Full- and Reduced-Order Observers for Image-Based Depth Estimation Using Concurrent Learning

Ghananeel Rotithor, *Graduate Student Member, IEEE*, Daniel Trombetta, Rushikesh Kamalapurkar, *Senior Member, IEEE*, and Ashwin P. Dani, *Senior Member, IEEE*

***Abstract*— In this brief, concurrent learning (CL)-based full- and reduced-order observers for a perspective dynamical system (PDS) are developed. The PDS is a widely used model for estimating the depth of a feature point from a sequence of camera images. Building on the current progress of CL for parameter estimation in adaptive control, state observers are developed for the PDS model where the inverse depth appears as a time-varying parameter in the dynamics. The data recorded over a sliding time window in the near past are used in the CL term to design the full- and reduced-order state observers. A Lyapunov-based stability analysis is carried out to prove the uniformly ultimately bounded (UUB) stability of the developed observers. Simulation results are presented to validate the accuracy and convergence of the developed observers in terms of convergence time, root mean square error (RMSE), and mean absolute percentage error (MAPE) metrics. Real-world depth estimation experiments are performed to demonstrate the performance of the observers using the aforementioned metrics on a 7-DoF manipulator with an eye-in-hand configuration.**

***Index Terms*— Concurrent learning (CL), depth estimation, reduced-order observer design, visual servo control.**

I. Introduction

ESTIMATING the 3-D coordinates of feature points using observations from a sequence of camera images is referred to as the Structure from Motion (SfM) problem in computer vision literature. The 3-D coordinates of feature points can be estimated by estimating the depth of the features. The estimated 3-D coordinates of feature points or structure information can be used in a variety of automatic control, autonomy, and intelligent control applications. Existing solutions to this problem include off-line [1] and online [2]–[16] methods. The focus of this brief is on online methods where the problem is formulated as a state estimation problem of a perspective dynamical system (PDS). The PDS is a class of nonlinear systems that use inverse depth parameterization, which is widely used in observer-based methods, and simultaneous localization and mapping (SLAM) [17].

Online methods often rely on the use of an extended Kalman filter (EKF) [5], [6]. In comparison to EKF-based approaches, nonlinear observers are developed for SfM with analytical proofs of stability. Under the assumption that the camera motion is known, continuous and discontinuous observers are developed to estimate the depth, which can then be used to estimate the range to the object. A high-gain observer called the identifier-based observer is presented for range estimation in [7]. A semiglobally asymptotically stable reduced-order observer is presented in [8] to estimate the range based on immersion and invariance (I&I) methodology, which is extended to the range and orientation identification observer design in [18]. A continuous observer, which guarantees asymptotic range estimation, is presented in [16] under the assumption that camera motion is known. In [9], an asymptotically converging nonlinear observer is developed based on Lyapunov's indirect method. In [10], a discontinuous sliding-mode observer is developed, which guarantees exponential convergence of the estimation error. In [11], a nonlinear observer is developed that achieves local exponential convergence of estimation error. A range observer design based on nonlinear contraction and synchronization theory is presented in [19]. In [4], a globally exponentially stable observer is designed for the PDS. Extensions of these observers for PDS with moving objects are presented in [20] and [21]. All these observers require persistence of excitation (PE) condition to be satisfied by the camera motion to achieve the convergence of the estimation error.

Drawing parallels to the adaptive control design, in the PDS, the inverse depth appears as a parameter where the parameter is time-varying with known dynamics associated with it. Concurrent learning (CL) is used in adaptive control for parameter estimation, where the knowledge of past trajectory memory data is leveraged to estimate the constant parameter and achieve state tracking [22], [23]. CL has also been used for target tracking applications in [24] and for target size estimation in [25]. The use of CL relaxes the PE condition to a finite excitation condition, which depends upon the rank of the regressor matrix [22]. Recently, in [26], adaptive controllers with memory-augmentation are also designed.

Full- and reduced-order depth observers are designed in this brief that builds on the work in [4] by adding CL-terms. The observer design guarantees the boundedness of the depth estimation error even when the PE condition is not satisfied by the camera motion in a time window given that finite

Manuscript received August 13, 2020; accepted November 1, 2020. Date of publication November 25, 2020; date of current version October 8, 2021. Manuscript received in final form November 3, 2020. This work was supported in part by the Space Technology Research Institute under Grant 80NSSC19K1076 from the NASA Space Technology Research Grants Program, in part by the Office of Naval Research under Award N00014-20-1-2040, and in part by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Advanced Manufacturing Office Award DE-EE0007613. The work of Rushikesh Kamalapurkar was supported by Air Force Research Lab (AFRL) under Grant FA8651-19-2-0009. Recommended by Associate Editor G. Hu. *(Corresponding author: Ashwin P. Dani.)*

Ghananeel Rotithor, Daniel Trombetta, and Ashwin P. Dani are with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269 USA (e-mail: ghananeel.rotithor@uconn.edu; daniel.trombetta@uconn.edu; ashwin.dani@uconn.edu).

Rushikesh Kamalapurkar is with the Mechanical and Aerospace Engineering Department, Oklahoma State University, Stillwater, OK 74078 USA (e-mail: rushikesh.kamalapurkar@okstate.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCST.2020.3036369.

