An adaptive optimal control approach to monocular depth observability maximization

Tochukwu Elijah Ogri[1] Muzaffar Qureshi[1] Zachary I. Bell[2] Kristy Waters[3] Rushikesh Kamalapurkar[1]

***Abstract*— This paper presents an integral concurrent learning (ICL)-based observer for a monocular camera to accurately estimate the Euclidean distance to features on a stationary object, under the restriction that state information is unavailable. Using distance estimates, an infinite horizon optimal regulation problem is solved, which aims to regulate the camera to a goal location while maximizing feature observability. Lyapunov-based stability analysis is used to guarantee exponential convergence of depth estimates and input-to-state stability of the goal location relative to the camera. The effectiveness of the proposed approach is verified in simulation, and a table illustrating improved observability through better conditioning of the regressor is provided.**

I. Introduction

The use of drones and other micro air vehicle systems has seen rapid growth in recent years due to their ability to perform dangerous or complex tasks such as surveillance, search and rescue, and weather monitoring, that are challenging or even impossible for human pilots [1], [2]. In the absence of state-feedback information from a positioning system, these robotic systems are forced to navigate, relying solely on local sensing data (e.g., camera images, inertial measurement units, and wheel encoders). The poses of objects in the surrounding environment relative to a robot must thus be determined from sensor data to inform estimation of its pose; otherwise, the performance of the controller may be affected, and the robotic system may fail to achieve its objective.

Accurately estimating the pose of a robot using cameras to reconstruct the environment using scaled Euclidean coordinates of an object is a key challenge, commonly referred to as simultaneous localization and mapping (SLAM) [3]–[6]. A significant challenge in SLAM is determining the scale of objects in a 2D image, given the loss of depth information. Several image-based methods estimate depth by reconstructing the structure of an object by using multiple images and scale information [7], [8], or by estimating motion using the camera's linear or angular velocities [9]–[24] where scales can be recovered using multiple calibrated cameras [7], [8]. However, the performance of motion-based methods is limited when the objects lack parallax between successive camera images. Alternative approaches include the use of extended Kalman filters (EKFs) [9], [11]–[13], [25] for depth estimation. For discrete time systems [26] developed EKFs with convergence guarantees. However, due to the nature of EKFs their convergence result is local, and the corresponding propagation equations are only valid if the estimates are within a small neighborhood of the actual state. With techniques, such as those proposed in [14], [16], [17], [19], [21], [23], asymptotic convergence of structure estimation errors is guaranteed, and some of them guarantee exponential convergence of scale estimates [10], [15], [18], [20], [22]. However, these methods rely on stringent conditions such as persistence of excitation (PE) or extended output Jacobian (EOJ), which may be difficult to satisfy in practice.

This paper extends the results of [24], [27], which developed exponentially converging observers using concurrent learning (CL) and integral concurrent learning (ICL) [24], [28]–[30] to estimate the Euclidean distance to features on a stationary object in the camera's field of view (FOV) under the assumption that the velocities of the camera are known. The CL-based techniques in [24], [27] guarantee exponential convergence of depth estimates while relaxing the PE assumption in favor of a finite excitation (FE) condition, which can be monitored and verified online. Without sufficient excitation, depth estimation is affected as monocular cameras cannot observe the scale of objects.

Excitation conditions require the motion of the camera to be non-parallel to the line joining the camera and the object [31]. To achieve such motion, this paper develops an adaptive optimal control scheme that plans velocities for depth observability maximization by penalizing non-orthogonal motion of a monocular camera as it attempts to reach a goal location. The paper demonstrates that feature observability can be improved through velocity planning, without the need for added excitation, by introducing a novel cost function that yields controllers with theoretical stability guarantees. To the best of our knowledge, this is the first study in the current path planning literature where adaptive optimal control is employed to plan velocities for maximizing depth observability.

II. Camera Motion Model

Consider a monocular camera that tracks features on a stationary object while the features are within its FOV using techniques similar to those in [32], [33]. Leveraging these tracked features, the camera estimates the relative distances between the features and itself, subsequently utilizing these

This research was supported by the Air Force Research Laboratories under contract number AFRL-FA8651-23-1-0006. Any opinions, findings, or recommendations in this article are those of the author(s), and do not necessarily reflect the views of the sponsoring agencies.

[1] School of Mechanical and Aerospace Engineering, Oklahoma State University, email: {tochukwu.ogri, muzaffar.qureshi, rushikesh.kamalapurkar} @okstate.edu.

[2] Air Force Research Laboratories, Florida, USA, email: zachary.bell.10@us.af.mil.

[2] Autonomous Vehicles Laboratory, University of Florida, Gainesville, Florida, USA, email: watersk@ufl.edu.

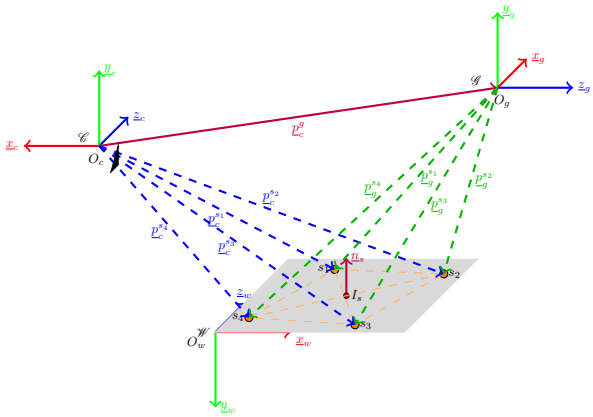

Fig. 1. Camera tracking four planar features on an object while moving from $\mathscr{C}$ to $\mathscr{G}$.

estimates to reach a user-specified goal location. Let the world frame be a fixed inertial reference frame, denoted by $\mathscr{W} := \{\vec{x}_w, \vec{y}_w, \vec{z}_w\}$, with its origin located at O_w. Let the camera frame, denoted by $\mathscr{C} := \{\vec{x}_c, \vec{y}_c, \vec{z}_c\}$ be fixed to the camera, with its origin O_c located at the principal point of the camera. Let the goal frame $\mathscr{G} := \{\vec{x}_g, \vec{y}_g, \vec{z}_g\}$, be fixed to the goal, with its origin located at O_g. The three frames are illustrated in Figure 1.

