

Position Estimation Using a Monocular Camera Subject to Intermittent Feedback

Zachary I. Bell, Runhan Sun, Ryan A. Licitra, Emily A. Doucette, J. Willard Curtis, Warren E. Dixon

Abstract—An estimator is developed that uses a monocular camera to estimate the position of an agent when state-feedback from a positioning system is unavailable. Often, a predictor-only strategy is used to estimate the state of the agent while state-feedback is unavailable; however, the estimation error of the state can be reduced through the use of a stable image-based observer. The developed estimator uses a monocular camera while traveling over a large distance and doesn't require objects to remain in the camera's field-of-view (FOV). While a stationary object is in the FOV, a set of observers is used to estimate the Euclidean scale of a stationary object's features while allowing those features to permanently leave the FOV after a finite period of time. If the set of observers for an object's features satisfy a minimum dwell-time condition, developed using a Lyapunov-based analysis, the learned Euclidean structure of the features is used in an observer for the Euclidean position of the camera. When the object leaves the camera's FOV, a predictor is used to estimate the position of the camera until another object's structure is learned. This new strategy allows a camera to travel over large distances without keeping a specific object in the FOV for all time, while maintaining a continuous estimate of the state of the camera. A Lyapunov-based stability analysis shows the estimator strategy exponentially converges. An experimental study is presented to demonstrate the effectiveness of the observer and predictor strategy when compared to a predictor-only strategy.

Index Terms—Nonlinear observers, dwell-time conditions, intermittent state feedback, hybrid systems theory, structure from motion, vision-based localization

I. INTRODUCTION

In many applications, an agent must operate in feedback-denied regions where feedback of the agent's state (e.g., position) is temporarily or permanently unavailable because of task definition, operating environment, or sensor vulnerabilities (e.g., position from a global positioning system (GPS) is unavailable while operating in contested environments). In these scenarios, predictor strategies can be used to estimate the agent's state (cf., [1] and [2]); however, relying on predictor-only strategies in feedback-denied regions results in error growth in state estimates, reducing dwell-times. A predictor and observer strategy using local sensing data (e.g., camera

Zachary I. Bell, Runhan Sun, Ryan A. Licitra, and Warren E. Dixon are with the Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL, 32611-6250, USA. Email:{bellz121, runhansun, rllicitra, wdixon}@ufl.edu. Emily A. Doucette and J. Willard Curtis are with the Munitions Directorate, Air Force Research Laboratory, Eglin AFB, FL, USA. Email:{emily.doucette, jess.curtis}@us.af.mil.

This research is supported in part by NEEC award number N00174-18-1-0003, a Task Order contract with the Air Force Research Laboratory, Munitions Directorate at Eglin AFB, AFOSR award number FA9550-18-1-0109, and AFOSR award number FA9550-19-1-0169. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the sponsoring agency.

systems, inertial measurement units, and wheel encoders) can be used to reduce state estimation error in feedback-denied environments, and hence lengthen the dwell-times. In these scenarios, the relative positions of objects in the surrounding environment must be determined from sensor data to inform an observer estimating the state of the agent. This problem is well known as simultaneous localization and mapping (SLAM) (cf., [3]–[8]). Many of these tasks require traveling over large distances implying the local environment for an agent is always changing which complicates estimating the agent's position; however, image-based observers can be used to reduce the error in the estimated position of an agent by accurately determining the scaled Euclidean coordinates of features on objects in the surrounding environment. Specifically, agents equipped with camera systems can use image-based observers to improve pose estimates (i.e., camera-in-hand estimation).

Using image-based methods to reconstruct the surrounding environment (i.e., determine the Euclidean scale of objects in the environment) requires the assumption that while an object's features are in the camera's field-of-view (FOV) they can be extracted and tracked through a sequence of images. However, a significant challenge arises in determining the scale of objects in an image given the lack of depth information. Specifically, images of objects are 2D projections of the 3D environment. Approaches that use image-based methods to estimate the structure of objects (i.e., determine the scaled Euclidean coordinates of features) use multiple images of an object along with scale information (cf., [9], [10]) or motion (cf., [11]–[27]), such as, linear and angular velocities of the camera. The latter of these methods is referred to as structure from motion (SfM). Generally, the Euclidean scale of an object is unknown; however, multiple calibrated cameras may be used in a stereo vision approach to recover the scale (cf., [9], [10]). However, relying on multiple cameras is limiting because some objects may have limited or no parallax between the camera images. In SfM approaches, the potential for limited parallax still exists; however, a camera can generate enough parallax by traveling over larger distances while tracking the object, which is generally not possible in stereo vision.

Image-based methods using SfM are generally online iterative methods (cf., [11]–[27]) and offline batch methods (cf., [9], [10], [28] and the references contained within). Offline approaches perform an optimization over an image sequence, but only show convergence for limited cases (cf., [29], [30]). Most online SfM approaches assume continuous measurements of objects or only update when a new image is received (cf., [11]–[21], [23]–[27]); however, recent more general results allow objects to temporarily leave the FOV

(cf., [31]–[33]). Many results apply the extended Kalman filter (EKF) to estimate depth, (cf., [11], [13]–[15], [22]); however, the EKF generally does not guarantee convergence unless strict excitation conditions are satisfied (cf., [34], [35]). Compared to the EKF, techniques such as [16], [18], [19], [21], [24], [26], show asymptotic convergence of the structure estimation errors. Furthermore, results such as [12], [17], [20], [23], [25] show exponential convergence of the scale estimate assuming some form of a persistence of excitation (PE), condition or the more strict Extended Output Jacobian (EOJ), is satisfied. Specifically, exponential convergence is obtained in [20] assuming the PE condition is met and either the initial estimation error is small or the relative velocities are limited. Furthermore, the development in [23] yields exponential convergence assuming the observer satisfies the EOJ condition. In [25], an exponentially stable observer is developed that requires the motion along at least one axis to be nonzero, and the observer remains ultimately bounded if the PE assumption does not hold, while in [23] the observer becomes singular. Typically, SfM approaches require the motion (e.g., linear and angular velocities) to be known; however, the design in [26], extending an approach similar to [25], demonstrates a partial solution to the more challenging problem (i.e. compared to SfM) of structure and motion (SaM) estimation where not only are the feature Euclidean coordinates estimated, but also two of the linear velocities and the three angular velocities of the camera assuming PE and the linear velocity and acceleration is measurable along one axis.

In [27] and [36], exponentially converging observers are developed that use a camera to estimate the Euclidean distance to features on a stationary object in the camera FOV while also estimating the Euclidean trajectory of the camera tracking the object. Unlike previous methods that assume a PE condition, the developed estimator only requires finite excitation. The finite excitation condition results from the use of concurrent learning (CL) (cf., [37]–[40]). The CL strategy is to use recorded input and output data from system trajectories to identify uncertain constant parameters of the system in real time under the assumption that the system is sufficiently excited for a finite amount of time. This approach relaxes the PE assumption and can be monitored and verified online. To eliminate the need to measure the highest order derivative of the state, we specifically use integral concurrent learning (ICL) (cf., [27], [36], [41]–[44]). ICL removes the necessity to estimate the highest order derivative of the system required in traditional concurrent learning.

Although ICL removes the need for measuring the state derivative, it still requires the state to be measurable; yet, a unique challenge in [27] and [36] is that the state depends on the unmeasurable distance to the target. Moreover, the traditional state used in results such as [11]–[21], [23]–[26] include an inherent singularity when one of the coordinates becomes zero (i.e., the so-called depth to the target). Specifically, previous results assume a positive depth constraint where the distance from the focal point of the camera to the target along the axis perpendicular to the image plane remains positive. The positive depth constraint is satisfied if the features remain in the FOV; however, the constraint can be violated for some

camera rotations that cause the feature to leave the FOV. Recently developed vision based switching theory in [31]–[33] allow features to temporarily leave the FOV, given dwell time conditions are satisfied on the amount of time a feature is in and out of the FOV. While this approach may solve issues that arise from occlusions over small periods or for reconstructing environments in a local region, it still may not be possible or desirable to return to some regions in large environments, motivating the need for techniques that can provide estimates of features that permanently leave the FOV. An additional drawback to [31]–[33] is the requirement that the features remain in front of the camera (i.e., positive depth constraint) for all time, even if they leave the FOV.

One approach that allows features to leave the FOV for extended periods of time or permanently, enabling the camera to travel over large distances, is the use of multiple sets of tracked features (cf., [3]–[8], [45]–[56]); however, most of these approaches assume that each of the new sets of features can be approximated or extracted before older sets of features leave the FOV (cf., [45], [46], [49], [50], [52], [54], [55]). In many applications, it is not possible to ensure a new set of features is observed and estimated before the older sets leave the camera's FOV. Additionally, many approaches use known information about features, the desired trajectory, or assume features lie on a plane to develop relationships between poses or features using homography relationships between planes (cf., [45], [46], [49], [50], [52], [54], [55]). In general, requiring a homographic relationship can introduce error in position estimates, especially in environments comprised of multiple objects. Additionally, many of these approaches do not take an observer based approach (cf., [45], [46], [49], [50], [52], [54], [55]); instead, known geometry or direct measurements are used. Furthermore, many approaches apply variations of Kalman filters or other optimization methods are used which may not converge or only converge up to a scale factor (cf., [3]–[8]). Methods that rely on direct measurements or geometry reduce robustness to noise because there is no ability to remove outliers; however, most optimization methods reject outliers and consider noise. Stochastic homography based approaches are developed in [54] and [55] to handle process and measurement noise; however, these approaches will still suffer from the same issues that arise from the planar assumption. For agents to travel over unknown environments, the limitations presented by the planar assumption and those introduced by not using an observer must be removed (i.e., it is not always possible to directly reconstruct features if they are nonplanar and so the structure of each set of features must be learned).

This paper extends [27] and [36] by developing a novel estimator strategy that maintains a continuous estimate of the pose of the camera and estimates the structure of features as they come into the FOV. Additionally, the developed learning strategy allows simulated measurements of features from objects that are no longer in the FOV enabling a continuous estimate of the distance to features with respect to the camera. This approach removes the positive depth constraint required by all previous SfM approaches. Using this estimator, a camera may travel over large distances without

keeping specific features in the FOV for all time and enables an object to permanently leave the FOV after satisfying a dwell-time condition developed using a Lyapunov based stability analysis. The developed observer for each feature's structure is shown to be exponentially stable and after satisfying a minimum dwell-time (i.e., the minimum time a feature must provide feedback by remaining in the FOV), the unmeasurable structure estimation error associated with the feature is shown to be lower than a user-defined threshold. Once all features on an object remaining in the camera's FOV have satisfied this dwell-time condition, they are used to provide position feedback for the camera, ensuring the estimation error of the position remains bounded until that object leaves the camera's FOV. If all of the features on the object leave the camera's FOV before satisfying the minimum dwell-time condition, that object is never used to provide feedback for the position of the camera. The strategy for estimating the position of the camera is similar to the recently developed work in [1] and [2] where a reset map is used when position feedback is available and a predictor is used when no position feedback is available. However, unlike [1] and [2], when feedback of the camera's position is unavailable an observer is used to estimate the position after an object in the FOV satisfies the minimum dwell-time condition. When no object is in the FOV a predictor strategy is used to estimate the position which is shown to remain bounded using a Lyapunov-based analysis. The approach does not require a new object to be in the camera's FOV when another object leaves the camera's FOV. If a recognized landmark enters the camera's FOV (i.e., an object with structure exactly known that can provide position feedback directly (cf., [57]–[60])), a reset map similar to [2] is used given feedback of the position will be directly available. This developed estimator strategy is compared to a predictor-only strategy like those in [1] and [2] in an experiment where a ground vehicle equipped with a single camera travels through an environment where hundreds of objects enter and leave the camera's FOV over a large trajectory, but only a third of those objects satisfy the dwell-time condition.

