) - 2\bar{\gamma} \delta(\bar{e}(\tau)) \dot{\bar{e}}(\tau)\} d\tau \\
&\quad + \int_{t_0}^t |\bar{\alpha} \bar{e}|^T \cdot \{|\bar{\eta}(\tau)| - \bar{\gamma}_v\} d\tau \quad (20a)
\end{aligned}$$

where $\delta(\bar{e}(\tau))$ is a delta function satisfying $\delta(\bar{e}(\tau)) = (\partial/2\partial\bar{e}(\tau))\operatorname{sgn}(\bar{e}(\tau))$. Since the right-hand side of (19) contains the discontinuous function, its absolutely continuous Filippov solution $V(t)$ in (20a) exists and satisfies the following inequality almost everywhere [44]:

$$\begin{aligned}
V(t) &\leq V(t_0) - \int_{t_0}^t \bar{r}(\tau)^T \bar{k}_s \bar{r}(\tau) d\tau \\
&\quad + |\bar{e}(t)|^T \cdot (|\bar{\eta}(t)| - \bar{\gamma}_v) - \bar{e}(t_0)^T \bar{\eta}(t_0) + \bar{\gamma}_v^T |\bar{e}(t_0)| \\
&\quad + \int_{t_0}^t |\bar{\alpha} \bar{e}|^T \cdot \{|\bar{\eta}(\tau)| + |\bar{\alpha}^{-1} \dot{\bar{\eta}}(\tau)| - \bar{\gamma}_v\} d\tau. \quad (20b)
\end{aligned}$$

Under Filippov's framework, a generalized Lyapunov stability theory can be used (see [45], [46] for further details) to establish strong stability of the closed-loop system. Since (16) implies the nonpositivity of each element of the third and sixth terms in the right-hand side of (20b) (i.e., $|\bar{\eta}(t)| - \bar{\gamma}_v$ and $|\bar{\eta}(\tau)| + |\bar{\alpha}^{-1} \dot{\bar{\eta}}(\tau)| - \bar{\gamma}_v$), (20b) can be arranged as

$$V(t) \leq V(t_0) - \int_{t_0}^t \bar{r}(\tau)^T \bar{k}_s \bar{r}(\tau) d\tau + \sigma \quad (21)$$

where $\sigma = -\bar{e}(t_0)^T \bar{\eta}(t_0) + \bar{\gamma}_v^T |\bar{e}(t_0)|$ is constant. From (18) and (21), $\bar{r} \in L_2 \cap L_\infty$, i.e., $\bar{e}, \dot{\bar{e}} \in L_\infty$ (respectively, $\hat{\bar{y}}, \dot{\hat{\bar{y}}} \in L_\infty$) from the definition of \bar{r} (respectively, \bar{e} and Assumption 1). Also, from (12) and (13), $\hat{\bar{g}}, \dot{\hat{\bar{g}}} \in L_\infty$. In particular, from the definition of $\bar{\eta}$ and the boundedness of $\bar{g}(\cdot)$, $\dot{\bar{g}}(\cdot)$, and $\ddot{\bar{g}}(\cdot)$, it follows from (15) that $\dot{\bar{r}} \in L_\infty$. Thus, Barbalat's lemma [47] can be used to show that $\bar{r} \rightarrow 0$ as $t \rightarrow \infty$, which leads to the asymptotic convergence of \bar{e} (respectively, $\dot{\bar{e}}$) to zero, such that $\hat{\bar{y}}$ (respectively, $\hat{\bar{g}}$) converges to \bar{y} (respectively, \bar{g}) as time goes to ∞ from

the definition of \bar{e} [respectively, (14)]. Also, due to the nonsingularity of Ω_{s1} , Θ_s can be estimated from $\hat{\bar{g}}$. ■

Although Θ_s can be estimated from Theorem 1, the estimation of the camera velocity requires the estimate of d_s to be known, which will be presented in Section V.

V. ESTIMATION OF CAMERA VELOCITY FOR DYNAMIC OBJECT AND RANGE

In this section, the information of a dynamic object, instead of a static object, is used to estimate the unknown constant scale factor, which, in turn, can be used to estimate the exact camera velocity.