To facilitate the development of the camera model, the following assumptions are necessary.

Assumption 1: The stationary object has features that can be detected and tracked, provided it is within the camera's FOV. Specifically, $\forall t \in \mathbb{R}_{\geq 0}$, a set of at least four trackable planar features are in the camera's FOV [34].

Assumption 2: While the goal location may lie outside the camera's FOV, the position of the ith feature on the object relative to the goal position, denoted by $\underline{p}_g^{s_i} \in \mathbb{R}^3$, is known.

Assumption 3: The camera's intrinsic matrix $A \in \mathbb{R}^{3\times 3}$ is known and invertible [8].

Given a stationary object s with its ith feature denoted by s_i for all $i = 1, \ldots, n$, Assumptions 1 and 2, the position of the ith feature on the object relative to the goal, $\underline{p}_g^{s_i} \in \mathbb{R}^3$, is known. Consequently, the position of the goal relative to the camera can be determined as

$$\underline{p}_c^g(t) = \underline{p}_c^{s_i}(t) - R_c^g(t)\underline{p}_g^{s_i}, \tag{1}$$

where $\underline{p}_c^g(t) \in \mathbb{R}^3$ denotes the unknown position of the goal relative to the camera, $\underline{p}_c^{s_i}(t) \in \mathbb{R}^3$ denotes the unknown position of the ith feature on the object, with respect to $\mathscr{C}$, and $R_c^g(t) \in \mathbb{R}^{3\times 3}$ is the rotation matrix describing the orientation of $\mathscr{G}$ with respect to $\mathscr{C}$. The kinematics of the moving monocular camera relative to the goal location are given by

$$\dot{\underline{p}}_c^g(t) = \underline{v}_c^g(t) \text{ and } \dot{q}_c^g(t) = \frac{1}{2}B(q_c^g(t))\underline{\omega}_c^g(t), \tag{2}$$

where $\underline{v}_c^g(t) \in \mathbb{R}^3$ and $\underline{\omega}_c^g(t) \in \mathbb{R}^3$ represent the linear (unknown) and angular (known) velocities of $\mathscr{G}$ with respect to $\mathscr{C}$, respectively, $B(q(t)) := \begin{bmatrix} -q_v{}^\top \\ q_0\,\mathrm{I}_{3\times 3} + q_v^\times \end{bmatrix} \in \mathbb{R}^{4\times 3}$ is an orthogonal matrix which has the pseudoinverse property $B^\top(q(t))B(q(t)) = \mathrm{I}_{3\times 3}$, where $\mathrm{I}_{3\times 3}$ is a 3 by 3 identity matrix, and $q_c^g(t) \in \mathbb{R}^4$ represents the quaternion parameterization of the rotational matrix $R_c^g(t)$, describing the orientation of $\mathscr{G}$ with respect to $\mathscr{C}$, with $q := \begin{bmatrix} q_0 & q_v^\top \end{bmatrix}^\top \in \mathcal{S}^4$ which has the standard basis $\{1, i, j, k\}$, where $\mathcal{S}^4 := \{x \in \mathbb{R}^4 | x^\top x = 1\}$, and $q_0 \in \mathbb{R}$ and $q_v \in \mathbb{R}^3$ represent the scalar and vector components of q, respectively. The angular velocity of $\mathscr{G}$ with respect to $\mathscr{C}$ given as $\underline{\omega}_c^g(t)$ is assumed to be known for the rest of the development of the paper.

Equation (1) can be equivalently expressed in the form $\begin{bmatrix} \underline{u}_c^{s_i}(t) & -\underline{u}_c^g(t) \end{bmatrix} \begin{bmatrix} d_c^{s_i}(t) \\ d_c^g(t) \end{bmatrix} = R_c^g(t)\underline{u}_g^{s_i} d_g^{s_i}$, by rearranging the terms, where $d_c^{s_i}(t) \in \mathbb{R}_{>0}$ and $\underline{u}_c^{s_i}(t) \in \mathbb{R}^3$ are the magnitude and direction of the position vector $\underline{p}_c^{s_i}(t)$ of feature s_i expressed in $\mathscr{C}$, respectively; $d_c^g(t) \in \mathbb{R}_{>0}$ and $\underline{u}_c^g(t) \in \mathbb{R}^3$ are the magnitude and direction of the position vector $\underline{p}_c^g(t)$ of the goal $\mathscr{G}$ expressed in $\mathscr{C}$, respectively; and $d_g^{s_i} \in \mathbb{R}_{>0}$ and $\underline{u}_g^{s_i} \in \mathbb{R}^3$ are the magnitude and direction of the position vector $\underline{p}_g^{s_i}$ of feature s_i expressed in $\mathscr{C}$, respectively. Under Assumptions 1 - 3, the rotation matrix $R_c^g(t)$ and unit vector $\underline{u}_c^g(t)$ can be determined from a general set of features on the object using techniques such as planar homography decomposition or essential decomposition. In addition, the unit vectors $\underline{u}_g^{s_i}$ and $\underline{u}_c^{s_i}(t)$ can be obtained from $\underline{u}_g^{s_i} := \frac{P_g^{s_i}}{\|P_g^{s_i}\|}$ and $\underline{u}_c^{s_i}(t) := \frac{A^{-1}P_c^{s_i}(t)}{\|A^{-1}\underline{p}_c^{s_i}(t)\|}$ where $P_g^{s_i}$, $P_c^{s_i}(t)$ are the homogeneous coordinates of feature s_i in $\mathscr{G}$ and $\mathscr{C}$, respectively. The only remaining unknowns are the distances $d_c^{s_i}(t)$, $d_c^g(t)$ and $d_g^{s_i}$. These unknowns are estimated in the following using an ICL-based observer. To simplify the notation, let $H_{s_i}(t) := \begin{bmatrix} \underline{u}_c^{s_i}(t) & -\underline{u}_c^g(t) \end{bmatrix} \in \mathbb{R}^{3\times 2}$. While $d_c^g(t) > 0$, the term $H_{s_i}^\top(t)H_{s_i}(t)$ is invertible such that (cf. [27])