II. MOTION MODEL USING STATIONARY FEATURES

A. Motion Model for Stationary Features

The following definitions and assumptions are presented to aid in the development of the subsequent observers. The frame at which the current image is taken, denoted by \mathcal{F}_c , has its origin at the principal point of the current image, denoted by c , and basis $\{\underline{x}_c, \underline{y}_c, \underline{z}_c\}$, where \underline{z}_c is aligned with the normal to the image plane, \underline{x}_c is aligned with the horizontal of the image plane (i.e., to the right in the image), and \underline{y}_c is aligned with the vertical of the image plane (i.e., downward in the image). Let $\mathcal{V}_c \subset \mathbb{R}^3$ represent the Euclidean space of the camera's FOV expressed in \mathcal{F}_c .

Assumption 1. The Euclidean space of the camera's FOV, \mathcal{V}_c , is compact and the norm of each point in \mathcal{V}_c is bounded using a known constant $\bar{d} \in \mathbb{R}_{>0}$.

Remark 1. Assumption 1 is necessary for the subsequent development of the maximum and minimum dwell-time con-

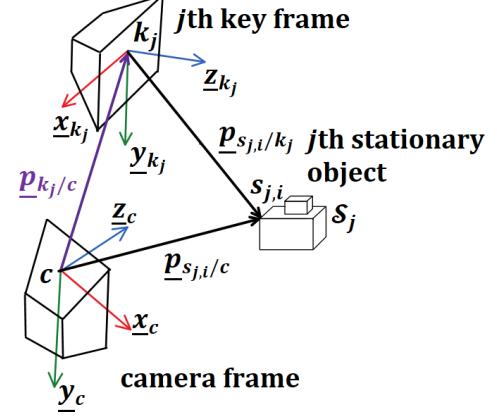


Figure 1. Example geometry for tracking the position of the i th feature in S_j where the key image taken from the top left and the camera traveling downward from the upper left to the lower left while tracking S_j on the right.

ditions and is inherently required when estimating distances to an unknown object's features that may leave the camera's FOV.

Assumption 2. There exists a set of stationary objects, $\{\mathcal{S}_j\}_{j=1}^{\bar{p}_s(t)}$, where $\mathcal{S}_j \subset \mathbb{R}^3$ represents the minimum Euclidean sphere enclosing the j th stationary object expressed in \mathcal{F}_c . Furthermore, there exists a set of trackable features on each stationary object, $\{\mathcal{O}_{s_j}\}_{j=1}^{\bar{p}_s(t)}$, where $\mathcal{O}_{s_j} \triangleq \{s_{j,i}\}_{i=1}^{p_{s_j}}$ is the j th stationary object's feature set, $s_{j,i}$ represents the i th feature on the j th stationary object, and $\bar{p}_{s_j} \in \mathbb{Z}_{\geq 4}$ represents the total number of trackable features on the j th stationary object.

Remark 2. Features on an object in the camera's FOV can be tracked using descriptor and matching techniques such as [9], [61]–[67] or feature extraction and tracking techniques such as [68]–[71].

A key frame is defined as the camera frame at which features are first extracted from an image of an object. For the following development, let $\{\mathcal{F}_{k_j}\}_{j=1}^{\bar{p}_s(t)}$ and $\{\zeta_j^a\}_{j=1}^{\bar{p}_s(t)}$ be a sequence of key frames and times at which key frames are established, respectively, where $\zeta_j^a \in \mathbb{R}_{\geq 0}$ denotes the time, $t \in \mathbb{R}_{\geq 0}$, that the j th object has feedback available (i.e., features are first extracted from the object after entering the FOV establishing the key frame), $\zeta_j^u \in \mathbb{R}_{>\zeta_j^a}$ denotes the time the when feedback for the j th object is unavailable (i.e., the object is no longer tracked because too many features on the object leave the camera's FOV), $\bar{p}_s(t) \in \mathbb{Z}_{>0}$ denotes the total number of key frames, and the j th key frame, denoted by \mathcal{F}_{k_j} , has its origin at the principal point of that image, denoted by k_j , and has the basis $\{\underline{x}_{k_j}, \underline{y}_{k_j}, \underline{z}_{k_j}\}$, which are selected such that \mathcal{F}_{k_j} is established to coincide with \mathcal{F}_c at time $t = \zeta_j^a$.

As illustrated in Figure 1, the position of feature $s_{j,i} \in \mathcal{O}_{s_j}$, can be described as

$$\underline{p}_{s_{j,i}/c}(t) = \underline{p}_{k_j/c}(t) + R_{k_j/c}(t) \underline{p}_{s_{j,i}/k_j}, \quad (1)$$

where $\underline{p}_{s_{j,i}/c}(t) \in \mathbb{R}^3$ is the position of feature $s_{j,i}$ with respect to c expressed in \mathcal{F}_c , $\underline{p}_{k_j/c}(t) \in \mathbb{R}^3$ is the position

of k_j with respect to c expressed in \mathcal{F}_c , $p_{s_{j,i}/k_j}(t) \in \mathbb{R}^3$ is the position of feature $s_{j,i}$ with respect to k_j expressed in \mathcal{F}_{k_j} , and $R_{k_j/c}(t) \in \mathbb{R}^{3 \times 3}$ is the rotation matrix describing the orientation of \mathcal{F}_{k_j} with respect to \mathcal{F}_c . Rearranging (1) gives

$$Y_{s_{j,i}}(t) \begin{bmatrix} d_{s_{j,i}/c}(t) \\ d_{k_j/c}(t) \end{bmatrix} = R_{k_j/c}(t) \underline{u}_{s_{j,i}/k_j} d_{s_{j,i}/k_j}, \quad (2)$$

where $Y_{s_{j,i}}(t) \triangleq [\underline{u}_{s_{j,i}/c}(t) \quad -\underline{u}_{k_j/c}(t)]$, $d_{s_{j,i}/c}(t) \in \mathbb{R}_{>0}$, $\underline{u}_{s_{j,i}/c}(t) \in \mathbb{R}^3$ are the distance and unit vector of feature $s_{j,i}$ with respect to c expressed in \mathcal{F}_c , $d_{k_j/c}(t) \in \mathbb{R}_{>0}$, $\underline{u}_{k_j/c}(t) \in \mathbb{R}^3$ are the distance and unit vector of k_j with respect to c expressed in \mathcal{F}_c , and $d_{s_{j,i}/k_j} \in \mathbb{R}_{>0}$, $\underline{u}_{s_{j,i}/k_j} \in \mathbb{R}^3$ are the distance and unit vector of feature $s_{j,i}$ with respect to k_j expressed in \mathcal{F}_{k_j} .

Since features on the object may leave the camera FOV over time, let $\mathcal{P}_{s_j}(t) \subseteq \mathcal{O}_{s_j}$ represent the remaining set of features in the camera's FOV, specifically, $\mathcal{P}_{s_j}(t) \triangleq \{s_{j,i} \in \mathcal{O}_{s_j} : p_{s_{j,i}/c}(t) \in \mathcal{S}_j \cap \mathcal{V}_c\}$, and let $p_{s_j}(t) \in \mathbb{Z}_{\geq 0}$ represent the number of features in $\mathcal{P}_{s_j}(t)$, and $\mathcal{P}_{s_j}^c(t) \triangleq \mathcal{O}_{s_j} \setminus \mathcal{P}_{s_j}(t)$ represent the compliment of $\mathcal{P}_{s_j}(t)$ (i.e., the features). Let $\zeta_{s_{j,i}}^u \in \mathbb{R}_{\geq \zeta_j^a}$, represent the time the feature indexed by $s_{j,i}$ leaves the FOV, specifically, the time instance when $s_{j,i} \notin \mathcal{P}_{s_j}(t)$, and let $\Delta t_{s_{j,i}}^a \triangleq \zeta_{s_{j,i}}^u - \zeta_j^a$ represent the total time the feature indexed by $s_{j,i}$ was tracked by the camera. Let $\zeta_j^u \triangleq \max \{\zeta_{s_{j,i}}^u\}_{i=1}^{p_{s_j}(t)}$ (i.e., the time the last feature leaves the FOV).

Assumption 3. The camera intrinsic matrix $A \in \mathbb{R}^{3 \times 3}$ is known and invertable [10].

Under Assumptions 2 and 3, the rotation matrix $R_{k_j/c}(t)$ and unit vector $\underline{u}_{k_j/c}(t)$ can be determined from the set of stationary features in $\mathcal{P}_{s_j}(t)$ while $p_{s_j}(t) \geq 4$.¹ Additionally, $\underline{u}_{s_{j,i}/k_j}$ and $\underline{u}_{s_{j,i}/c}(t)$ can be determined from $\underline{u}_{s_{j,i}/k_j} = \frac{A^{-1} p_{s_{j,i}/k_j}}{\|A^{-1} p_{s_{j,i}/k_j}\|}$ and $\underline{u}_{s_{j,i}/c}(t) = \frac{A^{-1} p_{s_{j,i}/c}(t)}{\|A^{-1} p_{s_{j,i}/c}(t)\|}$ where $p_{s_{j,i}/k_j}, p_{s_{j,i}/c}(t) \in \mathbb{R}^3$ are the homogeneous pixel coordinates of feature $s_{j,i}$ in \mathcal{F}_{k_j} and \mathcal{F}_c , respectively. Based on the definition of $Y_{s_{j,i}}(t)$, the minimum eigenvalue of $Y_{s_{j,i}}^T(t) Y_{s_{j,i}}(t)$ is $1 - \|\underline{u}_{k_j/c}(t) \underline{u}_{s_{j,i}/c}(t)\|$.

Assumption 4. The origins k_j and c are not coincident while $t > \zeta_j^a$, specifically, $d_{k_j/c}(t) > \underline{d}_1$ where $\underline{d}_1 \in \mathbb{R}_{>0}$ is a constant.

In practice, the camera can remain stationary after selecting feature points; however, the time ζ_j^a will be delayed until the camera moves. Also, since the camera and object are never coincident, $\underline{d}_2 \in \mathbb{R}_{>0}$ is selected such that $d_{s_{j,i}/c}(t) > \underline{d}_2$ given there is always some minimum distance for image formation based on the focal length. Additionally, for each feature $s_{j,i} \in \mathcal{P}_{s_j}(t)$, $d_{s_{j,i}/c}(t) \leq \bar{d}$ under Assumption 1, implying $\underline{d}_2 \leq d_{s_{j,i}/c}(t) \leq \bar{d}$ for each feature $s_{j,i} \in \mathcal{P}_{s_j}(t)$.

Given Assumption 4, when the motion of the camera is not parallel to the direction to a feature,

¹See [9], [10], and [72] for examples on calculating the rotation and normalized translation from planar and nonplanar features.