Digital Object Identifier 10.1109/TCST.2020.3036369

excitation is present. The observer can be used to estimate the feature point depth to the desired accuracy. Two cases are analyzed for the convergence and stability of the observer design. The first case is when the camera motions satisfy the PE condition, and the second case is when the camera motions do not satisfy the PE condition. A Lyapunov stability analysis is carried out for the switched observer error system (i.e., when PE is satisfied and when it is not) using multiple Lyapunov functions [27]. Although, compared with the existing depth/range observers in the literature, the observers in this brief cannot achieve asymptotic or exponential depth estimation convergence, the observers can achieve finite estimation errors in practical scenarios when the existing observers may not yield finite estimation error. Practical examples include the motion of the camera along the projected ray while grasping an object or when an aerial robot is moving in the direction of camera view during landing and takeoff or robots moving in the direction of view while docking. For these cases, the camera motion will not satisfy the PE condition for a certain time window. Compared with the recent work in [28] and [29], rigorous stability analysis and detailed simulation and experimental evaluations are presented in this brief. A history stack update procedure based on the Lyapunov analysis, which stores the camera motion and feature point data, is presented. The results of a performance evaluation of the CL-based full-order and reduced-order observers with a benchmark observer are presented using real-world experiments conducted on a camera mounted on the hand of a 7-DoF Baxter robot. For brevity, only the stability analysis proofs of the reduced-order observer are shown in the Appendix. The stability analysis proofs of the full-order observer are given in [30].

II. Perspective Camera Motion Model

The motion of a perspective camera capturing a scene results in the change of image plane coordinates of a feature point belonging to a static object. Let $\bar{m}(t) = [X(t)Y(t)Z(t)]^T \in \mathbb{R}^3$ and $m_n(t) = [(X(t)/Z(t))\,(Y(t)/Z(t))\,1] \in \mathbb{R}^3$ be the Euclidean and normalized Euclidean coordinates of a feature point belonging to a static object captured by a moving camera in the camera reference frame with known camera velocities. To estimate the depth, define an auxiliary vector $[x(t)y(t)\chi(t)]^T \in \mathcal{Y}$ such that $\mathcal{Y} \subset \mathbb{R}^3$ is a closed and bounded set where $x(t) = (X(t)/Z(t)), y(t) = (Y(t)/Z(t)), \chi(t) = (1/Z(t))$. Let $s(t) = [x(t)\ y(t)]^T$ be the state associated with an image plane feature point with two components and $\chi(t)$ be the inverse depth of the feature point.

Remark 1: The state variables $x(t)$ and $y(t)$ are image plane coordinates of a feature point whose pixel coordinates are bounded by the resolution of the camera. As a result, the state variables $x(t)$ and $y(t)$ are bounded by known constants $\underline{x} \le x(t) \le \overline{x}$ and $\underline{y} \le y(t) \le \overline{y}$. The Euclidean distance $Z(t)$ between the camera and the feature point can be lower bounded by the focal length of the camera λ measured in meters and is not assumed to be upper bounded. Therefore, the inverse depth $\chi(t)$ can be upper and lower bounded as in [4] and [11] using the known constants $0 < \underline{y_3} < \chi(t) \le (1/\lambda)$.

Assumption 1: The depth of the feature point $Z(t)$ is invertible in the compact set $\mathcal{Y}$.

The feature point dynamics can be written as a function of the linear and angular velocities as

$$\dot{s} = f_m(s,\omega) + \Omega^T(s,v)\chi \tag{1}$$

$$\dot{\chi} = f_u(s,\chi,u) \tag{2}$$

where $v(t) = [v_X(t)\ v_Y(t)\ v_Z(t)]^T \in \mathcal{V}$ and $\omega(t) = [\omega_X(t)\ \omega_Y(t)\ \omega_Z(t)]^T \in \mathcal{W}$ are the linear velocities in (m/s) and the angular velocities in (rad/s) of the camera in the body frame, and $u(t) = [v^T(t)\ \omega^T(t)]^T$. The sets $\mathcal{V}$ and $\mathcal{W}$ are bounded such that $\mathcal{V} \subset \mathbb{R}^3$ and $\mathcal{W} \subset \mathbb{R}^3$. In (1), $f_m(s,\omega) \in \mathbb{R}^2$ and $\Omega(s,v) \in \mathbb{R}^{1\times 2}$ are functions of measurable quantities. The state derivative $\dot{s}(t) \in \mathbb{R}^2$ is not measurable and can only be estimated. The time argument is dropped in the rest of this brief for ease of notation unless stated. Individually, $f_m(s,\omega)$, $\Omega(s,v)$, and $f_u(s,\chi,u)$ are defined as

$$\begin{aligned} f_m(s,\omega) &= \begin{bmatrix} xy & -(1+x^2) & y \\ 1+y^2 & -xy & -x \end{bmatrix}\omega \\ \Omega(s,v) &= \begin{bmatrix} xv_Z - v_x & yv_Z - v_Y \end{bmatrix} \\ f_u(s,\chi,u) &= v_Z\chi^2 + (y\omega_X - x\omega_Y)\chi. \end{aligned} \tag{3}$$

Problem Definition: Given the measurements of feature points in the image plane $s(t)$, the linear and angular velocities of the camera $u(t)$, and the linear acceleration of the camera $\dot{v}(t)$ in the camera reference frame, it is desired to estimate the inverse depth of the feature point $\chi(t)$ using the dynamics in (1) and (2). To this end, full-order and reduced-order depth observers are designed in Sections III and V using CL.

Assumption 2: The camera velocities are bounded, and the linear velocities are $\mathcal{C}^1$ with respect to time.

III. CL-Based Full-Order Observer

The depth estimation schemes in the existing literature require a strong observability condition called the PE condition, and the estimation error converges to zero only if the PE condition is satisfied. The PE condition is satisfied if there exist constants $T_0, \rho \in \mathbb{R}^+$ such that

$$\int_t^{t+T_0} \Omega(s(\tau),v(\tau))\Omega^T(s(\tau),v(\tau))d\tau \ge \rho > 0 \quad \forall t > t_0. \tag{4}$$

CL-based parameter estimation techniques use a history stack of recorded data generated by the dynamical system to make updates to the parameter estimation scheme. CL is based on the premise that, even if the PE condition cannot be guaranteed, input can be exciting over a finite interval of time. For the full-order CL observer, the history stack is a tuple $\mathcal{H}_f = \{(\dot{\hat{s}}_j, s_j, u_j)\}_{j=1}^{M}$ containing the past data points up to the index $M-1$ chosen by the method proposed in Section VII, where M is the index of the data point at the current time instant. Let $\{t_j\}_{j=1}^{M-1}$ denote the corresponding time instances at which the jth entry in the history stack is recorded. Then, by the definition of the history stack, $\dot{\hat{s}}_j := \dot{\hat{s}}(t_j)$ and $s_j := s(t_j), u_j := u(t_j)\ \forall j = 1,\ldots,M-1$.