$$\begin{bmatrix} d_c^{s_i}(t) \\ d_c^g(t) \end{bmatrix} = Y_{s_i}(t)d_g^{s_i}, \tag{3}$$

where $Y_{s_i}(t) := \left(H_{s_i}^\top(t)H_{s_i}(t)\right)^{-1} H_{s_i}^\top(t)R_c^g(t)\underline{u}_g^{s_i}$ is invertible and measurable under Assumptions 1-3. Furthermore, since the goal and the stationary object are stationary, the time derivatives of the unknown distances are known and given by

$$\dot{d}_c^{s_i}(t) = -\underline{u}_c^{s_i\top}(t)\underline{v}_c(t), \tag{4}$$

$$\dot{d}_c^g(t) = -\underline{u}_c^{g\top}(t)\underline{v}_c(t), \text{ and} \tag{5}$$

$$\dot{d}_g^{s_i} = 0, \tag{6}$$

where $\underline{v}_c(t) \in \mathbb{R}^3$ represents the velocity of the camera, expressed in $\mathscr{C}$. Since the goal location is fixed in $\mathscr{W}$, the relationship between $\underline{v}_c(t)$ and $\underline{v}_c^g(t)$ is given by $\underline{v}_c(t) = -\underline{v}_c^g(t)$.

The control objective is to design the camera velocity $\underline{v}_c(t)$ to improve feature observability by maximizing orthogonal

motion of the camera with respect to the plane containing the features. The objective is achieved by using the ICL-based observer to generate estimates of the distances denoted by $\hat{d}_c^{s_i}(t) \in \mathbb{R}$, $\hat{d}_c^g(t) \in \mathbb{R}$ and $\hat{d}_g^{s_i}(t) \in \mathbb{R}$. Using these estimates and given the known position of the ith feature of the stationary object relative to the goal location, the position of the goal relative to the camera expressed in $\mathscr{C}$, $\hat{\underline{p}}_c^g$, can be estimated using

$$\hat{\underline{p}}_c^g(t) = \hat{\underline{p}}_c^{s_i}(t) - R_c^g(t)\underline{p}_g^{s_i}, \tag{7}$$

where $\hat{\underline{p}}_c^{s_i}(t) \in \mathbb{R}^3$ denotes the estimate of the position of the ith feature on the object with respect to $\mathscr{C}$. Let I_s denote the origin of the feature frame and select any three features out of the number of features on the plane that surrounds I_s as depicted in Figure 1. Let $\underline{n}_s \in \mathbb{R}^3$ represents the normal vector to the plane containing s_1, s_2, s_3, and I_s which can be expressed as $\underline{n}_s = \left(\underline{p}_w^{s_1} - \underline{p}_w^{s_2}\right) \times \left(\underline{p}_w^{s_3} - \underline{p}_w^{s_2}\right)$, where $\underline{p}_w^{s_1}$, $\underline{p}_w^{s_2}$, and $\underline{p}_w^{s_3} \in \mathbb{R}^3$ represents the position of s_1, s_2 and s_3 expressed in $\mathscr{W}$, respectively, and the notation $\times$ represents the cross product. An optimal control problem is then formulated to generate the desired linear velocity commands $\underline{v}_c^g$ for the camera, online, to minimize the cost functional

$$J(\underline{p}_c^g(\cdot), \underline{v}_c^g(\cdot)) = \int_0^\infty r_{LQR}(\underline{p}_c^g(\tau), \underline{v}_c^g(\tau)) + r_{ORTHO}\left(\underline{v}_c^g(\tau)\right) d\tau, \tag{8}$$

over the set $\mathcal{U}$ of piece-wise continuous functions and under the dynamic constraint in (2). The linear quadratic regulator (LQR) cost denoted by $r_{LQR} : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ is designed to drive the camera to the goal while the orthogonality cost denoted by $r_{ORTHO} : \mathbb{R}^3 \to \mathbb{R}$ is designed to improve estimates of $\underline{p}_c^g$ by encouraging orthogonal motion of the camera to the plane that contains the features. The LQR cost r_{LQR} is defined as $r_{LQR}(\underline{p}_c^g(t), \underline{v}_c^g(t)) := \underline{p}_c^{g\top}(t) Q_c \underline{p}_c^g(t) + \underline{v}_c^{g\top}(t) R_c \underline{v}_c^g(t)$, where $Q_c \in \mathbb{R}^{3\times 3}$ and $R_c \in \mathbb{R}^{3\times 3}$ are constant positive definite symmetric matrices. The orthogonality cost r_{ORTHO} is designed as $r_{ORTHO}(\underline{v}_c^g(t)) := \gamma_c \left(\langle \underline{v}_c^g(t), \underline{n}_s \rangle\right)^2$, where $\gamma_c \in \mathbb{R}_{>0}$ is a user-defined constant designed to maximize orthogonality of the motion of the camera relative to the feature plane. The goal to move the camera orthogonally relative to the feature plane is captured in (8) via minimization of the dot product $\langle \underline{v}_c^g(t), \underline{n}_s \rangle$.