$\left(1 - \|\underline{u}_{k_j/c}^T(t) \underline{u}_{s_{j,i}/c}(t)\|\right) > \lambda_a$, where $\lambda_a \in (0, 1)$ is a selected constant. Let $\{\sigma_{s_{j,i}}(t)\}_{s_{j,i} \in \mathcal{P}_{s_j}(t)}$ be the set of switching signals for the features in $\mathcal{P}_{s_j}(t)$, where $\sigma_{s_{j,i}}(t) \in \{u, a\}$ indicates whether $\left(1 - \|\underline{u}_{k_j/c}^T(t) \underline{u}_{s_{j,i}/c}(t)\|\right) \leq \lambda_a$ or $\left(1 - \|\underline{u}_{k_j/c}^T(t) \underline{u}_{s_{j,i}/c}(t)\|\right) > \lambda_a$, respectively. When $\sigma_{s_{j,i}}(t) = a$ (i.e., not parallel motion), $Y_{s_{j,i}}(t)$ is full column rank and (2) can be written

$$\begin{bmatrix} d_{s_{j,i}/c}(t) \\ d_{k_j/c}(t) \end{bmatrix} = \psi_{s_{j,i}}^a(t) d_{s_{j,i}/k_j}, \quad (3)$$

where

$$\psi_{s_{j,i}}^a(t) \triangleq \left(Y_{s_{j,i}}^T(t) Y_{s_{j,i}}(t)\right)^{-1} Y_{s_{j,i}}^T(t) R_{k_j/c}(t) \underline{u}_{s_{j,i}/k_j}$$

is measurable based on Assumptions 2 and 3. When $\sigma_{s_{j,i}}(t) = u$ (i.e., parallel motion), $Y_{s_{j,i}}(t)$ can not be guaranteed to have full column rank but (2) can always be written as $d_{s_{j,i}/c}(t) = \underline{u}_{s_{j,i}/c}^T(t) \underline{u}_{k_j/c}(t) d_{k_j/c}(t) + \underline{u}_{s_{j,i}/c}^T(t) R_{k_j/c}(t) \underline{u}_{s_{j,i}/k_j} d_{s_{j,i}/k_j}$ yielding

$$d_{s_{j,i}/c}(t) = \psi_{s_{j,i}}^u(t) \begin{bmatrix} d_{k_j/c}(t) \\ d_{s_{j,i}/k_j} \end{bmatrix}, \quad (4)$$

where

$$\psi_{s_{j,i}}^u(t) \triangleq \left[\underline{u}_{s_{j,i}/c}^T(t) \underline{u}_{k_j/c}(t) \quad \underline{u}_{s_{j,i}/c}^T(t) R_{k_j/c}(t) \underline{u}_{s_{j,i}/k_j}\right].$$

Assumption 5. The camera linear and angular velocities, $\underline{v}_c(t), \underline{\omega}_c(t) \in \mathbb{R}^3$, are measurable and expressed in \mathcal{F}_c and are upper bounded as $\|\underline{v}_c(t)\| \leq \bar{v}_c$ and $\|\underline{\omega}_c(t)\| \leq \bar{\omega}_c$, where $\bar{v}_c, \bar{\omega}_c \in \mathbb{R}_{>0}$ are known constants.

Using Assumption 5 and while $p_{s_j}(t) \geq 4$, the time derivatives of the unknown distances $d_{s_{j,i}/c}(t)$, $d_{k_j/c}(t)$, and d_{s_i/k_j} are measurable for $s_{j,i} \in \mathcal{P}_{s_j}(t)$ as

$$\frac{d}{dt}(d_{s_{j,i}/c}(t)) = -\underline{u}_{s_{j,i}/c}^T(t) \underline{v}_c(t), \quad (5)$$

$$\frac{d}{dt}(d_{k_j/c}(t)) = -\underline{u}_{k_j/c}^T(t) \underline{v}_c(t), \quad (6)$$

and

$$\frac{d}{dt}(d_{s_i/k_j}) = 0. \quad (7)$$

III. LEARNING FEATURE STRUCTURE

In general, there is no relationship between any two objects that may be exploited to immediately estimate the structure of the i th feature on the j th object (i.e., $s_{j,i}$) when $t = \zeta_j^a$, since the objects are unknown and there may only be one object in the FOV. While the i th feature on the j th object has the eigenvalue condition satisfied (i.e., $\sigma_{s_{j,i}}(t) = a$), an approach motivated by the development in [27] and [36] is used to learn a constant unknown distance, $d_{s_{j,i}/k_j}$. Specifically, (5) and (6) are integrated over a time window $\varsigma \in \mathbb{R}_{>0}$ yielding

$$\begin{bmatrix} d_{s_{j,i}/c}(t) \\ d_{k_j/c}(t) \end{bmatrix} - \begin{bmatrix} d_{s_{j,i}/c}(t-\varsigma) \\ d_{k_j/c}(t-\varsigma) \end{bmatrix} = - \int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{s_{j,i}/c}^T(\iota) \\ \underline{u}_{k_j/c}^T(\iota) \end{bmatrix} \underline{v}_c(\iota) d\iota, \quad t > \varsigma, \quad (8)$$

where ς may be constant in size or change over time. As described in (5) and (6), while $s_{j,i} \in \mathcal{P}_{s_j}(t)$, $-\int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{s_{j,i}/c}^T(t) \\ \underline{u}_{k_j/c}^T(t) \end{bmatrix} \underline{v}_c(\iota) d\iota$ is known, but $\begin{bmatrix} d_{s_{j,i}/c}(t) \\ d_{k_j/c}(t) \end{bmatrix}$ and $\begin{bmatrix} d_{s_{j,i}/c}(t-\varsigma) \\ d_{k_j/c}(t-\varsigma) \end{bmatrix}$ are unknown indicating the left side of the equality in (8) is unknown. However, when $\sigma_{s_{j,i}}(t) = a$, the relationship in (3) may be utilized in (8) yielding

$$\mathcal{Y}_{s_{j,i}}(t) d_{s_{j,i}/k_j} = \mathcal{U}_{s_{j,i}}(t), \quad t > \zeta_j^a, \quad (9)$$

where $\mathcal{Y}_{s_{j,i}}(t), \mathcal{U}_{s_{j,i}}(t) \in \mathbb{R}^2$ are defined as

$$\begin{aligned} \mathcal{Y}_{s_{j,i}}(t) &\triangleq \begin{cases} \left(\psi_{s_{j,i}}^a(t) - \psi_{s_{j,i}}^a(\zeta_{s_{j,i}}^{a_l}) \right), & t \in \left(\zeta_{s_{j,i}}^{a_l}, \zeta_{s_{j,i}}^{a_l} + \varsigma \right], \\ \left(\psi_{s_{j,i}}^a(t) - \psi_{s_{j,i}}^a(t-\varsigma) \right), & t \in \left(\zeta_{s_{j,i}}^{a_l} + \varsigma, \zeta_{s_{j,i}}^{u_l} \right], \\ 0_{2 \times 1}, & t \in \left(\zeta_{s_{j,i}}^{u_l}, \zeta_{s_{j,i}}^{a_{l+1}} \right], \end{cases} \\ \mathcal{U}_{s_{j,i}}(t) &\triangleq \begin{cases} - \int_{\zeta_{s_{j,i}}^{a_l}}^t \begin{bmatrix} \underline{u}_{s_{j,i}/c}^T(\iota) \\ \underline{u}_{k_j/c}^T(\iota) \end{bmatrix} \underline{v}_c(\iota) d\iota, & t \in \left(\zeta_{s_{j,i}}^{a_l}, \zeta_{s_{j,i}}^{a_l} + \varsigma \right], \\ - \int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{s_{j,i}/c}^T(\iota) \\ \underline{u}_{k_j/c}^T(\iota) \end{bmatrix} \underline{v}_c(\iota) d\iota, & t \in \left(\zeta_{s_{j,i}}^{a_l} + \varsigma, \zeta_{s_{j,i}}^{u_l} \right], \\ 0_{2 \times 1}, & t \in \left(\zeta_{s_{j,i}}^{u_l}, \zeta_{s_{j,i}}^{a_{l+1}} \right], \end{cases} \end{aligned}$$

and $\zeta_{s_{j,i}}^{a_l}, \zeta_{s_{j,i}}^{u_l} \in [\zeta_j^a, \zeta_{s_{j,i}}^u]$ represent time instances when $\sigma_{s_{j,i}}(t) = a$ and $\sigma_{s_{j,i}}(t) = u$, respectively, and $l \in \mathbb{Z}_{>0}$, represents the index corresponding to each switch for feature $s_{j,i}$. Multiplying both sides of (9) by $\mathcal{Y}_{s_{j,i}}^T(t)$ yields

$$\mathcal{Y}_{s_{j,i}}^T(t) \mathcal{Y}_{s_{j,i}}(t) d_{s_{j,i}/k_j} = \mathcal{Y}_{s_{j,i}}^T(t) \mathcal{U}_{s_{j,i}}(t). \quad (10)$$

In general, $\mathcal{Y}_{s_{j,i}}(t)$ will not full column rank while $\sigma_{s_{j,i}}(t) = a$ (e.g. when the camera is stationary implying $\mathcal{Y}_{s_{j,i}}^T(t) \mathcal{Y}_{s_{j,i}}(t) \geq 0$) and cannot be determined while $\sigma_{s_{j,i}}(t) = u$. However, the equality in (10) may be evaluated at any instance in time and summed together (i.e., history stacks) yielding

$$\Sigma_{\mathcal{Y}_{s_{j,i}}} d_{s_{j,i}/k_j} = \Sigma_{\mathcal{U}_{s_{j,i}}}, \quad (11)$$

where $\Sigma_{\mathcal{Y}_{s_{j,i}}} \triangleq \sum_{h=1}^N \mathcal{Y}_{s_{j,i}}^T(t_h) \mathcal{Y}_{s_{j,i}}(t_h)$, $\Sigma_{\mathcal{U}_{s_{j,i}}} \triangleq \sum_{h=1}^N \mathcal{Y}_{s_{j,i}}^T(t_h) \mathcal{U}_{s_{j,i}}(t_h)$, $t_h \in (\zeta_j^a, \zeta_{s_{j,i}}^u]$, and $N \in \mathbb{Z}_{>1}$.

Assumption 6. The camera motion occurs so there exists a set of features $\mathcal{A}_{s_j}(t) \subseteq \mathcal{O}_{s_j}$, constant $\lambda_\tau \in \mathbb{R}_{>0}$, and a set of times $\tau_j \triangleq \{\tau_{s_{j,i}}\}_{s_{j,i} \in \mathcal{A}_{s_j}}$, such that for all time $t > \tau_{s_{j,i}}$, $\lambda_{\min}\{\Sigma_{\mathcal{Y}_{s_{j,i}}}\} > \lambda_\tau$, where $\tau_{s_{j,i}} \in (\zeta_j^a, \zeta_{s_{j,i}}^u)$ and $\lambda_{\min}\{\cdot\}$ is the minimum eigenvalue of $\{\cdot\}$. Let $\mathcal{A}_{s_j}^c(t) \triangleq \mathcal{O}_{s_j} \setminus \mathcal{A}_{s_j}(t)$. Furthermore, $a_{s_j}(\zeta_j^u) \geq 4$, where $a_{s_j}(t) \in \mathbb{Z}_{\geq 0}$ represents the number of features in $\mathcal{A}_{s_j}(t)$.²

Learning the subset in $\mathcal{A}_{s_j}(t)$ is less restrictive than assuming all of the features in \mathcal{O}_{s_j} are learned because there is

²See [73], [74] for some examples of methods for selecting data to satisfy the assumption.

no guarantee that the motion of the camera will be sufficient before every feature leaves the camera's FOV permanently. Camera motion in Assumption 6 can be verified online and is heuristically easy to satisfy because it only requires a finite collection of sufficiently exciting $\mathcal{Y}_{s_{j,i}}(t)$ and $\mathcal{U}_{s_{j,i}}(t)$ to yield $\lambda_{\min}\{\Sigma_{\mathcal{Y}_{s_{j,i}}}\} > \lambda_\tau$. The times in τ_j are unknown; however, they can be determined online by checking the minimum eigenvalue of $\Sigma_{\mathcal{Y}_{s_{j,i}}}$ for each feature.