Assumption 3: The term $\dot{\bar{s}}_j$ is the approximation of $\dot{s}_j$, computed numerically, such that $\|\dot{\bar{s}} - \dot{s}\| < \bar{d}$ and $\bar{d} \in [0, \infty)$ is an unknown constant.

The estimates of s and χ are denoted by $\hat{s}$ and $\hat{\chi}$, respectively, and the state and depth estimation errors as $z = \chi - \hat{\chi}$ and $\xi = s - \hat{s}$. Using the dynamics in (1) and (2), the observer for estimating the state and the depth is designed as follows:

$$\dot{\hat{s}} = f_m(s, \omega) + \Omega^T(s, v)\hat{\chi} + H\xi \tag{5}$$

$$\dot{\hat{\chi}} = f_u(s, \hat{\chi}, u) + \Gamma\Omega(s, v)\xi + K_{CL}\Gamma \sum_{j=1}^{M} \Omega(s_j, v_j) \times (\dot{\bar{s}}_j - f_m(s_j, \omega_j) - \Omega^T(s_j, v_j)\hat{\chi}) \tag{6}$$

where $H \in \mathbb{R}^{2\times 2}$ is positive-definite diagonal gain matrix, and $\Gamma \in \mathbb{R}^+$ and $K_{CL} \in \mathbb{R}^+$ are suitable observer gains. Since $d_j = \dot{\bar{s}}_j - \dot{s}_j$, the approximated state derivative term $\dot{\bar{s}}_j$ is substituted as $\dot{\bar{s}}_j = f_m(s_j, \omega_j) + \Omega^T(s_j, v_j)\chi_j + d_j$ to compute the estimation error dynamics. Using the dynamics in (1) and (2), observer equations in (5) and (6), adding and subtracting $K_{CL}\Gamma \sum_{j=1}^{M} \Omega(s_j, v_j)\Omega^T(s_j, v_j)\chi$, and grouping χ and $\hat{\chi}$, the estimation error dynamics can be written as

$$\begin{aligned}
\dot{\xi} &= -H\xi + \Omega^T(s, v)z \\
\dot{z} &= -\Gamma\Omega(s, v)\xi + g(s, z, u) \\
&\quad -K_{CL}\Gamma\Bigg(\sum_{j=1}^{M} \Omega(s_j, v_j)d_j \\
&\qquad + \sum_{j=1}^{M} \Omega(s_j, v_j)\Omega^T(s_j, v_j)\big(z + \chi_j - \chi\big)\Bigg)
\end{aligned} \tag{7}$$

where $g(s, z, u) = f_u(s, \chi, u) - f_u(s, \hat{\chi}, u)$.

Assumption 4: The history stack contains recent information and the change in depth over a short period of time remains bounded, i.e., $\exists \bar{\chi} \geq 0$ such that $\sup_{t\geq 0} \max_{j\in\{1,\ldots,M-1\}} \|\chi_j - \chi\| \leq \bar{\chi}$, where $\chi_j = \chi(t_j)$ and χ are the past and current true depth values at the time instants t_j and t such that $t > t_j \forall j = 1, \ldots, M-1$ for a suitably chosen value of M.

Remark 2: The main implication of Assumption 4 is that the history stack should be frequently updated to contain information about the current true depth from the past feature point and camera motion data. In addition, the upper bound $\bar{\chi}$ will be smaller if the object is not too close to the camera and if the camera linear velocities are slow.

IV. Stability Analysis for Full-Order Observer

Since the history stack is initialized with zeros, the stability analysis is carried out in two phases: the initial phase when the data are being collected in the history stack and the phase when the history stack is fully populated with informative points. In Theorem 1, it is shown that the estimation error dynamics in (7) is uniformly ultimately bounded (UUB) when the PE condition is satisfied and the history stack is incomplete. In Theorem 2, it is shown that the estimation error dynamics in (7) is UUB when the PE condition is not satisfied and the history stack is complete. To facilitate the analysis, let $\exists \bar{\sigma} > 0$ such that $\sup_{t\geq 0} \max_{j\in\{1,\ldots,M\}} \|\Omega(s_j, v_j)\Omega^T(s_j, v_j)\| \leq \bar{\sigma}$ and $\sigma_1 = \sum_{j=1}^{M-1} \Omega(s_j, v_j)\Omega^T(s_j, v_j)$ such that $\sigma_1 \in \mathbb{R}_{\geq 0}$.

Definition 1: The history stack is defined to be incomplete when the stack is not completely populated with informative points such that $\sigma_1 \geq 0$.

Definition 2: The history stack is defined to be complete when the history stack is completely populated with informative points such that $\sigma_1 > 0$.

Theorem 1: When the history stack is incomplete, the error system in (7) is UUB if Assumption 4 and the PE condition in (4) are satisfied. Furthermore, the ultimate bound on the estimation error is given by $((c_2/c_1))^{1/2}(\gamma_1 K_{CL}\Gamma((M-1)\bar{\sigma}\bar{\chi} + M\bar{d}(\bar{\sigma})^{1/2})/k_2)$, where c_1, c_2, k_2, γ_1 are the positive constants.