In the case of a dynamic object as well, the state variable for the SaM model is introduced as $Y = [Y_1, Y_2, Y_3]^T$, which is defined in the same way as y in (1) and is denoted differently just for the differentiation of the static and dynamic state vectors, along with the velocity of the dynamic object defined as $v_o = [v_{ox}, v_{oy}, v_{oz}]^T$ instead of v_p in (5). Then, the SaM model for a dynamic object can be described based on (5) as

$$\dot{Y} = \Phi_d^T \Psi_d \quad (22)$$

where $\Psi_d := [v_c, \omega, v_o]^T$ and

$$\begin{aligned}
\Phi_d^T &= \begin{pmatrix} Y_3 & 0 & -Y_1 Y_3 & -Y_1 Y_2 & 1 + Y_1^2 & -Y_2 & -Y_3 & 0 & Y_1 Y_3 \\ 0 & Y_3 & -Y_2 Y_3 & -(1 + Y_2^2) & Y_1 Y_2 & Y_1 & 0 & -Y_3 & Y_2 Y_3 \\ 0 & 0 & -Y_3^2 & -Y_2 Y_3 & Y_1 Y_3 & 0 & 0 & 0 & Y_3^2 \end{pmatrix}.
\end{aligned}$$

In the case of a dynamic object, not only the range information Y_3 is unmeasurable, but also v_o is a nonzero velocity vector of a dynamic object.

Since the nonlinear observer will be designed to estimate Y_3 , Y_3 does not need to be constant or slowly time varying. Due to the asymptotic convergence of the estimation error to zero in Theorem 1, v_c can be replaced by $\hat{\Theta}_s/d_s$. Then, the number of the unknown variables in (22) reduces two, which is the number of d_s and Y_3 . Therefore, we just need to use only the first and second equations of (22). To this end, a measurable vector $\tilde{Y} := [Y_1, Y_2]^T$ is included instead of Y . That is, instead of the third-order SaM model in (7), we need to adopt the second-order SaM model given by

$$\dot{\tilde{Y}} = \Omega_{d1}(Y_3/d_s) + \Omega_{d2} + \Omega_{d3} Y_3 \quad (23)$$

where

$$\begin{aligned}
\Omega_{d1} &= \begin{pmatrix} 1 & 0 & -Y_1 \\ 0 & 1 & -Y_2 \end{pmatrix} \hat{\Theta}_s \\
\Omega_{d2} &= \begin{pmatrix} -Y_1 Y_2 & 1 + Y_1^2 & -Y_2 \\ -(1 + Y_2^2) & Y_1 Y_2 & Y_1 \end{pmatrix} \omega \\
\Omega_{d3} &= \begin{pmatrix} 1 & 0 & -Y_1 \\ 0 & 1 & -Y_2 \end{pmatrix} (-v_o).
\end{aligned}$$

Since the right-hand side of (23) is linear in the unknown parameters Y_3/d_s and Y_3 , the estimation of its unknown terms can yield the estimates of unknown parameters. Thus, (23) is rewritten as

$$\dot{\tilde{Y}} = \Omega_1 + g \quad (24)$$

where Ω_1 (respectively, g) is a measurable (respectively, unmeasurable) vector satisfying $\Omega_1 = \Omega_{d2}$ (respectively, $g = [\Omega_{d1}, \Omega_{d3}] \cdot [Y_3/d_s, Y_3]^T$). When $\hat{g} := [\hat{g}_1, \hat{g}_2]^T \in \mathbb{R}^{2 \times 1}$, which is an estimate of g , can be obtained by the nonlinear observer to be designed later in (27), the estimates of $[Y_3/d_s, Y_3]^T$ can be obtained by defining $[G_1, G_2]^T := [\Omega_{d1}, \Omega_{d3}]^{-1} \cdot g = [Y_3/d_s, Y_3]^T$ as

$$[\hat{G}_1, \hat{G}_2]^T = [\Omega_{d1}, \Omega_{d3}]^{-1} \cdot \hat{g} \quad (25)$$

provided that $[\Omega_{d1}, \Omega_{d3}]^{-1}$ exists. When we compare Ω_{d1} and Ω_{d3} , the nonsingularity of the matrix $[\hat{\Theta}_s, v_o]$ (i.e., the noncollinearity of the camera and object motion) is a sufficient condition for the existence of (25). As long as both camera and object move in at least one noncollinear direction, this condition can be satisfied; this makes sense when we consider the practical situation. As in Section IV, the assumption that $\|g(\cdot)\| \leq \zeta_1$, $\|\dot{g}(\cdot)\| \leq \zeta_2$, and $\|\ddot{g}(\cdot)\| \leq \zeta_3$ for positive constants ζ_1 , ζ_2 , and ζ_3 is made again.