If motion occurs as discussed in Assumption 6, the constant unknown distance, $d_{s_{j,i}/k_j}$, can be determined for feature $s_{j,i} \in \mathcal{A}_{s_j}(t)$ from (11) yielding

$$d_{s_{j,i}/k_j} = \mathcal{X}_{s_{j,i}}, \quad s_{j,i} \in \mathcal{A}_{s_j}(t), \quad (12)$$

where $\mathcal{X}_{s_{j,i}} \triangleq \Sigma_{\mathcal{Y}_{s_{j,i}}}^{-1} \Sigma_{\mathcal{U}_{s_{j,i}}}$, $s_{j,i} \in \mathcal{A}_{s_j}(t)$. Substituting (12) into (2) yields

$$Y_{s_{j,i}}(t) \begin{bmatrix} d_{s_{j,i}/c}(t) \\ d_{k_j/c}(t) \end{bmatrix} = R_{k_j/c}(t) \underline{u}_{s_{j,i}/k_j} \mathcal{X}_{s_{j,i}}, \quad s_{j,i} \in \mathcal{A}_{s_j}(t). \quad (13)$$

Since there will always be a delay before $\mathcal{X}_{s_{j,i}}$ is determined for $s_{j,i} \in \mathcal{A}_{s_j}(t)$, an additional relationship is developed in an effort to provide feedback based on the rate of change of the direction to the feature, motivated by the development in [25]. Specifically, the time rate of change of $\underline{u}_{s_{j,i}/c}(t)$ is approximated and used to provide feedback. Taking the time derivative of $\underline{u}_{s_{j,i}/c}(t)$ yields

$$\begin{aligned} \frac{d}{dt}(\underline{u}_{s_{j,i}/c}(t)) &= -\underline{\omega}_c^\times(t) \underline{u}_{s_{j,i}/c}(t) \\ &\quad + \frac{1}{d_{s_{j,i}/c}(t)} (\underline{u}_{s_{j,i}/c}(t) \underline{u}_{s_{j,i}/c}^T(t) - I_{3 \times 3}) \underline{v}_c(t), \end{aligned} \quad (14)$$

implying

$$\xi_{s_{j,i}}(t) d_{s_{j,i}/c}(t) = \rho_{s_{j,i}}(t), \quad (15)$$

where $\xi_{s_{j,i}}(t) \triangleq \left(\frac{d}{dt}(\underline{u}_{s_{j,i}/c}(t)) + \underline{\omega}_c^\times(t) \underline{u}_{s_{j,i}/c}(t) \right)$, $\rho_{s_{j,i}}(t) \triangleq (\underline{u}_{s_{j,i}/c}(t) \underline{u}_{s_{j,i}/c}^T(t) - I_{3 \times 3}) \underline{v}_c(t)$, $\underline{\omega}_c^\times(t) \triangleq \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}$, and $I_{3 \times 3} \triangleq \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Substituting (4) into (15) yields

$$\xi_{s_{j,i}}(t) \psi_{s_{j,i}}^u(t) \begin{bmatrix} d_{k_j/c}(t) \\ d_{s_{j,i}/k_j} \end{bmatrix} = \rho_{s_{j,i}}(t). \quad (16)$$

Let a composite signal $\eta_{s_{j,i}}(t) \in \mathbb{R}^3$ be defined as $\eta_{s_{j,i}}(t) \triangleq \begin{bmatrix} d_{s_{j,i}/c}(t) \\ d_{k_j/c}(t) \\ d_{s_{j,i}/k_j} \end{bmatrix}$. Combining (12) and (13) yields

$$Y_{\mathcal{X}_{s_{j,i}}}(t) \eta_{s_{j,i}}(t) = u_{\mathcal{X}_{s_{j,i}}}(t), \quad (17)$$

where $Y_{\mathcal{X}_{s_{j,i}}}(t) \triangleq \begin{bmatrix} Y_{s_{j,i}}(t) & 0_{3 \times 1} \\ 0_{1 \times 2} & 1 \end{bmatrix}$ and $u_{\mathcal{X}_{s_{j,i}}}(t) \triangleq \mathcal{X}_{s_{j,i}} \begin{bmatrix} R_{k_j/c}(t) \underline{u}_{s_{j,i}/k_j} \\ 1 \end{bmatrix}$, and combining (15) and (16) yields

$$Y_{\xi_{s_{j,i}}}(t) \eta_{s_{j,i}}(t) = u_{\xi_{s_{j,i}}}(t), \quad (18)$$

$$\text{where } Y_{\xi_{s_{j,i}}}(t) \triangleq \begin{bmatrix} \xi_{s_{j,i}}(t) & 0_{3 \times 2} \\ 0_{3 \times 1} & \xi_{s_{j,i}}(t) \psi_{s_{j,i}}^u(t) \end{bmatrix} \quad \text{and} \\ u_{\xi_{s_{j,i}}}(t) \triangleq \begin{bmatrix} \rho_{s_{j,i}}(t) \\ \rho_{s_{j,i}}(t) \end{bmatrix}.$$

IV. FEATURE OBSERVER DESIGN WITHOUT OBJECT RETURN

The estimation errors for feature $s_{j,i} \in \mathcal{O}_{s_j}$, $\tilde{d}_{s_{j,i}/c}(t), \tilde{d}_{k_{j,i}/c}(t), \tilde{d}_{s_{j,i}/k_j}(t) \in \mathbb{R}$, are defined as

$$\tilde{d}_{s_{j,i}/c}(t) \triangleq d_{s_{j,i}/c}(t) - \hat{d}_{s_{j,i}/c}(t), \quad (19)$$

$$\tilde{d}_{k_{j,i}/c}(t) \triangleq d_{k_{j,i}/c}(t) - \hat{d}_{k_{j,i}/c}(t), \quad (20)$$

and

$$\tilde{d}_{s_{j,i}/k_j}(t) \triangleq d_{s_{j,i}/k_j}(t) - \hat{d}_{s_{j,i}/k_j}(t), \quad (21)$$

where $\hat{d}_{s_{j,i}/c}(t), \hat{d}_{s_{j,i}/k_j}(t) \in \mathbb{R}$ are the estimates of $d_{s_{j,i}/c}(t)$ and $d_{s_{j,i}/k_j}(t)$, respectively, and $\hat{d}_{k_{j,i}/c}(t)$ is the estimate of $d_{k_{j,i}/c}(t)$ by feature $s_{j,i}$. The combined error for feature $s_{j,i}$ is quantified using (19)-(21) as

$$\tilde{\eta}_{s_{j,i}}(t) \triangleq \eta_{s_{j,i}}(t) - \hat{\eta}_{s_{j,i}}(t), \quad (22)$$

where $\hat{\eta}_{s_{j,i}}(t) \triangleq \begin{bmatrix} \hat{d}_{s_{j,i}/c}(t) \\ \hat{d}_{k_{j,i}/c}(t) \\ \hat{d}_{s_{j,i}/k_j}(t) \end{bmatrix}$ is the estimate of $\eta_{s_{j,i}}(t)$ implying $\tilde{\eta}_{s_{j,i}}(t) \triangleq \begin{bmatrix} \tilde{d}_{s_{j,i}/c}(t) \\ \tilde{d}_{k_{j,i}/c}(t) \\ \tilde{d}_{s_{j,i}/k_j}(t) \end{bmatrix}$.

A. Feature Observer Design

If the j th object will never return to the camera's FOV, no updates can be guaranteed after $s_{j,i} \in \mathcal{P}_{s_j}^c(t)$. In this case, each feature is designed as though it would remain in the FOV and the last known estimate is used after the feature leaves the FOV (i.e., a zero-order hold). Motivated by the subsequent analysis, the estimator update law for $\hat{\eta}_{s_{j,i}}(t)$ is defined as

$$\frac{d}{dt}(\hat{\eta}_{s_{j,i}}(t)) \triangleq \begin{cases} 0_{3 \times 1}, & s_{j,i} \in \mathcal{P}_{s_j}^c(t), \\ \text{proj}\left(\mu_{\xi_{s_{j,i}}}(t)\right), & s_{j,i} \in \mathcal{P}_{s_j}(t), \\ \text{proj}\left(\mu_{\xi_{s_{j,i}}}(t) + \mu_{\mathcal{X}_{s_{j,i}}}(t)\right), & s_{j,i} \in \mathcal{A}_{s_j} \cap \mathcal{P}_{s_j}(t), \end{cases} \quad (23)$$

where $\mu_{\xi_{s_{j,i}}}(t) \triangleq \begin{bmatrix} -\underline{u}_{s_{j,i}/c}^T(t) \underline{v}_c(t) \\ -\underline{u}_{k_{j,i}/c}^T(t) \underline{v}_c(t) \\ 0 \end{bmatrix} + K_\xi Y_{\xi_{s_{j,i}}}^T(t) u_{\xi_{s_{j,i}}}(t) - K_\xi Y_{\xi_{s_{j,i}}}^T(t) Y_{\xi_{s_{j,i}}}(t) \hat{\eta}_{s_{j,i}}(t)$, $\mu_{\mathcal{X}_{s_{j,i}}}(t) \triangleq K_\mathcal{X} Y_{\mathcal{X}_{s_{j,i}}}^T(t) u_{\mathcal{X}_{s_{j,i}}}(t) - K_\mathcal{X} Y_{\mathcal{X}_{s_{j,i}}}^T(t) Y_{\mathcal{X}_{s_{j,i}}}(t) \hat{\eta}_{s_{j,i}}(t)$, and $K_\xi, K_\mathcal{X} \in \mathbb{R}^{3 \times 3}$ are positive definite gain matrices, and $\text{proj}(\cdot)$ is a projection operator to bound $\underline{d}_2 \leq \hat{d}_{s_{j,i}/c}(t) \leq \bar{d}$, $\underline{d}_2 \leq \hat{d}_{s_{j,i}/k_j}(t) \leq \bar{d}$, and $\underline{d}_1 \leq \hat{d}_{k_{j,i}/c}(t)$.³

³See [32, Appendix E] or [33, Remark 3.7] for examples on implementing a smooth projection operator

Taking the time derivative of (22), substituting (5)-(7), (17), (18), (22), and (23), and simplifying yields

$$\frac{d}{dt}(\tilde{\eta}_{s_{j,i}}(t)) = \begin{cases} \begin{bmatrix} -\underline{u}_{s_{j,i}/c}^T(t) \underline{v}_c(t) \\ -\underline{u}_{k_{j,i}/c}^T(t) \underline{v}_c(t) \\ 0 \end{bmatrix}, & s_{j,i} \in \mathcal{P}_{s_j}^c(t), \\ -\Psi_{\xi_{s_{j,i}}}(t) \tilde{\eta}_{s_{j,i}}(t), & s_{j,i} \in \mathcal{P}_{s_j}(t), \\ -\Psi_{s_{j,i}}(t) \tilde{\eta}_{s_{j,i}}(t), & s_{j,i} \in \mathcal{A}_{s_j}(t) \cap \mathcal{P}_{s_j}(t), \end{cases} \quad (24)$$

where $\Psi_{\xi_{s_{j,i}}}(t) \triangleq \Psi_{\xi_{s_{j,i}}}(t) + \Psi_{\mathcal{X}_{s_{j,i}}}(t)$, $\Psi_{\xi_{s_{j,i}}}(t) \triangleq K_\xi Y_{\xi_{s_{j,i}}}^T(t) Y_{\xi_{s_{j,i}}}(t)$, and $\Psi_{\mathcal{X}_{s_{j,i}}}(t) \triangleq K_\mathcal{X} Y_{\mathcal{X}_{s_{j,i}}}^T(t) Y_{\mathcal{X}_{s_{j,i}}}(t)$. While feature $s_{j,i} \in \mathcal{P}_{s_j}(t)$, there may be a set of times where $\Psi_{\xi_{s_{j,i}}}(t)$ can improve the estimate if a PE assumption is satisfied. Let $\mathcal{B}_{s_j}(t) \triangleq \{s_{j,i} \in \mathcal{A}_{s_j}(t) \cap \mathcal{P}_{s_j}(t) : \sigma_{s_{j,i}}(t) = a\}$ and $\mathcal{B}_{s_j}^c(t) \triangleq \mathcal{O}_{s_j} \setminus \mathcal{B}_{s_j}(t)$. If feature $s_{j,i} \in \mathcal{B}_{s_j}(t)$, $\Psi_{s_{j,i}}(t) > 0$ and $\lambda_{\min}\{\Psi_{s_{j,i}}(t)\} > \lambda_a$; however, if feature $s_{j,i} \in \mathcal{A}_{s_j}(t) \cap \mathcal{P}_{s_j}(t) \cap \mathcal{B}_{s_j}^c(t)$, $\Psi_{s_{j,i}}(t) \geq 0$ given $Y_{s_{j,i}}^T(t) Y_{s_{j,i}}(t) \geq 0$. After feature $s_{j,i} \in \mathcal{P}_{s_j}^c(t)$, the object never returns to the FOV and the error will grow given no update is available.