Proof: Refer [30, Appendix A]. ■

Theorem 2: When the history stack is complete, the error system in (7) is UUB if Assumption 4 is satisfied, the PE condition in (4) is not satisfied, and the adjustable observer gain is selected according to the sufficient condition: $K_{CL} > (L_g/\sigma_1\Gamma)$. Furthermore, the ultimate bound on the estimation error is given by $((c_4/c_3))^{1/2}(K_{CL}((M-1)\bar{\sigma}\bar{\chi} + M\bar{d}(\bar{\sigma})^{1/2})/(2k_3\alpha_1)^{1/2})$, where c_3, c_4, k_3, and α_1 are the positive constants.

Proof: Refer [30 Appendix B]. ■

V. CL-Based Reduced-Order Observer

For the reduced-order CL observer, the history stack is a tuple $\mathcal{H}_r = \{(s_j, u_j, \dot{v}_j)\}_{j=1}^{M}$ containing the past data points up to the index $M-1$, and M is the index of the data point at the current time instant. The reduced-order depth observer is defined as

$$\hat{\chi}(t) = \kappa(s, \hat{\chi}, u, \dot{v}) + \gamma(s, v) \tag{8}$$

where

$$\begin{aligned}
\dot{\kappa} &= f_u(s, \hat{\chi}, u) + \bar{K}\sum_{j=1}^{M}\Bigg(\theta_j^T\dot{v}_j - \Omega(s_j, v_j)\big(f_m(s_j, \omega_j) \\
&\qquad + \Omega^T(s_j, v_j)\hat{\chi}\big)\Bigg) \\
\gamma &= -\bar{K}\sum_{j=1}^{M}\theta_j^T v_j
\end{aligned} \tag{9}$$

where $\theta_j = [x_j \; y_j \; (-(x_j^2 + y_j^2)/2)]^T$ for $j = 1, \ldots, M$. The initial condition of the observer is selected as $\kappa(t_0) = \kappa_0$, where $\kappa_0 > 0$ is a constant.

VI. Stability Analysis for Reduced-Order Observer

Differentiating (8) and using (9), to obtain the dynamics of $\hat{\chi}(t)$ as

$$\dot{\hat{\chi}} = f_u(s, \hat{\chi}, u) + \bar{K}\sum_{j=1}^{M}\Omega\big(s_j, v_j\big)\big(\dot{s}_j - f_m(s_j, \omega_j) - \Omega^T(s_j, v_j)\hat{\chi}\big). \tag{10}$$

The error dynamics for the reduced-order observer can be derived by using (3) and (10), substituting $\dot{s}_j$ from (1), adding

and subtracting $\bar{K}\sum_{j=1}^{M}\Omega(s_j,v_j)\Omega^T(s_j,v_j)\chi$, and grouping χ and $\hat{\chi}$ as

$$\dot{z}=-\bar{K}\left(\sum_{j=1}^{M}\Omega(s_j,v_j)\Omega^T(s_j,v_j)\big(z+\chi_j-\chi\big)\right)+g(s,z,u) \quad (11)$$

where $g(s,z,u) = f_u(s,\chi,u) - f_u(s,\hat{\chi},u)$. The stability analysis of the reduced-order CL-based observer is carried out in two phases: the initial phase when the data are being collected in the history stack and the phase when the history stack is fully populated with informative points. In Theorem 3, it is shown that the estimation error dynamics in (11) is UUB when the PE condition is satisfied and the history stack is incomplete. In Theorem 4, it is shown that the estimation error dynamics in (11) is UUB when the PE condition is not satisfied and the history stack is complete.

Theorem 3: When the history stack is incomplete, the error system in (11) is UUB if Assumption 4 and the PE condition in (4) are satisfied. Furthermore, the ultimate bound on the estimation error is given by $((c_6/c_5))^{1/2}(\gamma_2\bar{K}\bar{\sigma}(M-1)\bar{\chi}/k_4)$, where c_5, c_6, k_4, and γ_2 are the positive constants.

Proof: Refer Appendix A. ∎

Theorem 4: When the history stack is complete, the error system in (11) is UUB if Assumption 4 is satisfied, the PE condition in (4) is not satisfied, and the observer gain is selected according to the sufficient condition: $\bar{K} > (L_g/\sigma_1)$. Furthermore, the ultimate bound on the estimation error is given by $(\bar{K}\bar{\sigma}(M-1)\bar{\chi}/k_5)$, where k_5 is a positive constant.

Proof: Refer Appendix *B*. ∎

Remark 3: For the full- and reduced-order observers, the estimation error decreases exponentially to an ultimate bound as $t\to\infty$. The ultimate bound on the estimation error can be made arbitrarily small by selecting appropriate gain values H, Γ, and K_{CL} for the full-order observer, or $\bar{K}$ for the reduced-order observer, and the size of the history stack M. The optimal observer gains may be efficiently computed by solving a linear matrix inequality using incremental quadratic constraints, as demonstrated in [31] and [32].

Remark 4: Old data can be replaced with new data in the history stack even after the history stack is full as long as σ_1 is greater than zero. Using the procedure in Section VII, σ_1 always stays positive even after the old points are replaced from the full history stack. Hence, the upper bound on the derivative of the Lyapunov functions for the full- and reduced-order observers holds at any given time after the history stack is full. Thus, the ultimate bound on the switched systems can be derived from the analysis of multiple Lyapunov functions, as demonstrated in [27, Th. 3.1].

Remark 5: The ultimate bound on the estimation error increases linearly with $\bar{\chi}$, as defined in Assumption 4. As a result, if the history stack contains old points with previous values, the ultimate bound on the estimation error will grow linearly with $\bar{\chi}$. Hence, it is essential for the history stack to be updated frequently to avoid the growth of the ultimate bound. Based on the presented analysis, the procedure in Section VII is designed to frequently update the history stack and ensure that σ_1 stays greater than zero.