Based on the SaM model in (24), the estimator for g should be designed by defining $e := [e_1, e_2]^T = \bar{Y} - \hat{Y}$. Here, \bar{Y} is measurable and \hat{Y} comes from the observer to be designed later in (26). Also, $r = \dot{e} + \alpha e$ is defined for a diagonal constant matrix $\alpha \in \mathbb{R}^{2 \times 2}$. Then, a nonlinear observer can be designed for (24) as

$$\dot{\hat{Y}} = \Omega_1 + \hat{g} \quad (26)$$

where \hat{g} can be obtained as in (13) as

$$\dot{\hat{g}} = -(k_s + \alpha)\hat{g} + \gamma \text{sgn}(e) + \alpha k_s e \quad (27)$$

where $k_s, \gamma \in \mathbb{R}^{2 \times 2}$ are positive definite constant diagonal matrices. Unlike the derivation of the error dynamics in (14), the nonzero object velocity is no hindrance in deriving the error dynamics

$$\dot{e} = g - \hat{g} \quad (28)$$

and

$$\dot{r} = \eta - k_s r - \gamma \text{sgn}(e) \quad (29)$$

where $\eta(t) := [\eta_1, \eta_2]^T = \dot{g} + (k_s + \alpha)g \in \mathbb{R}^{2 \times 1}$. Since both g and \dot{g} are bounded, it is easy to see that $\|\eta(\cdot)\| \leq \zeta_4$ and $\|\dot{\eta}(\cdot)\| \leq \zeta_5$ hold for positive constants ζ_4 and ζ_5 . Then, using the nonlinear observer in (26) and (27) for dynamic object, we can estimate both camera velocity and range as in the following theorem.

Theorem 2: When each element of the diagonal constant matrix γ in (13) satisfies

$$\gamma_i \geq \zeta_4 + \frac{1}{\alpha_i} \zeta_5 \quad (30)$$

for $i = 1, 2, 3$, \hat{g} asymptotically converges to g , in such a way that the estimates of the unknown variables d_s and Y_3 (i.e., \hat{d}_s and \hat{Y}_3) in the SaM dynamics (23) can be obtained using the nonlinear observer in (26) and (27) as

$$\hat{d}_s = \hat{G}_2 / \hat{G}_1, \quad \hat{Y}_3 = \hat{G}_2. \quad (31)$$

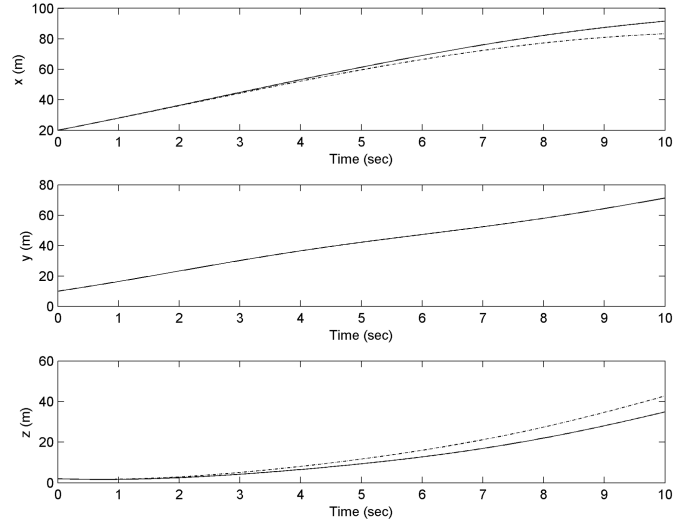


Fig. 3. Time histories of the static and dynamic objects position in the single-camera images. Solid line: static. Dashed-dotted line: dynamic.

In addition, the estimate of the camera velocity v_c (i.e., \hat{v}_c) can be estimated as

$$\hat{v}_c = \hat{\Theta}_s / \hat{d}_s \quad (32)$$

where $\hat{\Theta}_s$ can be obtained as (17) by using the nonlinear observers in (12) and (13).