B. Observer Design Stability Analysis

To facilitate the subsequent development, let $\mathcal{L}_{s_{j,i}} \triangleq \{l \in \mathbb{Z}_{>0} : \zeta_{s_{j,i}}^{a_l} > \tau_{s_{j,i}}\}$ and $\Delta\zeta_{s_{j,i}}^{a_l} \triangleq \zeta_{s_{j,i}}^{u_l} - \zeta_{s_{j,i}}^{a_l}$. Let $V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t)) : \mathbb{R}^3 \rightarrow \mathbb{R}$ be a candidate Lyapunov function defined as

$$V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t)) \triangleq \frac{1}{2} \tilde{\eta}_{s_{j,i}}^T(t) \tilde{\eta}_{s_{j,i}}(t), \quad (25)$$

which can be bounded as $\frac{1}{2} \|\tilde{\eta}_{s_{j,i}}(t)\|^2 \leq V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t)) \leq \frac{1}{2} \|\tilde{\eta}_{s_{j,i}}(t)\|^2$.

Lemma 1. *The observer update law defined in (23) ensures the estimation error $\tilde{\eta}_{s_{j,i}}(t)$ is bounded for the feature $s_{j,i} \in \mathcal{P}_{s_j}(t)$ in the sense that*

$$\|\tilde{\eta}_{s_{j,i}}(t)\| \leq \|\tilde{\eta}_{s_{j,i}}(\zeta_j^a)\|, \quad s_{j,i} \in \mathcal{P}_{s_j}(t). \quad (26)$$

Proof: Taking the time derivative of (25), substituting (24) for the case when feature $s_{j,i} \in \mathcal{P}_{s_j}(t)$ and using $\lambda_{\min}\{\Psi_{\xi_{s_{j,i}}}(t)\} \geq 0$ for $s_{j,i} \in \mathcal{P}_{s_j}(t)$ yields

$$\frac{d}{dt}(V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t))) \leq 0. \quad (27)$$

Invoking [75, Theorem 8.4] on (27) yields $\|\tilde{\eta}_{s_{j,i}}(t)\|^2 \leq \|\tilde{\eta}_{s_{j,i}}(\zeta_j^a)\|^2$ and taking the square root yields (26). ■

Lemma 2. *The observer update law defined in (23) ensures the estimation error $\tilde{\eta}_{s_{j,i}}(t)$ is bounded for feature $s_{j,i} \in \mathcal{A}_{s_j}(t) \cap \mathcal{P}_{s_j}(t) \cap \mathcal{B}_{s_j}^c(t)$ in the sense that*

$$\|\tilde{\eta}_{s_{j,i}}(t)\| \leq \|\tilde{\eta}_{s_{j,i}}(\zeta_{s_{j,i}}^{u_l})\|, \quad s_{j,i} \in \mathcal{A}_{s_j}(t) \cap \mathcal{P}_{s_j}(t) \cap \mathcal{B}_{s_j}^c(t). \quad (28)$$

Proof: Taking the time derivative of (25), substituting (24) for the case when feature $s_{j,i} \in \mathcal{A}_{s_j}(t) \cap \mathcal{P}_{s_j}(t)$ and using

$\lambda_{\min}\{\Psi_{s_{j,i}}(t)\} \geq 0$ for $s_{j,i} \in \mathcal{A}_{s_j}(t) \cap \mathcal{P}_{s_j}(t) \cap \mathcal{B}_{s_j}^c(t)$ yields

$$\frac{d}{dt}(V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t))) \leq 0. \quad (29)$$

Invoking [75, Theorem 8.4] on (29) yields $\|\tilde{\eta}_{s_{j,i}}(t)\|^2 \leq \|\tilde{\eta}_{s_{j,i}}(\zeta_{s_{j,i}}^{u_l})\|^2$ and taking the square root yields (28). ■

Lemma 3. *The observer update law defined in (23) ensures the estimation error $\tilde{\eta}_{s_{j,i}}(t)$ is exponentially converging for feature $s_{j,i} \in \mathcal{B}_{s_j}(t)$ in the sense that*

$$\|\tilde{\eta}_{s_{j,i}}(t)\| \leq \|\tilde{\eta}_{s_{j,i}}(\zeta_{s_{j,i}}^{a_l})\| \exp(-\beta(t - \zeta_{s_{j,i}}^{a_l})), \quad s_{j,i} \in \mathcal{B}_{s_j}(t). \quad (30)$$

Proof: Taking the time derivative of (25), substituting (24) for the case when feature $s_{j,i} \in \mathcal{A}_{s_j}(t) \cap \mathcal{P}_{s_j}(t)$, and using $\lambda_{\min}\{\Psi_{s_{j,i}}(t)\} > \lambda_a$ for $s_{j,i} \in \mathcal{B}_{s_j}(t)$ yields

$$\frac{d}{dt}(V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t))) \leq -2\beta V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t)), \quad (31)$$

where $\beta \triangleq \lambda_a \lambda_{\min}\{K_X\}$. Invoking [75, Theorem 4.10] on (31) yields $\|\tilde{\eta}_{s_{j,i}}(t)\|^2 \leq \|\tilde{\eta}_{s_{j,i}}(\zeta_{s_{j,i}}^{a_l})\|^2 \exp(-2\beta(t - \zeta_{s_{j,i}}^{a_l}))$ and taking the square root yields (30). ■

Theorem 1. *When feature $s_{j,i} \in \mathcal{A}_{s_j}(t) \cap \mathcal{P}_{s_j}(t)$ leaves the FOV, the switched system defined by $\sigma_{s_{j,i}}(t)$ and the observer update law defined in (23) is globally uniformly ultimately bounded (GUUB) as*

$$\|\tilde{\eta}_{s_{j,i}}(\zeta_{s_{j,i}}^u)\| \leq \|\tilde{\eta}_{s_{j,i}}(\zeta_{s_{j,i}}^a)\| \exp\left(-\beta \sum_{l \in \mathcal{L}_{s_{j,i}}} \Delta \zeta_{s_{j,i}}^{a_l}\right). \quad (32)$$

Proof: Using the bounds in (26), (28), and (30) implies $\|\tilde{\eta}_{s_{j,i}}(\zeta_{s_{j,i}}^{u_l})\| \leq \|\tilde{\eta}_{s_{j,i}}(\zeta_{s_{j,i}}^{a_l})\| \exp(-\beta \Delta \zeta_{s_{j,i}}^{a_l})$ and $\|\tilde{\eta}_{s_{j,i}}(\zeta_{s_{j,i}}^{u_{l+1}})\| \leq \|\tilde{\eta}_{s_{j,i}}(\zeta_{s_{j,i}}^{u_l})\| \exp(-\beta \Delta \zeta_{s_{j,i}}^{a_{l+1}})$. Substituting the first inequality into the second, and using the relationship for all $l \in \mathcal{L}_{s_{j,i}}$ leads to (32). ■

As shown in (32), the final error when a feature leaves is bounded; however, once $s_{j,i} \in \mathcal{P}_{s_j}^c(t)$, the estimation errors in $\tilde{d}_{s_{j,i}/c}(t)$ and $\tilde{d}_{k_{j,i}/c}(t)$ will diverge given no observations are made. For example, using $\frac{d}{dt}(V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t))) = \tilde{\eta}_{s_{j,i}}^T(t) \frac{d}{dt}(\tilde{\eta}_{s_{j,i}}(t))$ and substituting (22) and (24) for the case when feature $s_{j,i} \in \mathcal{P}_{s_j}^c(t)$ implies $\frac{d}{dt}(V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t))) \leq \|\tilde{d}_{s_{j,i}/c}(t)\| \|v_c(t)\| + \|\tilde{d}_{k_{j,i}/c}(t)\| \|v_c(t)\|$, implying the error grows. In applications where it is not possible to return to an object, this growth cannot be compensated for, implying the estimator for feature $s_{j,i}$ is not continued after $s_{j,i} \in \mathcal{P}_{s_j}^c(t)$.

V. FEATURE OBSERVER DESIGN WITH OBJECT RETURN

As discussed in recent work such as [1] and [2], the objective of exploring unknown environments where feedback is unavailable requires an agent to return to regions where feedback is available to compensate for error growth. This return will enable the ability to reduce the ultimate bound of the error described in (32) and compensate for the error growth through the development of dwell-time conditions.

A. Feature Predictor Design

As shown in (24), if feedback is unavailable (i.e., $s_{j,i} \in \mathcal{P}_{s_j}^c(t)$), $\underline{u}_{s_{j,i}/c}(t)$ is unknown. Therefore, a predictor is designed to estimate $\underline{u}_{s_{j,i}/c}(t)$ as

$$\frac{d}{dt}(\hat{u}_{s_{j,i}/c}(t)) \triangleq \mu_{\underline{u}_{s_{j,i}/c}}(t), \quad s_{j,i} \in \mathcal{P}_{s_j}^c(t), \quad (33)$$

where $\mu_{\underline{u}_{s_{j,i}/c}}(t) \triangleq -\omega_c^\times(t) \hat{u}_{s_{j,i}/c}(t) + \frac{1}{\tilde{d}_{s_{j,i}/c}(t)} (\hat{u}_{s_{j,i}/c}(t) \hat{u}_{s_{j,i}/c}^T(t) - I_{3 \times 3}) v_c(t)$. While $s_{j,i} \in \mathcal{P}_{s_j}^c(t)$, a reset map (cf., [1] and [2]) is used to set $\hat{u}_{s_{j,i}/c}(t) \rightarrow \underline{u}_{s_{j,i}/c}(t)$. Also, since $\underline{u}_{s_{j,i}/c}(t)$ is a unit vector, $\|\hat{u}_{s_{j,i}/c}(t)\| = 1$. Let the predictor error for $\underline{u}_{s_{j,i}/c}(t)$ be quantified as

$$\tilde{u}_{s_{j,i}/c}(t) \triangleq \underline{u}_{s_{j,i}/c}(t) - \hat{u}_{s_{j,i}/c}(t), \quad (34)$$

where $\|\tilde{u}_{s_{j,i}/c}(t)\| \leq 2$.