VII. History Stack Update

From the analysis in Sections IV and VI, a procedure is designed in this section to ensure that the ultimate bound on the error is small and that σ_1 stays positive. Auxiliary stacks $\mathcal{G}_f = \{(\dot{\bar{s}}_j, s_j, u_j)\}_{j=1}^{N}$ for the full-order observer and $\mathcal{G}_r = \{(s_j, u_j, \dot{v}_j)\}_{j=1}^{N}$ for the reduced-order observer, such that $N > M$, are used to select informative points. The auxiliary stack is a dynamic sliding window of the N most recent points. In each iteration, $M-1$ most informative points are selected from the auxiliary stack to replace the $M-1$ points in the history stack. The history stack and auxiliary stack are both initialized with zeros. At each time instance, the auxiliary stack is sorted in descending order based on the value of $\Omega(s_j,v_j)\Omega^T(s_j,v_j)\forall j=1,\ldots,N$ to select the top $M-1$ points. The points in the history stack are replaced only if the chosen $M-1$ points from the auxiliary stack satisfy $\{\sum_{j=1}^{M-1}\Omega\Omega^T\}_{\mathcal{G}} \geq \epsilon$ for a suitably chosen constant $\epsilon > 0$. This ensures that the value of σ_1 does not drop below ϵ, and the upper bound on the derivative of the Lyapunov functions for the full- and reduced-order observer holds at all times after the history stack is full. The choice of ϵ is critical as it maintains a balance between frequently updating the stack (thus, keeping the ultimate bound small) and ensuring that σ_1 remains greater than zero.

VIII. Simulation Results

A. Simulation 1

A simulation is performed using a simulated feature point to verify the performance of the CL full-order observer designed in Section III. An initial point with the Euclidean coordinates $\bar{m}(t_0) = [2.5\ 0.5\ 3]^T$ is selected. In this simulation, the linear velocities are designed as $v = [0.3, 0.2, \cos(\pi t/4) - 0.3]^T$ and $\omega = [0,\ -(\pi/30),\ 0]^T$. Additive white Gaussian noise with a signal-to-noise ratio (SNR) of 40 dB is added to the state, and the velocity measurements are corrupted with a Gaussian distributed measurement noise with zero mean and variance of 0.01. The CL full-order observer gain values used for the simulation are $K_{CL} = 0.15$, $\Gamma = 5$, and $H = \text{diag}\{10, 10\}$. The initial values for the state estimate and inverse depth estimate are selected as $\hat{s}(t_0) = [10,\ 5]^T$ and $\hat{\chi}(t_0) = 3$. The history stack and the auxiliary stack are initialized with three points and five points, respectively. The CL full-order observer converges to the true depth in 4.7 s compared with the observer in [2], which converges in 5.3 s, as shown in Fig. 1(a). The yellow dashed lines in Fig. 1(a) show the noisy performance of the least-squares (LS) depth estimation based on the formula $\hat{\chi}_{LS} = (\Omega^T(s,v))^{\dagger}(\dot{\bar{s}} - f_m(s,\omega))$. A simple LS estimation is not a good solution to the depth estimation problem due to the measurement noise and the singular value of Ω. The accuracy for the CL observer, batch LS estimator, and the observer in [2] is reported for 500 Monte Carlo runs of a 50-s simulation. Initial conditions are sampled from a normal distribution centered around $\hat{s}(t_0) = [10, 5]^T$ and $\hat{\chi}(t_0) = 3$. The CL full-order observer achieves steady-state root mean square error (RMSE) of 0.046 m and mean absolute percentage error (MAPE) of 1.83%. The LS estimation achieves RMSE of 1.05 m and MAPE of 23.18%. The observer in [2] achieves

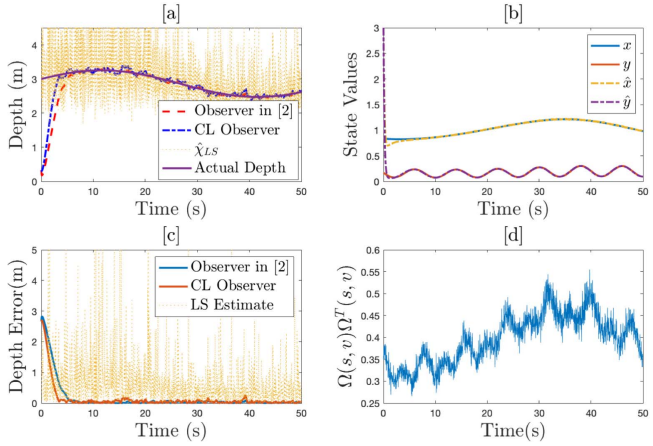

Fig. 1. (a) Comparison between the actual depth and estimated depth over a period of 50 s. (b) Comparison between the actual and estimated state values. (c) Error between the true and the estimated depth. (d) Evolution of $\Omega(s,v)\Omega^T(s,v)$ for the simulated velocities and described initial conditions.

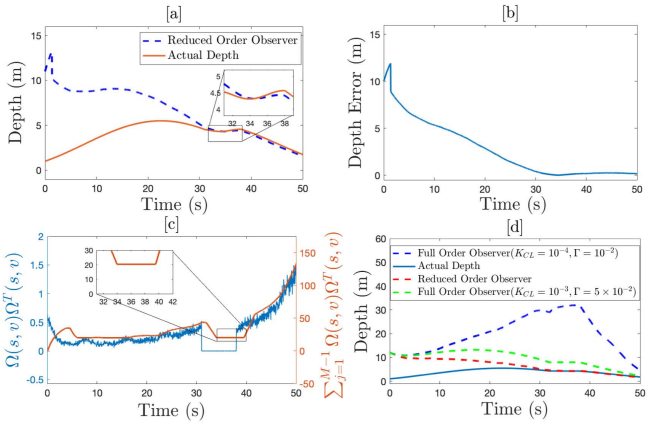

Fig. 2. (a) Comparison of true versus estimated depth using the reduced-order observer. (b) Depth estimation error for the reduced-order observer. (c) Evolution of $\Omega(s,v)\Omega^T(s,v)$ with time showing the PE violation from $t = 31$–38 s. (d) Comparison between the reduced-order integrated observer and the CL observer for different gain values in the presence of 20-dB noise.