Proof: Most of the proof follows from that of Theorem 1 and the remaining part is trivial. ■

VI. SIMULATION RESULTS

Simulations are performed to demonstrate the performance of the proposed method for camera velocity and range estimation. As far as we know, there have been no previous results except the authors' previous conference [30], where the object and the camera can move in a free way without any restriction. Thus, just the proposed method and the previous method in [30] will be compared, in such a way that the aforementioned limitation of the application of the RLS algorithm to the considered problem will be clarified. Whereas the usual speed of the camera is 30 frames/s, the nonlinear observers are valid for a continuous-time system. Thus, their performance of the camera velocity and range estimation should be evaluated under the following practical conditions. First, the camera and object velocities are assumed to be available at the sampling rate of 30 ms from the camera images, and the solutions of the nonlinear observers were calculated using the Runge–Kutta algorithm with the sampling rate of 1 ms.

Design parameters of nonlinear observers are chosen as $\bar{k}_s = k_s = \text{diag}\{50, 50\}$, $\bar{\alpha} = \alpha = \text{diag}\{50, 50\}$, and $\bar{\gamma} = \gamma = \text{diag}\{10^{-5}, 10^{-5}\}$, and the camera calibration matrix A_c in (2) is set to be

$$A_c = \begin{bmatrix} 720 & 0 & 320 \\ 0 & 720 & 240 \\ 0 & 0 & 1 \end{bmatrix}.$$

An initial target feature point position is set to be $\bar{m}(t_0) = [20, 10, 1]^T$ (m), and the camera and object velocities are

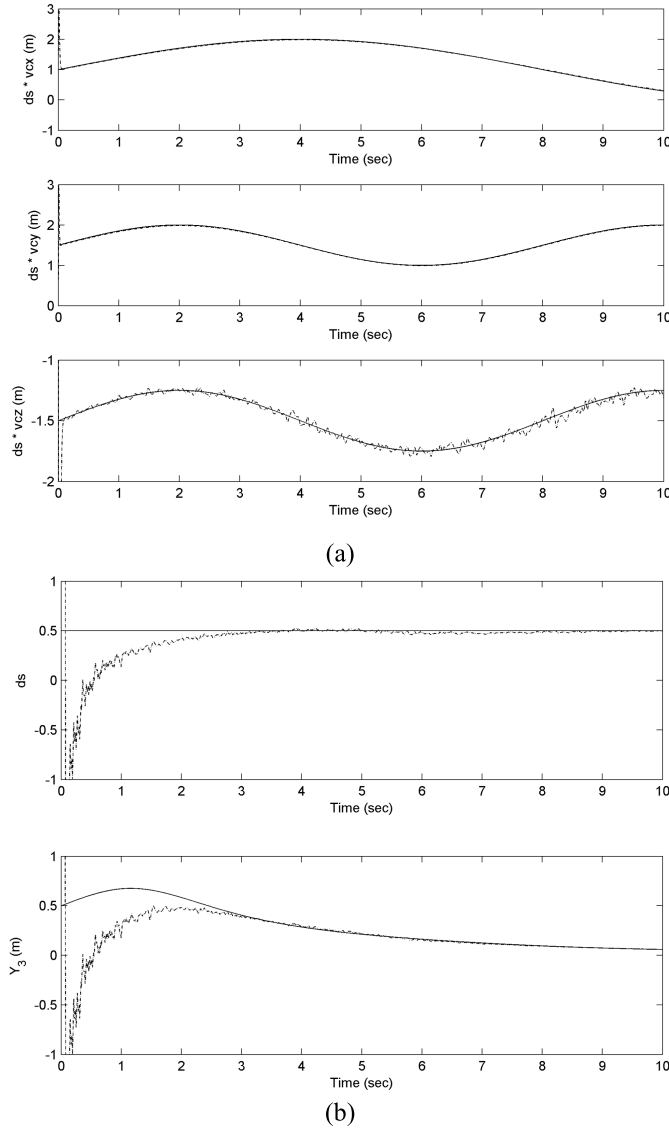


Fig. 4. Performance of the combination of the RLS algorithm for static object and nonlinear observer for dynamic object in the single camera images with pixel noise. Solid line: actual. Dashed-dotted line: estimated. (a) Estimation performance of Θ_s for static object. (b) Estimation performance of d_s and Y_3 for dynamic object.