III. ICL-Based Observer Design

An ICL update law is implemented to estimate the unknown distances $d_c^{s_i}(t)$, $d_c^g(t)$, and $d_g^{s_i}$ by integrating (4), (5) and (6), respectively, over a time delay $T \in \mathbb{R}_{>0}$ to obtain

$$\begin{bmatrix} d_c^{s_i}(t) \\ d_c^g(t) \end{bmatrix} - \begin{bmatrix} d_c^{s_i}(t-T) \\ d_c^g(t-T) \end{bmatrix} = -\int_{t-T}^{t} \begin{bmatrix} \underline{u}_c^{s_i\top}(\tau) \\ \underline{u}_c^{g\top}(\tau) \end{bmatrix} \underline{v}_c(\tau) d\tau, \tag{9}$$

for $t > T$. Substituting the relationship in equation (3) at current time t and previous time $t - T$ yields

$$\mathcal{Y}_{s_i}(t) d_g^{s_i} = \mathcal{U}_{s_i}(t) \tag{10}$$

where $\mathcal{Y}_{s_i} := \begin{cases} 0_{2\times 1}, & t \le T, \\ Y_{s_i}(t) - Y_{s_i}(t-T), & t > T, \end{cases}$ and $\mathcal{U}_{s_i}(t) := \begin{cases} 0_{2\times 1}, & t \le T, \\ -\int_{t-T}^{t} \begin{bmatrix} \underline{u}_c^{s_i\top}(\tau) \\ \underline{u}_c^{g\top}(\tau) \end{bmatrix} \underline{v}_c(\tau) d\tau, & t > T. \end{cases}$

Multiplying both sides of (10) by the term $\mathcal{Y}_{s_i}^\top(t)$ yields

$$\mathcal{Y}_{s_i}^\top(t)\mathcal{Y}_{s_i}(t) d_g^{s_i} = \mathcal{Y}_{s_i}^\top(t)\mathcal{U}_{s_i}(t) \tag{11}$$

In general, $\mathcal{Y}_{s_i}(t)$ will not have full column rank (e.g. when the camera is stationary) implying $\mathcal{Y}_{s_i}^\top(t)\mathcal{Y}_{s_i}(t)$ is positive semidefinite but not positive definite. However, the equality in (11) may be evaluated at several (possibly time-varying) time instances $t_1, \ldots, t_N$ and summed together to yield

$$\Sigma_{\mathcal{Y}_{s_i}}(t) d_g^{s_i} = \Sigma_{\mathcal{U}_{s_i}}(t) \tag{12}$$

where $\Sigma_{\mathcal{Y}_{s_i}}(t) := \sum_{j=1}^{N} \mathcal{Y}_{s_i}^\top(t_j(t))\mathcal{Y}_{s_i}(t_j(t))$, $\Sigma_{\mathcal{U}_{s_i}}(t) := \sum_{j=1}^{N} \mathcal{Y}_{s_i}^\top(t_j(t))\mathcal{U}_{s_i}(t_j(t))$, and $N \in \mathbb{Z}_{\ge 1}$. The following assumption is an observability-like condition that must be satisfied to guarantee convergence of distance estimates in finite time.

Assumption 4: The camera has sufficiently rich motion so that there exist constants $\tau \in \mathbb{R}_{>T}$ and $\lambda_\tau \in \mathbb{R}_{>0}$ such that for all $t \ge \tau$, $\lambda_{\min}\{\Sigma_{\mathcal{Y}_{s_i}}(t)\} > \lambda_\tau$, where $\lambda_{\min}\{\cdot\}$ denotes the minimum eigenvalue of $\{\cdot\}$.

Remark 1: Assumption 4 can be verified online and is easy to satisfy provided the trajectories contain sufficient information to make $\mathcal{Y}_{s_i}$ sufficiently exciting on a finite interval [27], [34], [35].

The time τ is unknown; however, it can be determined online by checking the minimum eigenvalue of $\Sigma_{\mathcal{Y}_{s_i}}(t)$. After $t = \tau$, $\lambda_{\min}\{\Sigma_{\mathcal{Y}_{s_i}}(t)\} > \lambda_\tau$ implies that the constant unknown distance $d_g^{s_i}$ can be determined from (12) and obtained as $d_g^{s_i} = \begin{cases} 0, & t < \tau, \\ \Sigma_{\mathcal{Y}_{s_i}}^{-1}(t)\Sigma_{\mathcal{U}_{s_i}}(t), & t \ge \tau. \end{cases}$

Substituting this expression into (3) yields $\begin{bmatrix} d_c^{s_i}(t) \\ d_c^g(t) \end{bmatrix} = \begin{cases} 0, & t < \tau, \\ Y_{s_i}(t)\Sigma_{\mathcal{Y}_{s_i}}^{-1}(t)\Sigma_{\mathcal{U}_{s_i}}(t), & t \ge \tau. \end{cases}$ Based on subsequent stability analysis, ICL update laws to generate the estimates $\hat{d}_c^{s_i}(t)$, $\hat{d}_c^g(t)$, and $\hat{d}_g^{s_i}$ are designed as

$$\dot{\hat{d}}_c^{s_i}(t) := \begin{cases} \eta_{s_i,1}(t), & t < \tau, \\ \eta_{s_i,1}(t) + \kappa_1 \left(\nu_{s_i,1}(t) - \hat{d}_c^{s_i}(t)\right), & t \ge \tau, \end{cases} \tag{13}$$

$$\dot{\hat{d}}_c^{g}(t) := \begin{cases} \eta_{s_i,2}(t), & t < \tau, \\ \eta_{s_i,2}(t) + \kappa_2 \left(\nu_{s_i,2}(t) - \hat{d}_c^{g}(t)\right), & t \ge \tau, \end{cases} \tag{14}$$