RMSE of 0.024 m and MAPE of 1.05% when the gains are set to $\Lambda = 9$, $\boldsymbol{H} = \text{diag}\{10, 10\}$, and $\boldsymbol{Q} = \text{diag}\{4, 4\}$.

B. Simulation 2

In this simulation, the PE condition is violated between 31 and 38 s. An initial point with the Euclidean coordinates $\bar{m}(t_0) = [1, 1, 1]^T$ is used to generate the trajectories using the velocities in simulation 1 from 0 to 31 and 38 to 50 s. The linear velocities are chosen to satisfy $\Omega(s,v)\Omega^T(s,v) = 0$ at each time instant during the period 31 s − 38 s, such that $v_X(t) = x(t)v_Z(t)$, $v_Y(t) = y(t)v_Z(t)$ with $v_Z(t) = 0.1\cos(\pi t/4)$m/s and $\omega = [0, 0, 0]$ rad/s. Additive white Gaussian noise with SNR 20 dB is added to the pixel measurements, and the Gaussian distributed noise with zero mean and variance 0.01 is added to the velocity measurements. The numerically approximated value of the PE between 31 and 38 s is $\int_{31}^{38}\Omega(s(\tau),v(\tau))\Omega^T(s(\tau),v(\tau))d\tau = 7.83\times10^{-4}$. The history stack is chosen to hold 120 points, and the auxiliary stack is chosen to hold 150 points. The observer is initialized at $\hat{s}(t_0) = [1,\ 1]^T$ and $\hat{\chi}(t_0) = 0.08$. The gain $\bar{K}$ for the reduced-order observer is set to 2×10^{-3}. From Fig. 2(b), the CL reduced-order observer error exponentially converges to an ultimate bound at 35.9 s, during the period when the PE condition is violated. The value of σ_1 always remains greater than zero once the history stack is full and does not drop below ϵ. The value of ϵ is chosen to be 20, and to maintain this value, the history stack is not updated between 7–16 s and 34–40 s, as shown in Fig. 2(c). The steady-state RMSE achieved is 0.129 m, and the steady-state MAPE achieved is 3.61%. Fig. 2(d) shows the comparison of the CL full-order observer presented in Section III with the CL reduced-order observer when the state measurements are noisy.

IX. Experiments

A. Experimental Platform

The camera in the wrist of the right arm of a Baxter research robot is used to capture images containing the feature point at a rate of 30 frames/s with resolution 640×400. The centroid of white circle is used as a feature point. The processing of the images and depth estimation is done in MATLAB 2019a at 30 frames/s using a desktop with an Intel Core2Duo CPU with a clock speed of 2.26 GHz and 4-GB RAM running Ubuntu 14.04. The camera intrinsics for Baxter's right hand camera obtained through the Baxter API and Robot Operating System (ROS) are given by $f_x = f_y = 407.1$, $c_x = 323.4$, and $c_y = 205.6$, where (c_x, c_y) represents the camera center pixel. The ground-truth depth for comparing the results is obtained using simple pose transformations.

B. Results

The experiment is done for 16 s, wherein the camera is stationary for the first 1.5 s. After 1.5 s, the camera starts moving in a circular motion in the XY plane up to 5.5 s. After 5.5 s, the camera moves downward along the Z-direction and back for time up to 10 s. The motion from 10–15 s is circular in the XY plane followed by a downward motion along the Z-direction. The estimated depth by full- and reduced-order observers and the observer in [2] is shown in Fig. 3(a). The observers are initialized with initial conditions $\hat{s}(t_0) = [1,\ 1]^T$ and $\hat{\chi}(t_0) = 2.5$. The depth estimation error exponentially converges to an ultimate bound, as shown in Fig. 3(c). The camera moves in the XY plane from 1.5 s − 5.5 s, and the value of $\int_{1.5}^{5.5}\Omega(s(\tau),v(\tau))\Omega^T(s(\tau),v(\tau))d\tau = 4.12\times10^{-2}$ is approximated numerically using trapezoidal rule of integration. When PE is violated $\int_{6}^{10}\Omega(s(\tau),v(\tau))\Omega^T(s(\tau),v(\tau))d\tau = 2.15\times10^{-3}$, the violation of the PE condition is achieved by moving the camera along the Z-direction when the feature point is near the center of the image implying that $s \approx 0$. The value of the estimated image plane feature point coordinates by the observer in (5) compared with the true values is shown in Fig. 3(b). The value of the constant is chosen as $\epsilon = 0.03$ and $\sigma_1 \geq \epsilon$, as shown in Fig. 3(c), even when PE is not satisfied. The full-order observer converges in 4.7 s and achieves RMSE of 0.016 m and MAPE of 6.55%.

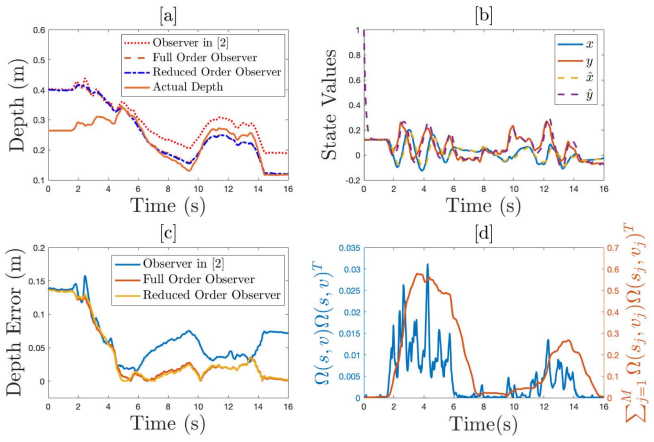

Fig. 3. (a) Comparison of true depth versus estimated depth using the full- and reduced-order observers. (b) Comparison between the actual and estimated state values using full-order observer. (c) Depth estimation error for the full- and reduced-order observers. (d) Evolution of $\Omega(s,v)\Omega^T(s,v)$ and $\sum_{j=1}^{M}\Omega(s_j,v_j)\Omega^T(s_j,v_j)$ with time.