chosen, respectively, as

$$v_c = \begin{bmatrix} 2 + 2 \sin(\pi t/8) \\ 3 + \sin(\pi t/4) \\ -3 + 0.5 \sin(\pi t/4) \end{bmatrix} \text{ (m/s)}$$

$\omega = [0, -\pi/30, 0]^T$ (rad/s), and $\bar{v}_0 = [-2, -1, -1]^T$ (m/s), which gives $v_o(t)$ as $v_o(t)\bar{R}\bar{v}_o$ and \bar{R} can be obtained by using the angular velocity. Depending on the choice of \bar{y}_3 with respect to y_3 , the true value of d_s in (6) becomes an appropriate positive constant. Here, d_s is chosen as $d_s = 0.5$. Since \bar{y} and \bar{Y} are known, we can set the initial values of $\hat{\bar{y}}$ in (12) and $\hat{\bar{Y}}$ in (26) to be their true values $\bar{y}(0)$ and $\bar{Y}(0)$. In addition, the initial estimates of \hat{g} in (13) and \hat{g} in (27) are selected as

$$\begin{aligned} \hat{g}(0) &= [\hat{g}_1(0), \hat{g}_2(0), \hat{g}_3(0)]^T = [1, 1, 1]^T \\ \hat{g}(0) &= [\hat{g}_1(0), \hat{g}_2(0)]^T = [1, 1]^T. \end{aligned}$$

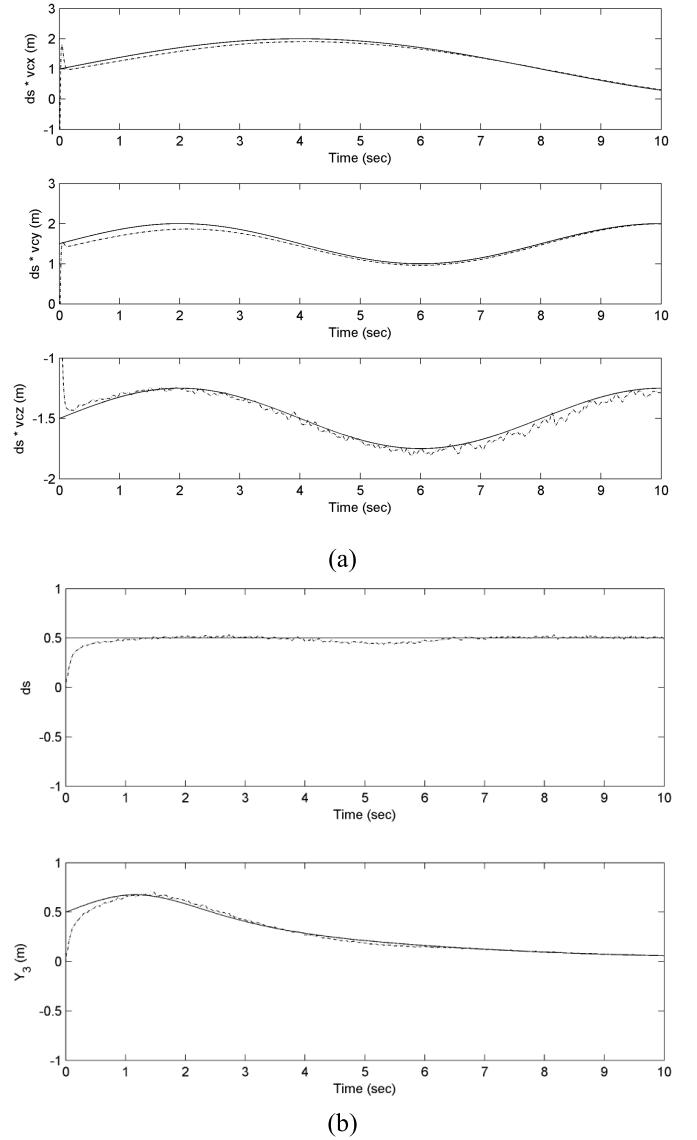


Fig. 5. Performance of the proposed nonlinear observer using both static and dynamic objects in the single camera images with pixel noise. Solid line: actual. Dashed-dotted line: estimated. (a) Estimation performance of Θ_s for static object. (b) Estimation performance of d_s and Y_3 for dynamic object.