and

$$\dot{\hat{d}}_g^{s_i}(t) := \begin{cases} 0, & t < \tau, \\ \kappa_3 \left(\Sigma_{\mathcal{Y}_{s_i}}^{-1}(t)\Sigma_{\mathcal{U}_{s_i}}(t) - \hat{d}_g^{s_i}(t)\right), & t \ge \tau, \end{cases} \tag{15}$$

respectively, where $\eta_{s_i}(t) := -\begin{bmatrix} \underline{u}_c^{s_i\top}(t) \\ \underline{u}_c^{g\top}(t) \end{bmatrix} \underline{v}_c(t)$, $\nu_{s_i}(t) := Y_{s_i}(t)\Sigma_{\mathcal{Y}_{s_i}}^{-1}(t)\Sigma_{\mathcal{U}_{s_i}}(t)$, and $\kappa_1 \in \mathbb{R}_{>0}$, $\kappa_2 \in \mathbb{R}_{>0}$, and $\kappa_3 \in \mathbb{R}_{>0}$ are user-selected gains. Let $\tilde{d}_c^{s_i}(t) \in \mathbb{R}, \tilde{d}_c^{g}(t) \in \mathbb{R}$ and $\tilde{d}_g^{s_i}(t) \in \mathbb{R}$ represent the distance estimation errors defined as $\tilde{d}_c^{s_i}(t) := d_c^{s_i}(t) - \hat{d}_c^{s_i}(t)$, $\tilde{d}_c^{g}(t) := d_c^{g}(t) - \hat{d}_c^{g}(t)$ and $\tilde{d}_g^{s_i}(t) := d_g^{s_i} - \hat{d}_g^{s_i}(t)$, respectively. Taking their derivatives and substituting the dynamics in (4), (5), and (6) and update laws in (13), (14), and (15) yields

$$\dot{\tilde{d}}_c^{s_i}(t) := \begin{cases} 0, & t < \tau, \\ -\kappa_1 \tilde{d}_c^{s_i}(t), & t \geq \tau, \end{cases} \tag{16}$$

$$\dot{\tilde{d}}_c^{g}(t) := \begin{cases} 0, & t < \tau, \\ -\kappa_2 \tilde{d}_c^{g}(t), & t \geq \tau, \end{cases} \tag{17}$$

and

$$\dot{\tilde{d}}_g^{s_i}(t) := \begin{cases} 0, & t < \tau, \\ -\kappa_3 \tilde{d}_g^{s_i}(t), & t \geq \tau, \end{cases} \tag{18}$$

The subsequent analysis in Section V shows that the error $\tilde{d}_c^{s_i}$ remains bounded for $t < \tau$ and decays exponentially for $t \geq \tau$, once sufficient data has been gathered.

IV. Design of feature observability maximizing Velocity

This section presents an analytical solution to the optimal control problem in (8) using estimates of the position of the goal relative to the camera $\hat{\underline{p}}_c^g(t)$ obtained from the results of the observer in Section III. The Hamilton-Jacobi-Bellman (HJB) equation for the optimal control problem in (8) can be expressed in the form,

$$0 = \min_{\underline{v}_c^g} \Big\{ J^{*\prime}(\underline{p}_c^g)\underline{v}_c^g(\underline{p}_c^g) + \underline{p}_c^{g\top} Q_c \underline{p}_c^g + \underline{v}_c^{g*\top}(\underline{p}_c^g) R_c \underline{v}_c^{g*}(\underline{p}_c^g) + \gamma_c \left(\langle \underline{v}_c^{g*}(\underline{p}_c^g), \underline{n}_s \rangle \right)^2 \Big\}, \tag{19}$$

where $J^* : \mathbb{R}^3 \to \mathbb{R}$ is the optimal cost-to-go. Since the position dynamics in (2) are linear and the cost in (8) is quadratic, the optimal cost-to-go is given by $J^*(\underline{p}_c^g) := \underline{p}_c^{g\top} S_c \underline{p}_c^g$, where $S_c \in \mathbb{R}^{3\times 3}$ is a constant positive definite symmetric matrix, and the notation $(\cdot)'$ is used to denote $\frac{\partial}{\partial(\cdot)}$. The optimal control policy, denoted by $\underline{v}_c^{g*} : \mathbb{R}^3 \to \mathbb{R}^3$, is given as

$$\underline{v}_c^{g*}(\underline{p}_c^g(t)) = -\overline{R}_c^{-1} S_c \underline{p}_c^g(t), \tag{20}$$

where $\overline{R}_c \in \mathbb{R}^{3\times 3}$ is a positive definite matrix defined as $\overline{R}_c := R_c + \gamma_c N_s$ and $N_s \in \mathbb{R}^{3\times 3}$ is a positive semi-definite symmetric defined as $N_s := \underline{n}_s \underline{n}_s^\top$. Since the matrix $\overline{R}_c$ is the sum of a symmetric positive definite matrix and a symmetric positive semi-definite matrix, it is also symmetric and positive definite. Substituting the (20) back into the HJB (19) equation and simplifying yields the following necessary and sufficient condition for optimality

$$-S_c \overline{R}_c^{-1} S_c + Q_c = 0, \tag{21}$$

where the objective is to find the matrix S_c. Given symmetric positive semi-definite matrices $\overline{R}_c$, Q_c and S_c, the solution to the quadratic equation in (21) is unique and is given as

$$S_c = \overline{R}_c^{1/2} (\overline{R}_c^{-1/2} Q_c \overline{R}_c^{-1/2})^{\frac{1}{2}} \overline{R}_c^{1/2}. \tag{22}$$

Since $\underline{p}_c^g(t)$ is unknown, the linear velocity of the camera is subsequently designed using the estimate $\hat{\underline{p}}_c^g(t)$ as

$$\underline{v}_c(t) := -\hat{\underline{v}}_c^g(t) = K_s \hat{\underline{p}}_c^g(t), \tag{23}$$

where $K_s \in \mathbb{R}^{3\times 3}$ is the feedback gain defined as $K_s := \overline{R}_c^{-1} S_c$. The velocity $\underline{v}_c(t)$, when represented in $\mathscr{W}$, is denoted by $\underline{v}_w^c(t) \in \mathbb{R}^3$ and given by $\underline{v}_w^c(t) = K_s R_w^c(t) \hat{\underline{p}}_c^g(t)$ where $R_w^c(t)$ is the orientation of $\mathscr{C}$ with respect to $\mathscr{W}$.