Since the time derivative of $\underline{u}_{k_j/c}(t)$ is

$$\begin{aligned} \frac{d}{dt}(\underline{u}_{k_j/c}(t)) &= -\omega_c^\times(t) \underline{u}_{k_j/c}(t) \\ &\quad + \frac{1}{\tilde{d}_{k_j/c}(t)} (\underline{u}_{k_j/c}(t) \underline{u}_{k_j/c}^T(t) - I_{3 \times 3}) v_c(t), \end{aligned} \quad (35)$$

a predictor similar to (33) is designed to estimate $\underline{u}_{k_j/c}(t)$ as

$$\frac{d}{dt}(\hat{u}_{k_j/c}(t)) \triangleq \mu_{\underline{u}_{k_j/c}}(t), \quad s_{j,i} \in \mathcal{P}_{s_j}^c(t), \quad (36)$$

where $\mu_{\underline{u}_{k_j/c}}(t) \triangleq -\omega_c^\times(t) \hat{u}_{k_j/c}(t) + \frac{1}{\tilde{d}_{k_j/c}(t)} (\hat{u}_{k_j/c}(t) \hat{u}_{k_j/c}^T(t) - I_{3 \times 3}) v_c(t)$. Additionally, while $s_{j,i} \in \mathcal{P}_{s_j}^c(t)$, a reset map is used to set $\hat{u}_{k_j/c}(t) \rightarrow \underline{u}_{k_j/c}(t)$. Also, since $\underline{u}_{k_j/c}(t)$ is a unit vector, $\|\hat{u}_{k_j/c}(t)\| = 1$. Let the predictor error for $\underline{u}_{k_j/c}(t)$ be quantified as

$$\tilde{u}_{k_j/c}(t) \triangleq \underline{u}_{k_j/c}(t) - \hat{u}_{k_j/c}(t) \quad (37)$$

where $\|\tilde{u}_{k_j/c}(t)\| \leq 2$.

An estimate of $R_{k_j/c}(t)$ is established using the unit quaternion form of the orientation, which can be represented as $q_{k_j/c}(t) \in \mathbb{R}^4$, where $q_{k_j/c}^T(t) q_{k_j/c}(t) = 1$. The time derivative of $q_{k_j/c}(t)$ is

$$\frac{d}{dt}(q_{k_j/c}(t)) = -\frac{1}{2} B(q_{k_j/c}(t)) \omega_c(t), \quad (38)$$

where

$$B(q) \triangleq \begin{bmatrix} -q_2 & -q_3 & -q_4 \\ q_1 & -q_4 & q_3 \\ q_4 & q_1 & -q_2 \\ -q_3 & q_2 & q_1 \end{bmatrix},$$

$q_1, q_2, q_3, q_4 \in \mathbb{R}$ are the four elements of a unit quaternion $q(t)$ and $B^T(q(t)) B(q(t)) = I_{3 \times 3}$.⁴ The rotation matrix representation of a unit quaternion $q(t)$ is

$$R(q) \triangleq \begin{bmatrix} 1 - 2(q_3^2 + q_4^2) & 2(q_2q_3 - q_4q_1) & 2(q_2q_4 + q_3q_1) \\ 2(q_2q_3 + q_4q_1) & 1 - 2(q_2^2 + q_4^2) & 2(q_3q_4 - q_2q_1) \\ 2(q_1q_4 - q_3q_1) & 2(q_3q_4 + q_2q_1) & 1 - 2(q_2^2 + q_3^2) \end{bmatrix}.$$

⁴Time dependence is suppressed except when needed for clarity or introducing terms.

Similar to (38), a predictor is designed for $q_{k_j/c}(t)$ as

$$\frac{d}{dt}(\hat{q}_{k_j/c}(t)) = -\frac{1}{2}B(\hat{q}_{k_j/c}(t))\underline{\omega}_c(t), \quad (39)$$

where $\hat{q}_{k_j/c}(t) \in \mathbb{R}^4$ is the estimate of $q_{k_j/c}(t)$. Only one predictor $\hat{q}_{k_j/c}(t)$ is necessary given $\frac{d}{dt}(q_{k_j/c}(t))$ is not dependent on feature estimates. Additionally, the orientation is often estimated through other methods and a predictor may not be necessary. Since $\frac{d}{dt}(q_{k_j/c}(t))$ is only a function of $q_{k_j/c}(t)$, initializing $\hat{q}_{k_j/c}(\zeta_j^u) = q_{k_j/c}(\zeta_j^u)$ implies (38) and (39) are equivalent for all $t > \zeta_j^u$.

Using the estimates from the predictors in (33), (36), and (39), a predictor is designed for $\eta_{s_{j,i}}(t)$ when $s_{j,i} \in \mathcal{P}_{s_j}^c(t)$ as

$$\frac{d}{dt}(\hat{\eta}_{s_{j,i}}(t)) \triangleq \text{proj}\left(\hat{\mu}_{\xi_{s_{j,i}}}(t)\right), \quad s_{j,i} \in \mathcal{P}_{s_j}^c(t), \quad (40)$$

where $\hat{\mu}_{\xi_{s_{j,i}}}(t) \triangleq \begin{bmatrix} -\hat{u}_{s_{j,i}/c}^T(t) \underline{v}_c(t) \\ -\hat{u}_{k_{j,i}/c}^T(t) \underline{v}_c(t) \\ 0 \end{bmatrix}$ and the projection operator is used to bound $\underline{d}_2 \leq \hat{d}_{s_{j,i}/c}(t) \leq \bar{d}$, $\underline{d}_2 \leq \hat{d}_{s_{j,i}/k_j}(t) \leq \bar{d}$, and $\underline{d}_1 \leq \hat{d}_{k_{j,i}/c}(t)$.

Taking the time derivative of (22), substituting (5)-(7), (34), (37), and (40), simplifying yields

$$\frac{d}{dt}(\tilde{\eta}_{s_{j,i}}(t)) = -\begin{bmatrix} \tilde{u}_{s_{j,i}/c}^T(t) \underline{v}_c(t) \\ \tilde{u}_{k_{j,i}/c}^T(t) \underline{v}_c(t) \\ 0 \end{bmatrix}. \quad (41)$$

B. Stability Analysis of Feature Predictor Design

To quantitatively describe the stability of the observer and predictor, let $\sigma_{s_{j,i}}^o(t) \in \{a, u\}$ describe whether an observer is activated or a predictor is activated, respectively. Specifically, when $\sigma_{s_{j,i}}^o(t) = a$, $s_{j,i} \in \mathcal{P}_{s_j}(t)$ indicating the feature $s_{j,i}$ is in the FOV and an observer is used for feature $s_{j,i}$. Similarly, when $\sigma_{s_{j,i}}^o(t) = u$, $s_{j,i} \in \mathcal{P}_{s_j}^c(t)$ indicating the feature $s_{j,i}$ leaves the FOV and a predictor is used for feature $s_{j,i}$. Let $t_{s_{j,i}}^{a_n}$ represent the n th instance in time feature $s_{j,i}$ enters the FOV (i.e., $s_{j,i} \in \mathcal{P}_{s_j}(t_{s_{j,i}}^{a_n})$) with $\sigma_{s_{j,i}}^o(t_{s_{j,i}}^{a_n}) = a \wedge \sigma_{s_{j,i}}^o(t_{s_{j,i}}^{a_n}) = a$, where $t_{s_{j,i}}^{a_1} \geq \zeta_j^a$. Furthermore, let $t_{s_{j,i}}^{u_n}$ represent the n th instance in time feature $s_{j,i}$ leaves the FOV (i.e., $s_{j,i} \in \mathcal{P}_{s_j}^c(t_{s_{j,i}}^{u_n})$).

Theorem 2. The predictor design in (40) for feature $s_{j,i} \in \mathcal{P}_{s_j}^c(t)$ ensures the estimation error $\tilde{\eta}_{s_{j,i}}(t)$ is bounded as

$$\|\tilde{\eta}_{s_{j,i}}(t)\| \leq \|\tilde{\eta}_{s_{j,i}}(t_{s_{j,i}}^{u_n})\| + 4\bar{v}_c(t - t_{s_{j,i}}^{u_n}). \quad (42)$$

Proof: Taking the time derivative of (25), substituting (41), and using the bounds $\|\tilde{u}_{k_{j,i}/c}(t)\| \leq 2$ and $\|\tilde{u}_{s_{j,i}/c}(t)\| \leq 2$ yields

$$\frac{d}{dt}(V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t))) \leq 4\sqrt{2}\bar{v}_c\sqrt{V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t))}. \quad (43)$$

Invoking the Comparison Lemma [75, Lemma 3.4] on (43) yields the result in (42). ■

C. Ensuring Stability Through Dwell-Time Conditions

To facilitate the subsequent development, let $\mathcal{L}_{s_{j,i}}^n \triangleq \{l \in \mathbb{Z}_{>0} : \zeta_{s_{j,i}}^{a_l} \in [t_{s_{j,i}}^{a_n}, t_{s_{j,i}}^{u_n}] \cap t > \tau_{s_{j,i}}\}$, where $\underline{l}_n, \bar{l}_n, \bar{l}_n \in \mathbb{Z}_{>0}$ represent the first and last $l \in \mathcal{L}_{s_{j,i}}^n$, respectively, and $\bar{l}_n \triangleq \bar{l}_n - l_n$. Additionally, let $l_n \in \{1, 2, \dots, \bar{l}_n + 1\}$ represent the index of $\mathcal{L}_{s_{j,i}}^n$, and let $\zeta_{s_{j,i}}^{a_{l_n,n}} \in [t_{s_{j,i}}^{a_n}, t_{s_{j,i}}^{u_n}] \cap t > \tau_{s_{j,i}}$ and $\zeta_{s_{j,i}}^{u_{l_n,n}} \in [t_{s_{j,i}}^{a_n}, t_{s_{j,i}}^{u_n}] \cap t > \tau_{s_{j,i}}$ represent the instances in time for the n th return such that $\sigma_{s_{j,i}}(t) = a$ and $\sigma_{s_{j,i}}(t) = u$, respectively. Furthermore, let $\Delta t_{s_{j,i}}^{a_n} \triangleq \sum_{l_n=1}^{\bar{l}_n} \Delta \zeta_{s_{j,i}}^{a_{l_n,n}}$ and $\Delta t_{s_{j,i}}^{u_n} \triangleq t_{s_{j,i}}^{a_{n+1}} - t_{s_{j,i}}^{u_n}$, where $\Delta \zeta_{s_{j,i}}^{a_{l_n,n}} \triangleq \zeta_{s_{j,i}}^{u_{l_n,n}} - \zeta_{s_{j,i}}^{a_{l_n,n}}$, $\Delta \zeta_{s_{j,i}}^{u_{l_n,n}} \triangleq \zeta_{s_{j,i}}^{a_{l_n+1,n}} - \zeta_{s_{j,i}}^{u_{l_n,n}}$, and $\Delta \zeta_{s_{j,i}}^{u_{\bar{l}_n,n}} \triangleq t_{s_{j,i}}^{u_{\bar{l}_n,n}} - \zeta_{s_{j,i}}^{u_{\bar{l}_n,n}}$.

To ensure that the system defined by the switching signals $\sigma_{s_{j,i}}(t)$ and $\sigma_{s_{j,i}}^o(t)$ remains bounded, minimum and maximum dwell-times must be developed for each observer and predictor, respectively. Approaches like those taken in [1] and [2] will not be possible for switched systems like those defined by $\sigma_{s_{j,i}}(t)$ and $\sigma_{s_{j,i}}^o(t)$ given $\tilde{\eta}_{s_{j,i}}(t_{s_{j,i}}^{a_n})$ and $\tilde{\eta}_{s_{j,i}}(t_{s_{j,i}}^{u_n})$ are unknown and cannot be reset. However, $\tilde{d}_{s_{j,i}/c}(t_{s_{j,i}}^{a_1}) \leq \bar{d}$, $\tilde{d}_{k_{j,i}/c}(t_{s_{j,i}}^{a_1}) \leq \bar{d}$, and $\tilde{d}_{s_{j,i}/k_j}(t_{s_{j,i}}^{a_1}) \leq \bar{d}$ implying $\|\tilde{\eta}_{s_{j,i}}(t_{s_{j,i}}^{a_1})\| \leq \sqrt{3}\bar{d}$.