The proposed method is validated for robustness by the addition of white Gaussian noise with a signal-to-noise ratio of 100 dB to the image pixel measurements, and 5% noise with zero mean and a variance of 0.1 to the measured velocities. We found that the estimation of the static object using the RLS algorithm was very sensitive to noise in the considered problem, such that considerable degradation of performance could be observed depending on the size of the noise.

Under the above conditions, the positions of static and dynamic objects captured in the single camera images in the inertial reference frame are shown in Fig. 3. The performance of the RLS algorithm and the proposed method can be compared, as shown in Figs. 4 and 5.

First, the RLS algorithm in [30] is employed to the SaM model for the static object in (9) to produce $\hat{\Theta}_s$ as in Section IV-B. In the case that there is no pixel noise, the estimation performance of the RLS algorithm is satisfactory,

as shown in [30]. However, the presence of the pixel noise can significantly degrade its performance. Therefore, it can be seen that the practical situation does require a more robust observer for the considered problem. In particular, the initial transient estimates of Θ_s , d_s , and Y_3 can make the estimation performance much degraded in the sense that the estimate of v_c that can be obtained from the division of Θ_s by d_s can be much worse in the present case.

Next, a proposed robust nonlinear observer is employed to the SaM model for the static object in (11) to produce $\hat{\Theta}_s$. Fig. 5(a) shows the satisfactory estimation performance of an up-to-a-scale camera velocity even in the presence of pixel noise. By using $\hat{\Theta}_s$ to obtain (23), the estimates of the nonlinear observer can be processed to obtain \hat{d}_s . Fig. 5(b) shows the asymptotic convergence of \hat{d}_s and \hat{Y}_3 to their true value. In addition, by combining the estimate $\hat{\Theta}_s$ from the static object and \hat{d}_s from the dynamic object, the camera velocity v_c can be estimated well. In particular, the performance in the estimation of these variables in Fig. 5(b) is much improved, compared with that of Fig. 4(b). These results demonstrate that the proposed nonlinear observer-based camera motion and the range estimation method can achieve robust and satisfactory performance even with the object motion and noise.

VII. CONCLUSION

The proposed method for a camera velocity and range estimation uses just the images of a monocular camera, and does not restrict the camera and object motion, unlike the previous studies. In the case of previous studies, they are based on the form of an algebraic equation, and the motion and structure estimations are solved by the least-square solution or an optimization method, which inevitably leads to weakness in the presence of object motion and noise. To solve these problems, we have derived the SaM dynamic equations in terms of measurable states, such that the robust nonlinear observers can be applied to the camera images of static and dynamic objects. Unlike the RLS method, which can show degraded performance in the presence of pixel noise, the proposed RISE-based nonlinear observer has been previously verified to work well against even the fast time-varying uncertainties. Since the camera velocity and range can show the changing characteristics in a practical situation, the performance of both camera velocity and range estimation becomes satisfactory using the proposed method in this situation. It is noticeable that our work can estimate the camera motion and range based on the feature points of the moving object, eliminating the restriction on the object or camera motion. Although simulations are performed to demonstrate the validity of the proposed method, an experimental study needs to be pursued as a further work by considering the following issues. The feature tracks may often be broken during the processing of the input frames, due to occlusion, blur, image noise, object motion leaving the camera's FOV, and so on, which may have to be considered in the actual work. In addition, the proposed method can be extended to more practical applications where both quadrotor with camera and the vehicle are in motion and only limited information of either camera or object velocities is available.