V. Stability Analysis

This section presents the main theoretical results of this paper. First, the convergence properties of the proposed observers in Section III are presented, and finally, the convergence of the position error trajectory $\underline{p}_c^g(t)$ to a given neighborhood of the origin is presented.

A. Analysis of Camera ICL Observer Error system

Let $\tilde{\vartheta}(t) \in \mathbb{R}^9$ denote a concatenated state vector containing the distance estimation errors, defined as $\tilde{\vartheta}(t) := \begin{bmatrix} \tilde{d}_c^{s_i}(t) & \tilde{d}_c^{g}(t) & \tilde{d}_g^{s_i}(t) \end{bmatrix}^\top$ and let $L : \mathbb{R}^9 \to \mathbb{R}$ be a candidate Lyapunov function defined as

$$L(\tilde{\vartheta}(t)) = \frac{1}{2} \tilde{\vartheta}^\top(t) \tilde{\vartheta}(t). \tag{24}$$

The following theorem establishes the exponential stability of the observer error system obtained in (16), (17), and (18).

Theorem 1: Provided Assumptions 1-4 hold, the update laws defined in (13), (14), and (15) ensure that the origin of the observer error system is globally exponentially stable and the trajectories of the estimation errors $\tilde{\vartheta}(\cdot)$ converge exponentially to the origin.

Proof: Taking the orbital derivative of the candidate Lyapunov function in (24), along the solutions of (16), (17), and (18), simplifying, and upper bounding, yields the inequality

$$\dot{L}(\tilde{\vartheta}(t)) \leq \begin{cases} 0, & t < \tau, \\ -2\kappa L(\tilde{\vartheta}(t)), & t \geq \tau, \end{cases} \tag{25}$$

where $\kappa = \min\{\kappa_1, \kappa_2, \kappa_3\}$. At $t < \tau$, it can be observed from (24) and (25) that the distance estimation errors in $\tilde{\vartheta}(t)$ are non-increasing, specifically $\tilde{\vartheta}(t) \leq \vartheta(0), \forall t < \tau$. Invoking [36, Theorem 4.10], it can be concluded that the observer error system is exponentially stable and by the Comparison Lemma [36, Lemma 3.4], the bound $\|\tilde{\vartheta}(t)\| \leq \|\vartheta(\tau)\| e^{-\kappa(t-\tau)}$ holds for all $t \geq \tau$. ■

B. Analysis of position error system

To facilitate the following analysis, let $\Gamma_s := S_c R_c^{-1} S_c$ and note that $\lambda_{\min}(\Gamma_s)\|\underline{p}_c^g(t)\|^2 \leq \underline{p}_c^{g\top}(t)\Gamma_s \underline{p}_c^g(t) \leq \lambda_{\max}(\Gamma_s)\|\underline{p}_c^g(t)\|^2$. Using the optimal cost-to-go function J^* as the candidate Lyapunov function, the following theorem

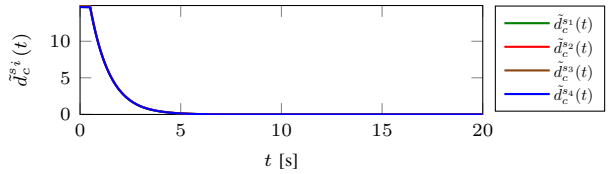

Fig. 2. Trajectory of the distance error of the features of the object relative to the camera.

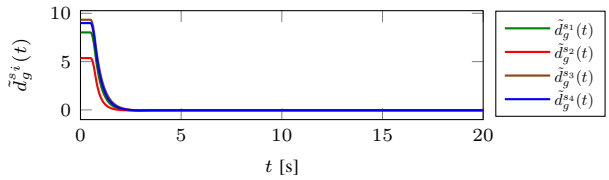

Fig. 3. Trajectory of the distance error of the features of the object relative to the goal.

establishes the input-to-state stability of the position error system.

Theorem 2: Provided Assumption 1-4 hold and the estimated distances $\hat{d}_c^{s_i}(\cdot)$, $\hat{d}_c^{g}(\cdot)$, and $\hat{d}_g^{s_i}(\cdot)$ are updated using the update laws defined in (13), (14), and (15) respectively such that the conditions of Theorem 1 are satisfied, then the system in (2) is input-to-state stable with state $\underline{p}_c^g(\cdot)$ and input $\sqrt{\|\tilde{d}_c^g(\cdot)\|}$.

Proof: The orbital derivative of the optimal cost-to-go function is bounded as $\dot{J}(\underline{p}_c^g(t)) \leq -2\lambda_{\min}(\Gamma_s)\|\underline{p}_c^g(t)\|^2 + 2\lambda_{\max}(\Gamma_s)\|\underline{p}_c^g(t)\|\|\tilde{\underline{p}}_c^g(t)\|$. Applying completion of squares and using the fact that $\sup_{t\in\mathbb{R}_{\geq 0}} \|\underline{u}_c^g(t)\| \leq 1$ since $\underline{u}_c^g(t)$ is a unit vector, the orbital derivative is bounded for all $t \geq 0$ as $\dot{J}(\underline{p}_c^g(t)) \leq -\lambda_{\min}(\Gamma_s)\|\underline{p}_c^g(t)\|^2, \forall \|\underline{p}_c^g(t)\| \geq \varrho\left(\sqrt{\|\tilde{d}_c^g(\cdot)\|}\right)$, where $\varrho\left(\sqrt{\|\tilde{d}_c^g(\cdot)\|}\right) := \sqrt{\frac{2\lambda_{\max}(\Gamma_s)}{\lambda_{\min}(\Gamma_s)}}\|\tilde{d}_c^g(\cdot)\|$. Therefore, the conditions of [36, Theorem 4.19] are satisfied and can be concluded that the system in (2) is input-to-state stable with state $\underline{p}_c^g(\cdot)$ and input $\sqrt{\|\tilde{d}_c^g(\cdot)\|}$. Since the distance error $\tilde{d}_c^g(\cdot)$ converges exponentially to the origin according to Theorem 1, the results of [36, Exercise 4.58] can be used to show that as $t \to \infty$ and the input $\sqrt{\|\tilde{d}_c^g(\cdot)\|}$ converges to zero, so does the state $\underline{p}_c^g(\cdot)$. ■