The reduced-order observer converges in 4.7 s and achieves RMSE of 0.015 m and MAPE of 6.28%. The estimation error keeps decreasing in the first 6 s for the observer in [2] when the camera motion is informative. However, the observer does not converge after 6 s when the motion is not informative. The observer in [2] achieves RMSE of 0.046 m and MAPE of 26.11% when the gains are set to $\boldsymbol{\Lambda} = 60$, $\boldsymbol{H} = \text{diag}\{30, 30\}$, and $\boldsymbol{Q} = \text{diag}\{6, 6\}$.

X. Conclusion

The CL-based full- and reduced-order nonlinear observers are presented for estimating the depth of a stationary feature point in an image using a moving camera. The estimation errors for the full-order observer and reduced-order observer are shown to be UUB. An analytical expression is derived for the ultimate bound on the error for the presented observers. Based on the stability analyses, a procedure to update the history stack is designed which ensures convergence of the error states and frequent updates of the history stack used for CL. The developed observers are verified through numerical simulations and are successfully tested in real-world experiments on a Baxter research robot when PE is violated and the camera motions are not informative. Despite the promising results, the choice of the sizes of the history and auxiliary stacks, the observer gains, and ϵ is purely empirical. The optimal choice of these parameters is a topic for the future. The depth observer design for discrete-time systems will also be explored as a part of future work.

Appendix

A. Proof of Theorem 3

In the subsequent development, the result of [4, Th. 2] is used, which proves the existence of a candidate Lyapunov function $V(t,z) : [0,\infty) \times \mathbb{R} \to \mathbb{R}^+$, which can be upper and lower bounded by $c_5\|z\|^2 \leq V(t,z) \leq c_6\|z\|^2$ such that $c_5, c_6 > 0$ and satisfies $\|(\partial V/\partial z)\| \leq \gamma_2\|z\|$. The derivative of the Lyapunov function guarantees the exponential stability of the estimation error dynamics without the CL terms when the PE condition in (4) is satisfied. When the stack is incomplete, $\sum_{j=1}^{M-1}\Omega(s_j,v_j)\Omega^T(s_j,v_j) \geq 0$. Using the dynamics in (11), Assumption 4, definition of the history stack, result of [4, Th. 2], and completing the squares $\dot{V}$ can be upper bounded as

$$\begin{aligned}\dot{V} &\leq -\frac{k_4}{2}\|z\|^2 + \frac{(\gamma_2\bar{K}\bar{\sigma}(M-1)\bar{\chi})^2}{2k_4}\\ &\leq -\frac{k_4}{2c_6}V + \frac{(\gamma_2\bar{K}\bar{\sigma}(M-1)\bar{\chi})^2}{2k_4}\end{aligned} \tag{12}$$

where $k_4 > 0$. Using the comparison [33, Lemma 3.4], the bound on the estimation error $\|z(t)\|$ when the history stack is incomplete can be given as

$$\begin{aligned}&\|z(t)\|\\ &\leq\sqrt{\frac{c_6}{c_5}\left(\|z(t_0)\|^2 e^{-\frac{k_4}{2c_6}(t-t_0)}+\beta_3^2\left(1-e^{-\frac{k_4}{2c_6}(t-t_0)}\right)\right)}\end{aligned} \tag{13}$$

where

$$\beta_3 = \frac{\gamma_2\bar{K}\bar{\sigma}(M-1)\bar{\chi}}{k_4}$$

yields an ultimate bound on estimation error $\|z(t)\|$ according to [33, Th. 4.18]. The error $\|z(t)\|$ is UUB with an ultimate bound $((c_6/c_5))^{1/2}\beta_3$.

B. Proof of Theorem 4

Consider the candidate Lyapunov function $V(z) : \mathbb{R} \to \mathbb{R}^+$ such that $V(z) = 1/2z^Tz$. When the history stack is complete $\sum_{j=1}^{M-1}\Omega(s_j,v_j)\Omega^T(s_j,v_j) > 0$, PE is not satisfied, using the Lipschitz continuity of $g(\cdot)$, considering the gain condition $\bar{K} > (L_g/\sigma_1)$ to be satisfied and completing the squares, $\dot{V}$ can be upper bounded as

$$\begin{aligned}\dot{V} &\leq -\frac{k_5}{2}\|z\|^2 + \frac{(\bar{K}\bar{\sigma}(M-1)\bar{\chi})^2}{2k_5}\\ &\leq -k_5V + \frac{(\bar{K}\bar{\sigma}(M-1)\bar{\chi})^2}{2k_5}.\end{aligned} \tag{14}$$

Using the comparison [33, Lemma 3.4], the bound on the estimation error $\|z(t)\|$ when the history stack is complete can be given as

$$\|z(t)\| \leq \sqrt{\|z(t_0)\|^2 e^{-k_5(t-t_0)} + \beta_4^2\left(1 - e^{-k_5(t-t_0)}\right)} \tag{15}$$

where $k_5 = (\bar{K}\sigma_1 - L_g)$ and $\beta_4 = (\bar{K}\bar{\sigma}(M-1)\bar{\chi}/k_5)$. Using (14) and invoking [33, Th. 4.18], the depth error $\|z(t)\|$ is UUB with an ultimate bound β_4.

References

[1] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision: From Images to Geometric Models*. New York, NY, USA: Springer, 2012.