Let $\tilde{\eta}_{s_{j,i}} \in (0, \bar{d})$ be a user-defined threshold such that $\|\tilde{\eta}_{s_{j,i}}(\zeta_{s_{j,i}}^{u_{l_n,n}})\| \leq \tilde{\eta}_{s_{j,i}}$ (i.e., the user will need the error below some threshold before allowing the object to leave the FOV). Furthermore, let $\overline{\tilde{\eta}_{s_{j,i}}} \in (\underline{\tilde{\eta}_{s_{j,i}}}, \bar{d})$ be a user-defined threshold such that $\tilde{\eta}_{s_{j,i}} < \|\tilde{\eta}_{s_{j,i}}(\zeta_{s_{j,i}}^{a_{l_n,n}})\| \leq \overline{\tilde{\eta}_{s_{j,i}}}$ for all $n > 1$. The thresholds $\underline{\tilde{\eta}_{s_{j,i}}}$ and $\overline{\tilde{\eta}_{s_{j,i}}}$ represent the acceptable amount of error for a user's application before a feature may leave and must return to the FOV, respectively. While the true error $\tilde{\eta}_{s_{j,i}}(t)$ is unknown, bounds on the distances are known and establishing the bounds $\underline{\tilde{\eta}_{s_{j,i}}}$ and $\overline{\tilde{\eta}_{s_{j,i}}}$ as described will ensure that the errors are within the thresholds provided the subsequently developed dwell-time conditions are satisfied. Specifically, the dwell-times are established such that the upper bound on the distance errors converge implying the true errors must also converge.

Assumption 7. It is possible for the system to satisfy the subsequently developed dwell-time conditions for the set of features $\mathcal{A}_{s_j}(t)$. Also, after $t > t_{s_{j,i}}^{u_1}$, the only tracked features from the j th object will be those contained in $\mathcal{A}_{s_j}(t)$. Specifically, $\mathcal{P}_{s_j}(t) \subseteq \mathcal{A}_{s_j}(t)$ for time $t > t_{s_{j,i}}^{u_1}$, implying $\mathcal{A}_{s_j}(t) \cap \mathcal{P}_{s_j}(t) = \mathcal{P}_{s_j}(t)$.

Under Assumptions 2 and 3, the rotation matrix $R_{k_{j,i}/c}(t)$ and unit vector $\underline{u}_{k_{j,i}/c}(t)$ can be determined from the set of stationary features in $\mathcal{P}_{s_j}(t)$ while $p_{s_j}(t) \geq 4$. This implies that it is not sufficient to only consider the dwell-times for individual features given the observers are used on the assumption $\underline{u}_{k_{j,i}/c}(t)$ is available. Let $\sigma_{s_j}^o(t) \in \{a, u\}$ be a switching signal that indicates when there are enough features in the FOV to determine $\underline{u}_{k_{j,i}/c}(t)$ and $R_{k_{j,i}/c}(t)$; specifically, the first and second mode of $\sigma_{s_j}^o(t)$ represents

when $p_{s_j}(t) \geq 4$ and $p_{s_j}(t) < 4$, respectively. However, for the object to be successfully recaptured, $\sigma_{s_j}^o(t) = a \wedge \bigcap_{s_{j,i} \in \mathcal{A}_{s_j}(t)} (\sigma_{s_{j,i}}^o(t) = a \wedge \sigma_{s_{j,i}}(t) = a)$ implying each feature $s_{j,i} \in \mathcal{A}_{s_j}(t)$ is in the FOV and the relative motion is sufficient for learning. Let $\bar{t}_{s_j}^{u_n} \triangleq \max\{\zeta_{s_{j,i}}^{a_{1,n}}\}$, $\underline{t}_{s_j}^{u_n} \triangleq \min\{\zeta_{s_{j,i}}^{u_{n,n}}\}$, and $\Delta t_{s_j}^{a_n} \triangleq \bar{t}_{s_j}^{u_n} - \underline{t}_{s_j}^{u_n}$ for the n th switching cycle across all features in $\mathcal{A}_{s_j}(t)$. Furthermore, let $\bar{t}_{s_j}^{u_{n+1}} \triangleq \max\{\zeta_{s_{j,i}}^{a_{1,n+1}}\}$ and $\underline{t}_{s_j}^{u_n} \triangleq \bar{t}_{s_j}^{u_{n+1}} - \underline{t}_{s_j}^{u_n}$.

Theorem 3. For each feature in the set of features $\mathcal{A}_{s_j}(t)$, the errors of the switched system defined by the switching signals $\sigma_{s_{j,i}}(t)$, $\sigma_{s_{j,i}}^o(t)$, and $\sigma_{s_j}^o(t)$, and the observer update law in (23) ensure the estimation error in $\tilde{\eta}_{s_{j,i}}(t)$ at time $t = t_{s_{j,i}}^{u_1}$ is GUUB as $\|\tilde{\eta}_{s_{j,i}}(t_{s_{j,i}}^{u_1})\| \leq \tilde{\eta}_{s_{j,i}}$ provided the switching signals satisfy the initial minimum feedback availability dwell-time condition

$$\underline{\Delta t}_{s_j}^{a_1} \geq -\frac{1}{\beta} \ln \left(\frac{\tilde{\eta}_{s_{j,i}}}{\sqrt{3} \bar{d}} \right) > -\frac{1}{\beta} \ln \left(\frac{1}{\sqrt{3}} \right). \quad (44)$$

Proof: Using (32) for the first instance implies $\|\tilde{\eta}_{s_{j,i}}(t_{s_{j,i}}^{u_1})\| \leq \|\tilde{\eta}_{s_{j,i}}(t_{s_{j,i}}^{a_1})\| \exp(-\beta \Delta t_{s_{j,i}}^{a_1})$. It is desired to have $\|\tilde{\eta}_{s_{j,i}}(t_{s_{j,i}}^{u_1})\| \leq \tilde{\eta}_{s_{j,i}}$ and the initial error is bounded as $\|\tilde{\eta}_{s_{j,i}}(t_{s_{j,i}}^{a_1})\| \leq \sqrt{3} \bar{d}$. Substituting these bounds into the first inequality and solving for $\Delta t_{s_{j,i}}^{a_1}$ yields $\Delta t_{s_{j,i}}^{a_1} \geq -\frac{1}{\beta} \ln \left(\frac{\tilde{\eta}_{s_{j,i}}}{\sqrt{3} \bar{d}} \right)$. Because $\underline{\Delta t}_{s_j}^{a_1}$ must lower bound the dwell-times to ensure all of the feature observers are implementable, $\underline{\Delta t}_{s_j}^{a_1} \geq -\frac{1}{\beta} \ln \left(\frac{\tilde{\eta}_{s_{j,i}}}{\sqrt{3} \bar{d}} \right)$. Since $\tilde{\eta}_{s_{j,i}} < \bar{d}$, $\frac{\tilde{\eta}_{s_{j,i}}}{\sqrt{3} \bar{d}} < \frac{1}{\sqrt{3}}$ yielding the bound in (44). ■

Theorem 4. For each feature in the set of features $\mathcal{A}_{s_j}(t)$, the errors of the switched system defined by the switching signals $\sigma_{s_{j,i}}(t)$, $\sigma_{s_{j,i}}^o(t)$, and $\sigma_{s_j}^o(t)$, and the observer update law in (23) ensure the estimation error in $\tilde{\eta}_{s_{j,i}}(t)$ is GUUB as $\|\tilde{\eta}_{s_{j,i}}(t_{s_{j,i}}^{u_n})\| \leq \tilde{\eta}_{s_{j,i}}$ provided the switching signals satisfy the minimum feedback availability dwell-time condition

$$\underline{\Delta t}_{s_j}^{a_n} \geq -\frac{1}{\beta} \ln \left(\frac{\tilde{\eta}_{s_{j,i}}}{\tilde{\eta}_{s_{j,i}}} \right) > 0, n > 1 \quad (45)$$

Proof: The proof follows Theorem 3 using the upper bound $\|\tilde{\eta}_{s_{j,i}}(t_{s_{j,i}}^{a_n})\| \leq \tilde{\eta}_{s_{j,i}}$. ■

Theorem 5. For each feature in the set of features $\mathcal{A}_{s_j}(t) \cap \mathcal{P}_{s_j}^c(t)$, the errors of the switched system defined by the switching signals $\sigma_{s_{j,i}}(t)$, $\sigma_{s_{j,i}}^o(t)$, and $\sigma_{s_j}^o(t)$, and the predictor update law in (40) ensure the estimation error in $\tilde{\eta}_{s_{j,i}}(t)$ is GUUB as $\|\tilde{\eta}_{s_{j,i}}(t_{s_{j,i}}^{a_{n+1}})\| \leq \tilde{\eta}_{s_{j,i}}$ provided the switching signals satisfy the maximum loss of feedback dwell-time condition

$$0 < \overline{\Delta t}_{s_j}^{u_n} \leq \frac{\tilde{\eta}_{s_{j,i}} - \tilde{\eta}_{s_{j,i}}}{4\bar{v}_c} \quad (46)$$

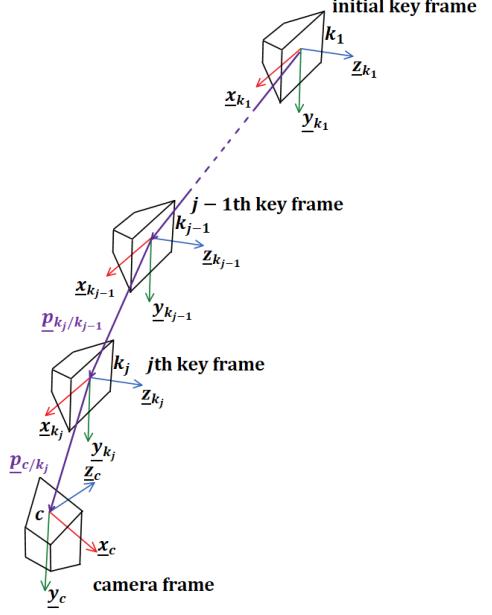


Figure 2. Example geometry for pose of the camera over time where the camera starts at the top where the first key frame is located and is traveling downward to the lower left.

Proof: Using (42) for the n th instance implies $\|\tilde{\eta}_{s_{j,i}}(t_{s_{j,i}}^{a_{n+1}})\| \leq \|\tilde{\eta}_{s_{j,i}}(t_{s_{j,i}}^{u_n})\| + 4\bar{v}_c \Delta t_{s_{j,i}}^{u_n}$. It is desired to have $\|\tilde{\eta}_{s_{j,i}}(t_{s_{j,i}}^{a_{n+1}})\| \leq \tilde{\eta}_{s_{j,i}}$ and $\|\tilde{\eta}_{s_{j,i}}(t_{s_{j,i}}^{u_n})\| \leq \tilde{\eta}_{s_{j,i}}$. Substituting the second bounds into the first and solving for $\Delta t_{s_{j,i}}^{u_n}$ yields $\Delta t_{s_{j,i}}^{u_n} \leq \frac{\tilde{\eta}_{s_{j,i}} - \tilde{\eta}_{s_{j,i}}}{4\bar{v}_c}$. Because $\overline{\Delta t}_{s_j}^{u_n}$ must upper bound the dwell-times to ensure all of the feature observers are implementable, $\overline{\Delta t}_{s_j}^{u_n} \leq \frac{\tilde{\eta}_{s_{j,i}} - \tilde{\eta}_{s_{j,i}}}{4\bar{v}_c}$. Given $\tilde{\eta}_{s_{j,i}} > \underline{\eta}_{s_{j,i}}$, $\tilde{\eta}_{s_{j,i}} - \underline{\eta}_{s_{j,i}} > 0$, yielding the bound in (46). ■

Ensuring that (44)-(46) are satisfied guarantees the error remains bounded as $\underline{\eta}_{s_{j,i}} < \|\tilde{\eta}_{s_{j,i}}(t)\| \leq \sqrt{3} \bar{d}$ for all time $t < t_{s_{j,i}}^{u_1}$, $\underline{\eta}_{s_{j,i}} < \|\tilde{\eta}_{s_{j,i}}(t)\| \leq \tilde{\eta}_{s_{j,i}}$ for all time $t \geq t_{s_{j,i}}^{u_1}$, and enables the ability to recapture the features in $\mathcal{A}_{s_j}(t)$ and use them to improve position estimates.