VI. Simulation Study

To demonstrate the performance of the developed observers and to test the effects of the orthogonality cost r_{ORTHO} on feature observability, consider a monocular camera with dynamics as defined in (2), which is tracking four co-planar features on a stationary object as described in Figure 1. The control objective is to move the camera from the initial position to the goal location using the control policy in (23), which uses estimates of the position of the camera obtained from the ICL-based observers developed Section III. The simulation parameters are omitted for brevity of the paper and are available in the *arXiv* version of this paper.

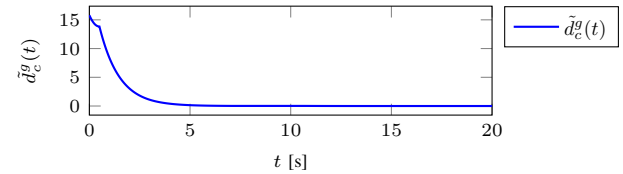

Fig. 4. Trajectory of the distance error of the goal relative to the camera.

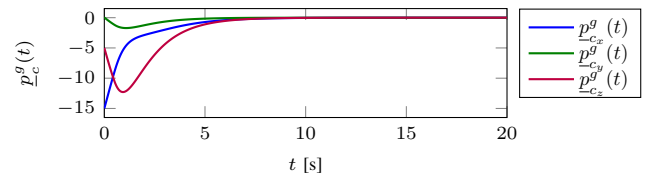

Fig. 5. The trajectories of the actual goal position relative to the camera expressed in $\mathscr{W}$.

A. Results

From Figures 2, 3, and 4, it can be observed that the trajectories of the distance errors converge exponentially to the origin which is consistent with the results of Theorem 1. Similarly, it can be observed in Figure 6 and in Figure 5 that the trajectories of the actual and estimated position of the goal relative to the camera $\underline{p}_c^g(\cdot)$ and $\hat{\underline{p}}_c^g(\cdot)$, respectively, decreases and eventually converges to the origin which is consistent with the results of Theorem 2.

TABLE I
Effect of Varied Orthogonality Penalty Gains (γ_c) on Regressor Conditioning.

γ_c	0	5	10	15	25	50
Avg. Cond. no	16.289	9.906	6.199	5.239	3.279	2.718

Table I illustrates how the conditioning of the regressor $\Sigma_{\mathcal{Y}si}$, as defined in (12), varies with increasing γ_c values. A lower condition number implies better numerical stability while estimating the Euclidean distance to the features, maximizing feature observability. Conversely, a higher condition indicates heightened sensitivity to measurement errors, indicating that the regressor is poorly conditioned, resulting in less accurate and reliable estimates. The result of Table I demonstrates the impact of the added cost r_{ORTHO} on obtaining better scale estimates; however, increasing the camera gain γ_c beyond a certain threshold can negatively affect the performance of the controller in achieving its

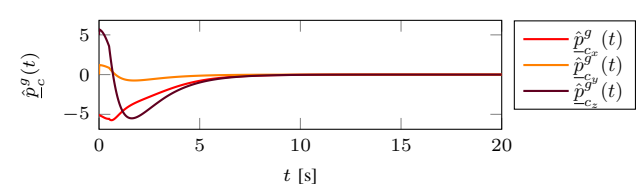

Fig. 6. The trajectories of the estimated goal position relative to the camera expressed in $\mathscr{W}$.

objective of reaching the goal position. A careful choice of the parameter γ_c allows for the right tradeoff between maximizing feature observability and achieving the goal, an important consideration in real-time systems, where the camera's objective to reach the goal must be balanced with the need to observe the features of landmarks within the operating environment.

VII. Conclusion

This paper develops a technique to plan trajectories for a monocular camera to maximize the observability of the features of a stationary object by formulating an optimal control problem whose objective is to reach a goal location while using estimates generated by ICL-based observers. The developed method does not require the positive depth constraint, which requires that the distance from the focal point of the camera to the target along the axis perpendicular to the image plane must remain positive, or the PE condition, which is difficult to satisfy in practice. As evidenced by the results described in Table I, noticeable improvements are obtained due to the added orthogonality cost designed to maximize observability. Future work will involve extending these results to nonplanar features on an object, as well as multiple objects that are non-stationary.

References

[1] H. Yu, R. Sharma, R. W. Beard, and C. N. Taylor, "Observability-based local path planning and obstacle avoidance using bearing-only measurements," *Robot. Autom. Sys.*, vol. 61, no. 12, pp. 1392–1405, 2013.

[2] J. Delaune, D. S. Bayard, and R. Brockers, "Range-visual-inertial odometry: Scale observability without excitation," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 2421–2428, 2021.

[3] G. Dubbelman and B. Browning, "COP-SLAM: Closed-form online pose-chain optimization for visual SLAM," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1194–1213, 2015.

[4] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, 2015.