REFERENCES

- [1] A. Mora, D. F. Glas, T. Kanda, and N. Hagita, "A teleoperation approach for mobile social robots incorporating automatic gaze control and three-dimensional spatial visualization," *IEEE Trans. Syst., Man, Cybern.*, vol. 43, no. 3, pp. 630–642, May 2013.
- [2] D. Raviv and M. Herman, "A unified approach to camera fixation and vision-based road following," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 8, pp. 1125–1141, Aug. 1994.
- [3] P. Jiang and R. Unbehauen, "Robot visual servoing with iterative learning control," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 32, no. 2, pp. 281–287, Mar. 2002.
- [4] F. Dornaika and B. Raducanu, "Three-dimensional face pose detection and tracking using monocular videos: Tool and application," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 4, pp. 935–944, Aug. 2009.
- [5] J. F. Ferreira, J. Lobo, P. Bessiere, M. Castelo-Branco, and J. Dias, "A Bayesian framework for active artificial perception," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 699–711, Apr. 2013.
- [6] 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.
- [7] W. E. L. Grimson, "Computational experiments with a feature based stereo algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 7, no. 1, pp. 17–33, Jan. 1985.
- [8] U. R. Dhond and J. K. Aggarwal, "Structure from stereo—A review," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 6, pp. 1489–1510, Nov./Dec. 1989.
- [9] S. T. Barnard and M. A. Fischler, "Computational stereo," *ACM Comput. Surv.*, vol. 14, no. 4, pp. 553–572, Dec. 1982.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.*, vol. 2, New York, NY, USA, Jun. 2006, pp. 2261–2268.
- [14] Y. Ma, S. Soatto, J. Kořecká, and S. S. Sastry, *An Invitation to 3-D Vision*. Springer, 2004.
- [15] A. P. Dani, N. R. Fischer, and W. E. Dixon, "Single camera structure and motion," *IEEE Trans. Autom. Control*, vol. 57, no. 1, pp. 241–246, Jan. 2012.
- [16] 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.
- [17] 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. 50th IEEE Conf. Decision Control Eur. Control Conf.*, Orlando, FL, USA, Dec. 2011, pp. 5005–5010.
- [18] 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 5th Annu. Dyn. Syst. Control Conf.*, Fort Lauderdale, FL, USA, Oct. 2012, pp. 597–606.
- [19] N. Gans, G. Hu, and W. E. Dixon, "Image based state estimation," in *Encyclopedia of Complexity and Systems Science*, vol. 5, Springer, 2009, pp. 4751–4776.
- [20] A. P. Dani, S. Velat, C. Crane, N. R. Gans, and W. E. Dixon, "Experimental results for image-based pose and velocity estimation," in *Proc. IEEE Int. Conf. Control Appl.*, San Antonio, TX, USA, Sep. 2008, pp. 1159–1164.
- [21] 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.
- [22] 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.
- [23] 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.*, Greece, Europe, Sep. 2010, pp. 158–171.

- [24] S. S. Mehta, P. Barooah, W. E. Dixon, E. L. Pasiliao, and J. W. Curtis, "PEGUS: An image-based robust pose estimation method," in *Proc. 9th Conf. Comput. Robot Vis.*, Toronto, ON, Canada, May 2012, pp. 78–85.
- [25] D. Nistér, N. Oleg, and J. Bergen, "Visual odometry for ground vehicle applications," *J. Field Robot.*, vol. 23, no. 1, pp. 3–20, Jan. 2006.
- [26] K. Konolige, M. Agrawal, and J. Solà, "Large-scale visual odometry for rough terrain," in *Robotics Research*, vol. 66. Springer, 2011, pp. 201–212.
- [27] J.-P. Tardif, Y. Pavlidis, and K. Daniilidis, "Monocular visual odometry in urban environments using an omnidirectional camera," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nice, France, Sep. 2008, pp. 2531–2538.
- [28] 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.
- [29] 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.
- [30] D. Chwa, A. Dani, H. Kim, and W. E. Dixon, "Camera motion estimation for 3-D structure reconstruction of moving objects," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Seoul, Korea, Oct. 2012, pp. 1788–1793.
- [31] K. J. Åström and B. Wittenmark, *Adaptive Control*, 2nd ed. Reading, MA, USA: Addison-Wesley, 1995.
- [32] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Artif. Intell.*, San Francisco, CA, USA, Aug. 1981, pp. 674–679.
- [33] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [34] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. 9th Eur. Conf. Comput. Vis.*, vol. 3951. May 2006, pp. 404–417.
- [35] 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.
- [36] A. De Luca, G. Oriolo, and P. R. Giordano, "Feature depth observation for image-based visual servoing: Theory and experiments," *Int. J. Robot. Res.*, vol. 27, no. 10, pp. 1093–1116, Oct. 2008.
- [37] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, Jun. 2007.
- [38] S. Dasgupta and Y.-F. Huang, "Asymptotically convergent modified recursive least-squares with data-dependent updating and forgetting factor for systems with bounded noise," *IEEE Trans. Inf. Theory*, vol. 33, no. 3, pp. 383–392, May 1987.
- [39] J. D. Bošković, L. Chen, and R. K. Mehra, "Adaptive tracking control of a class of non-affine plants using dynamic feedback," in *Proc. Amer. Control Conf.*, Arlington, VA, USA, Jun. 2001, pp. 2450–2455.
- [40] K. S. Narendra and A. M. Annaswamy, *Stable Adaptive Systems*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.
- [41] P. M. Patre, W. MacKunis, C. Makkar, and W. E. Dixon, "Asymptotic tracking for systems with structured and unstructured uncertainties," *IEEE Trans. Control Syst. Technol.*, vol. 16, no. 2, pp. 373–379, Mar. 2008.
- [42] G. Hu, D. Aiken, S. Gupta, and W. E. Dixon, "Lyapunov-based range identification for paracatadioptric systems," *IEEE Trans. Autom. Control*, vol. 53, no. 7, pp. 1775–1781, Aug. 2008.
- [43] 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.
- [44] A. F. Filippov, *Differential Equations With Discontinuous Righthand Sides*. Norwell, MA, USA: Kluwer, 1988.
- [45] D. Shevitz and B. Paden, "Lyapunov stability theory of nonsmooth systems," *IEEE Trans. Autom. Control*, vol. 39, no. 9, pp. 1910–1914, Sep. 1994.
- [46] F. H. Clarke, *Optimization and Nonsmooth Analysis*. Philadelphia, PA, USA: SIAM, 1990.
- [47] J. J. E. Slotine and W. Li, *Applied Nonlinear Control*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1991.