[2] R. Spica, P. R. Giordano, and F. Chaumette, "Active structure from motion: Application to point, sphere, and cylinder," *IEEE Trans. Robot.*, vol. 30, no. 6, pp. 1499–1513, Dec. 2014.

[3] 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.

[4] 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.
[5] L. Matthies, T. Kanade, and R. Szeliski, "Kalman filter-based algorithms for estimating depth from image sequences," *Int. J. Comput. Vis.*, vol. 3, no. 3, pp. 209–238, Sep. 1989.
[6] A. Chiuso, P. Favaro, H. Jin, and S. Soatto, "Structure from motion causally integrated over time," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 523–535, Apr. 2002.
[7] M. Jankovic and B. K. Ghosh, "Visually guided ranging from observations of points, lines and curves via an identifier based nonlinear observer," *Syst. Control Lett.*, vol. 25, no. 1, pp. 63–73, May 1995.
[8] D. Karagiannis and A. Astolfi, "A new solution to the problem of range identification in perspective vision systems," *IEEE Trans. Autom. Control*, vol. 50, no. 12, pp. 2074–2077, Dec. 2005.
[9] O. Dahl, Y. Wang, A. F. Lynch, and A. Heyden, "Observer forms for perspective systems," *Automatica*, vol. 46, no. 11, pp. 1829–1834, Nov. 2010.
[10] X. Chen and H. Kano, "State observer for a class of nonlinear systems and its application to machine vision," *IEEE Trans. Autom. Control*, vol. 49, no. 11, pp. 2085–2091, Nov. 2004.
[11] A. De Luca, G. Oriolo, and P. Robuffo 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.
[12] J. Keshavan and J. S. Humbert, "Robust structure and motion recovery for monocular vision systems with noisy measurements," *Int. J. Control*, vol. 91, no. 3, pp. 715–724, Mar. 2018.
[13] N. Gans, G. Hu, and W. E. Dixon, "Image based state estimation," in *Encyclopedia of Complexity and Systems Science*. New York, NY, USA: Springer-Verlag, 2009, pp. 4751–4776.
[14] T. Hatanaka and M. Fujita, "Cooperative estimation of averaged 3-D moving target poses via networked visual motion observer," *IEEE Trans. Autom. Control*, vol. 58, no. 3, pp. 623–638, Mar. 2013.
[15] 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.
[16] 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.
[17] J. Yang, A. Dani, S.-J. Chung, and S. Hutchinson, "Vision-based localization and robot-centric mapping in riverine environments," *J. Field Robot.*, vol. 34, no. 3, pp. 429–450, May 2017.
[18] M. Sassano, D. Carnevale, and A. Astolfi, "Observer design for range and orientation identification," *Automatica*, vol. 46, no. 8, pp. 1369–1375, Aug. 2010.
[19] I. Grave and Y. Tang, "A new observer for perspective vision systems under noisy measurements," *IEEE Trans. Autom. Control*, vol. 60, no. 2, pp. 503–508, Feb. 2015.
[20] 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.*, Dec. 2011, pp. 5005–5010.
[21] 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. 2010, pp. 6962–6967.
[22] G. Chowdhary, T. Yucelen, M. Mühlegg, and E. N. Johnson, "Concurrent learning adaptive control of linear systems with exponentially convergent bounds," *Int. J. Adapt. Control Signal Process.*, vol. 27, no. 4, pp. 280–301, Apr. 2013.
[23] R. Kamalapurkar, B. Reish, G. Chowdhary, and W. E. Dixon, "Concurrent learning for parameter estimation using dynamic state-derivative estimators," *IEEE Trans. Autom. Control*, vol. 62, no. 7, pp. 3594–3601, Jul. 2017.
[24] A. Parikh, R. Kamalapurkar, and W. E. Dixon, "Target tracking in the presence of intermittent measurements via motion model learning," *IEEE Trans. Robot.*, vol. 34, no. 3, pp. 805–819, Jun. 2018.
[25] L. D. Fairfax and P. A. Vela, "A concurrent learning approach to monocular, vision-based regulation of leader/follower systems," in *Proc. Annu. Amer. Control Conf. (ACC)*, Jun. 2018, pp. 3502–3507.
[26] D. Muthirayan and P. P. Khargonekar, "Working memory augmentation for improved learning in neural adaptive control," in *Proc. IEEE 58th Conf. Decis. Control (CDC)*, Dec. 2019, pp. 6785–6792.
[27] L. Vu, D. Chatterjee, and D. Liberzon, "Input-to-state stability of switched systems and switching adaptive control," *Automatica*, vol. 43, no. 4, pp. 639–646, Apr. 2007.
[28] G. Rotithor, R. Saltus, R. Kamalapurkar, and A. Dani, "Observer design for structure from motion using concurrent learning," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2019, pp. 2384–2389.
[29] G. Rotithor, D. Trombetta, R. Kamalapurkar, and A. P. Dani, "Reduced order observer for structure from motion using concurrent learning," in *Proc. IEEE 58th Conf. Decis. Control (CDC)*, Dec. 2019, pp. 6815–6820.
[30] G. Rotithor, D. Trombetta, R. Kamalapurkar, and A. Dani, "Extension of full and reduced order observers for image-based depth estimation using concurrent learning," 2020, *arXiv:2008.04809*. [Online]. Available: http://arxiv.org/abs/2008.04809
[31] A. Chakrabarty, M. J. Corless, G. T. Buzzard, S. H. Zak, and A. E. Rundell, "State and unknown input observers for nonlinear systems with bounded exogenous inputs," *IEEE Trans. Autom. Control*, vol. 62, no. 11, pp. 5497–5510, Nov. 2017.
[32] A. Chakrabarty and M. Corless, "Estimating unbounded unknown inputs in nonlinear systems," *Automatica*, vol. 104, pp. 57–66, Jun. 2019.
[33] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.