[5] T. Taketomi, H. Uchiyama, and S. Ikeda, "Visual SLAM algorithms: A survey from 2010 to 2016," *IPSJ Trans. Comput. Vis. Appl.*, vol. 9, no. 1, pp. 1–11, 2017.

[6] M. Karrer, P. Schmuck, and M. Chli, "CVI-SLAM—collaborative visual-inertial SLAM," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 2762–2769, 2018.

[7] R. I. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, 2003.

[8] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An invitation to 3-D vision*. Springer, 2004.

[9] L. Matthies, T. Kanade, and R. Szeliski, "Kalman filter-based algorithm for estimating depth from image sequences," *Int. J. Comput. Vis.*, vol. 3, pp. 209–236, 1989.

[10] M. Jankovic and B. K. Ghosh, "Visually guided ranging from observations points, lines and curves via an identifier based nonlinear observer," *Syst. Control Lett.*, vol. 25, no. 1, pp. 63–73, 1995.

[11] S. Soatto, R. Frezza, and P. Perona, "Motion estimation via dynamic vision," *IEEE Trans. Autom. Control*, vol. 41, no. 3, pp. 393–413, 1996.

[12] H. Kano, B. K. Ghosh, and H. Kanai, "Single camera based motion and shape estimation using extended Kalman filtering," *Math. Comput. Modell.*, vol. 34, pp. 511–525, 2001.

[13] 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.

[14] W. E. Dixon, Y. Fang, D. M. Dawson, and T. J. Flynn, "Range identification for perspective vision systems," *IEEE Trans. Autom. Control*, vol. 48, pp. 2232–2238, 2003.

[15] 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, 2004.

[16] 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, 2005.

[17] D. Braganza, D. M. Dawson, and T. Hughes, "Euclidean position estimation of static features using a moving camera with known velocities," in *Proc. IEEE Conf. Decis. Control*, New Orleans, LA, USA, Dec. 2007, pp. 2695–2700.

[18] 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, 2008.

[19] G. Hu, D. Aiken, S. Gupta, and W. E. Dixon, "Lyapunov-based range identification for a paracatadioptric system," *IEEE Trans. Autom. Control*, vol. 53, no. 7, pp. 1775–1781, 2008.

[20] F. Morbidi and D. Prattichizzo, "Range estimation from a moving camera: an immersion and invariance approach," in *Proc. IEEE Int. Conf. Robot. Autom.*, IEEE. IEEE, 2009, pp. 2810–2815.

[21] N. Zarrouati, E. Aldea, and P. Rouchon, "So(3)-invariant asymptotic observers for dense depth field estimation based on visual data and known camera motion," in *Proc. Am. Control Conf.*, Fairmont Queen Elizabeth, Montreal, Canada, Jun. 2012, pp. 4116–4123.

[22] A. Dani, N. Fischer, Z. Kan, and W. E. Dixon, "Globally exponentially stable observer for vision-based range estimation," *Mechatronics*, vol. 22, no. 4, pp. 381–389, 2012, special Issue on Visual Servoing.

[23] A. P. Dani, N. R. Fischer, and W. E. Dixon, "Single camera structure and motion," *IEEE Trans. Autom. Control*, vol. 57, no. 1, pp. 238–243, 2011.

[24] Z. I. Bell, H.-Y. Chen, A. Parikh, and W. E. Dixon, "Single scene and path reconstruction with a monocular camera using integral concurrent learning," in *Proc. IEEE Conf. Decis. Control*, IEEE. IEEE, 2017, pp. 3670–3675.

[25] K. Reif and R. Unbehauen, "The extended Kalman filter as an exponential observer for nonlinear systems," *IEEE Trans. Signal Process.*, vol. 47, no. 8, pp. 2324–2328, 1999.

[26] M. Boutayeb, H. Rafaralahy, and M. Darouach, "Convergence analysis of the extended Kalman filter used as an observer for nonlinear deterministic discrete -time systems," *IEEE Trans. Autom. Control*, vol. 42, no. 4, pp. 581–586, 1997.

[27] Z. I. Bell, P. Deptula, E. A. Doucette, J. W. Curtis, and W. E. Dixon, "Simultaneous estimation of Euclidean distances to a stationary object's features and the Euclidean trajectory of a Monocular camera," *IEEE Trans. Autom. Control*, vol. 66, no. 9, pp. 4252–4258, 2020.

[28] G. V. Chowdhary and E. N. Johnson, "Theory and flight-test validation of a concurrent-learning adaptive controller," *J. Guid. Control Dynam.*, vol. 34, no. 2, pp. 592–607, Mar. 2011.

[29] G. Chowdhary, M. Mühlegg, J. P. How, and F. Holzapfel, "Concurrent learning adaptive model predictive control," in *Proc. AIAA Guid. Navig. Control Conf.*, Springer. American Institute of Aeronautics and Astronautics, 2013, pp. 29–47.

[30] 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, 2013.

[31] Z. I. Bell, C. Harris, R. Sun, and W. E. Dixon, "Structure and velocity estimation of a moving object via synthetic persistence by a network of stationary cameras," in *Proc. IEEE Conf. Decis. Control*, IEEE. IEEE, 2019, pp. 1601–1606.

[32] J.-Y. Bouguet, "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm," *Intel corporation*, vol. 5, no. 1-10, p. 4, 2001.

[33] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artif. Intell.*, 1981, pp. 674–679.

[34] A. Parikh, R. Kamalapurkar, H.-Y. Chen, and W. E. Dixon, "Homography based visual servo control with scene reconstruction," in *Proc. IEEE Conf. Decis. Control*, Osaka, Japan, Dec. 2015, pp. 6972–6977.

[35] Z. Bell, A. Parikh, J. Nezvadovitz, and W. E. Dixon, "Adaptive control of a surface marine craft with parameter identification using integral concurrent learning," in *Proc. IEEE Conf. Decis. Control*, 2016.

[36] H. K. Khalil, *Nonlinear systems*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2002.