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, Korea, in 1995, 1997, and 2001, respectively.

He was a Post-Doctoral Researcher with Seoul National University from 2001 to 2003, where he was also a BK21 Assistant Professor in 2004. He was a Visiting Scholar with the Australian Defence Force Academy, University of New South Wales, Sydney, NSW, Australia, and the University of Melbourne, Melbourne, VIC, Australia, in 2003, and the University of Florida, Gainesville, FL, USA, in 2011. Since 2005, he has been with the Department of Electrical and Computer Engineering, Ajou University, Suwon, Korea, where he is currently a Professor. His current 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.



Ashwin P. Dani (M'11) received the M.S. and Ph.D. degrees from the University of Florida, Gainesville, FL, USA, in 2008 and 2011, respectively.

He was a Post-Doctoral Research Associate with the University of Illinois at Urbana-Champaign, Champaign, IL, USA, from 2011 to 2013. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT, USA. He has co-authored over 30 refereed papers and two book chapters, and holds two patents in single-camera ranging of stationary and moving objects. His current research interests include nonlinear estimation and control, stochastic nonlinear estimation, vision-based estimation and control, autonomous navigation, and human-robot interaction.



Warren E. Dixon (F'15) received the Ph.D. degree from the Department of Electrical and Computer Engineering, Clemson University, Clemson, SC, USA, in 2000.

He was selected as a Eugene P. Wigner Fellow with the Oak Ridge National Laboratory (ORNL), Oak Ridge, TN, USA. In 2004, he joined the Mechanical and Aerospace Engineering Department, University of Florida (UF), Gainesville, FL, USA. He has authored three books, an edited collection, 12 chapters, and over 100 journal and 200 conference papers. His current research interests include the development and application of Lyapunov-based control techniques for uncertain nonlinear systems.

Prof. Dixon serves as a member of the U.S. Air Force Science Advisory Board and the Director of Operations for the Executive Committee of the IEEE Control Systems Society (CSS) Board of Governors. His work was recognized by the American Automatic Control Council O. Hugo Schuck (Best Paper) Award in 2009 and 2015, the Fred Eilersick Award for Best Overall MILCOM Paper in 2013, the UF College of Engineering Doctoral Dissertation Mentoring Award from 2012 to 2013, the American Society of Mechanical Engineers Dynamics Systems and Control Division Outstanding Young Investigator Award in 2011, the IEEE Robotics and Automation Society Early Academic Career Award in 2006, an NSF CAREER Award from 2006 to 2011, the Department of Energy Outstanding Mentor Award in 2004, and the ORNL Early Career Award for Engineering Achievement in 2001. He is an IEEE CSS Distinguished Lecturer.