VI. ESTIMATOR DESIGN FOR POSE OF CAMERA

As previously discussed, the primary purpose of the dwell-time analysis is to ensure the object feature observers have converged below a user-defined threshold before using feature structure estimates in an observer for the position of the camera; however, there will be time periods where no objects remain in the camera's FOV. The position estimation objective is to express the camera pose in a fixed coordinate frame. As shown in Figure 2, the pose of the camera may be expressed through the sequence of objects. The starting location of the camera may be unknown but the pose of the camera over time can always be expressed relative to the first key frame; yet, in many applications, the region around the first object may directly have feedback available (cf., [1] and [2]). Let $\underline{p}_{c/k_1}(t) \in \mathbb{R}^3$ represent the position of the camera with respect to the first key frame expressed in the first key frame

and $q_{c/k_1}(t) \in \mathbb{R}^4$ be the quaternion representation of the orientation of the camera with respect to the first key frame.

Assumption 8. There exists a landmark around the first object that is unique compared to all other objects and $\underline{p}_{c/k_1}(t)$ and $q_{c/k_1}(t)$ are directly measurable through some means while the landmark is in the camera's FOV. Let $\sigma_L(t) \in \{a, u\}$ be a switching signal indicating whether the landmark is in the FOV or not in the FOV, respectively.

Considering the j th object only provides measurements of $\underline{u}_{k_j/c}(t)$ and $q_{k_j/c}(t)$ and estimates of $d_{k_j/c}$, an observer must be used to estimate $\underline{p}_{c/k_1}(t)$ and $q_{c/k_1}(t)$ while the j th object is in the FOV. When no object has feedback available, a predictor must be used to estimate $\underline{p}_{c/k_1}(t)$ and $q_{c/k_1}(t)$. Based on the minimum dwell-time analysis in Theorem 3, let $\sigma_{s_{j,i}}^{\Delta t}(t) \in \{a, u\}$ be a switching signal for i th feature on the j th object indicating when the total time converging has exceeded the minimum dwell-time condition or has not exceeded the minimum dwell time condition. Specifically, $\sigma_{s_{j,i}}^{\Delta t}(t) = a$ implies $\left(\sum_{l_n=1}^{l_n^*(t)} \Delta \zeta_{s_{j,i}}^{a_{l_n,n}} > \underline{\Delta t}_{s_j}^{a_n} \right) \wedge t \in [t_{s_{j,i}}^{a_n}, t_{s_{j,i}}^{u_n}]$, where $l_n^*(t) \in \mathbb{Z}_{>0}$ represents the current index of the n th cycle for the i th feature on the j th object. Similarly, let $\sigma_{s_j}^{\Delta t}(t) \in \{a, u\}$ be a switching signal indicating when all the remaining features on the j th object with feedback available have either satisfied or not satisfied the dwell-time condition, specifically, $\sigma_{s_j}^{\Delta t}(t) = a$ implies all remaining features on the j th object $s_{j,i} \in \mathcal{P}_{s_j}(t)$ have $\sigma_{s_{j,i}}^{\Delta t}(t) = a$. Furthermore, let $\tau_{s_j}^{a_n} \triangleq \min \{t > t_{s_{j,i}}^{a_n} : \sigma_{s_j}^{\Delta t} = a\}$, represent the time that all of the remaining features on the j th object $s_{j,i} \in \mathcal{P}_{s_j}(t)$ have satisfied the minimum dwell-time for the n th cycle. When the dwell-time condition has been satisfied for the j th object during the n th cycle (i.e., $t \in [\tau_{s_j}^{a_n}, t_{s_{j,i}}^{u_n}]$), the error in each of the feature observers is less than the desired threshold (i.e., $\|\tilde{\eta}_{s_{j,i}}(t)\| \leq \underline{\eta}_{s_{j,i}}$), implying $\tilde{d}_{k_{j,i}/c}(t) \leq \tilde{\eta}_{s_{j,i}}$.

Let the pose error be quantified as $\tilde{p}_{c/k_1}(t) \in \mathbb{R}^3$ and $\tilde{q}_{c/k_1}(t) \in \mathbb{R}^4$, where

$$\tilde{p}_{c/k_1}(t) \triangleq \underline{p}_{c/k_1}(t) - \hat{p}_{c/k_1}(t), \quad (47)$$

$$\tilde{q}_{c/k_1}(t) \triangleq q_{c/k_1}(t) - \hat{q}_{c/k_1}(t), \quad (48)$$

and $\hat{p}_{c/k_1}(t) \in \mathbb{R}^3$ and $\hat{q}_{c/k_1}(t) \in \mathbb{R}^4$ are the estimates of $\underline{p}_{c/k_1}(t)$ and $q_{c/k_1}(t)$, respectively. Taking the time derivative of $\underline{p}_{c/k_1}(t)$ yields

$$\frac{d}{dt}(\underline{p}_{c/k_1}(t)) = R(q_{c/k_1}(t)) \underline{v}_c(t). \quad (49)$$

Similarly, the time derivative of $q_{c/k_1}(t)$ is

$$\frac{d}{dt}(q_{c/k_1}(t)) = \frac{1}{2}B(q_{c/k_1}(t)) \underline{\omega}_c(t). \quad (50)$$

Taking the time derivative of $\tilde{p}_{c/k_1}(t)$ and $\tilde{q}_{c/k_1}(t)$, and substituting (49) and (50) yields

$$\frac{d}{dt}(\tilde{p}_{c/k_1}(t)) = R(q_{c/k_1}(t)) \underline{v}_c(t) - \frac{d}{dt}(\hat{p}_{c/k_1}(t)) \quad (51)$$

and

$$\frac{d}{dt}(\tilde{q}_{c/k_1}(t)) = \frac{1}{2}B(q_{c/k_1}(t)) \underline{\omega}_c(t) - \frac{d}{dt}(\hat{q}_{c/k_1}(t)). \quad (52)$$

When the landmark is in the FOV, $\sigma_L(t) = a$ and feedback of the pose of the camera is directly available under Assumption 8. While $\sigma_L(t) = a$, a reset map is used to reset both position and orientation as

$$\hat{p}_{c/k_1}(t) \rightarrow \underline{p}_{c/k_1}(t) \quad (53)$$

and

$$\hat{q}_{c/k_1}(t) \rightarrow q_{c/k_1}(t). \quad (54)$$

However, while $\sigma_L(t) = u$, an observer or predictor is used to estimate the pose of the camera depending on the set of switching signals, $\{\sigma_{s_j}^{\Delta t}(t)\}_{j=1}^{\bar{p}_s(t)}$.

When the feature estimators for the j th object have satisfied the minimum dwell-time condition, $\sigma_{s_j}^{\Delta t}(t) = a$, and enough features are remaining on the object $\sigma_{s_j}^o(t) = a$, $p_{s_j}(t) \geq 4$, both $R_{k_j/c}(t)$ and $\underline{u}_{k_j/c}(t)$ are measurable and the estimation error for each remaining feature $\|\tilde{d}_{k_{j,i}/c}(t)\| \leq \underline{\eta}_{s_{j,i}}$. While $\sigma_{s_j}^{\Delta t}(t) = a \wedge \sigma_{s_j}^o(t) = a$, an estimate of the position of the camera with respect to the j th key frame is available as

$$\hat{p}_{c/k_j}^*(t) \triangleq \underline{u}_{c/k_j}(t) \hat{d}_{k_j/c}(t), \quad (55)$$

where $\hat{d}_{k_j/c}(t) \triangleq \frac{1}{p_{s_j}(t)} \sum_{i=1}^{p_{s_j}(t)} \hat{d}_{k_{j,i}/c}(t)$, $\underline{u}_{c/k_j}(t) = -R_{c/k_j}(t) \underline{u}_{k_j/c}(t)$, $R_{c/k_j}(t) = R_{k_j/c}^T(t)$, and $\hat{p}_{c/k_j}^*(t)$ is expressed in the j th key frame based on (55). Using (55), an observer is designed for the pose of the camera while $\sigma_L(t) = u \wedge \sigma_{s_j}^{\Delta t}(t) = a \wedge \sigma_{s_j}^o(t) = a$ as

$$\begin{aligned} \frac{d}{dt}(\hat{p}_{c/k_1}(t)) &\triangleq R(\hat{q}_{c/k_1}(t)) \underline{v}_c(t) \\ &+ K_p (\hat{p}_{k_j/k_1} + R(\hat{q}_{k_j/k_1}) \hat{p}_{c/k_j}^*(t) - \hat{p}_{c/k_1}(t)), \end{aligned} \quad (56)$$

and

$$\begin{aligned} \frac{d}{dt}(\hat{q}_{c/k_1}(t)) &\triangleq \frac{1}{2}B(\hat{q}_{c/k_1}(t)) \underline{\omega}_c(t) \\ &+ K_q (Q(\hat{q}_{k_j/k_1}) q_{c/k_j}(t) - \hat{q}_{c/k_1}(t)), \end{aligned} \quad (57)$$

where $\hat{q}_{k_j/k_1} = \hat{q}_{c/k_1}(\zeta_1^a)$, $\hat{p}_{k_j/k_1} = \hat{p}_{c/k_1}(\zeta_1^a)$, $K_p \in \mathbb{R}^{3 \times 3}$ and $K_q \in \mathbb{R}^{4 \times 4}$ are constant positive gain matrices, and

$$Q(q) \triangleq \begin{bmatrix} q_1 & -q_2 & -q_3 & -q_4 \\ q_2 & q_1 & -q_4 & q_3 \\ q_3 & q_4 & q_1 & -q_2 \\ q_4 & -q_3 & q_2 & q_1 \end{bmatrix}.$$

However, if the minimum dwell-time condition for the object is unsatisfied or the object has too few features in the FOV, a predictor is used to update the pose estimate. Specifically, when $\sigma_L(t) = u \wedge (\sigma_{s_j}^{\Delta t}(t) = u \vee \sigma_{s_j}^o(t) = u)$, a predictor is designed for the pose as

$$\frac{d}{dt}(\hat{p}_{c/k_1}(t)) \triangleq R(\hat{q}_{c/k_1}(t)) \underline{v}_c(t) \quad (58)$$

and

$$\frac{d}{dt}(\hat{q}_{c/k_1}(t)) \triangleq \frac{1}{2}B(\hat{q}_{c/k_1}(t))\omega_c(t). \quad (59)$$

Only an analysis of the position estimator design is considered since the primary result of this work is estimating the position of the camera. Additionally, the orientation is often estimated through other methods and a predictor may not be necessary. Since $\frac{d}{dt}(q_{c/k_1}(t))$ is only a function of $q_{c/k_1}(t)$, initializing $\hat{q}_{c/k_1}(\zeta_1^a) = [1 \ 0 \ 0 \ 0]^T$ implies (50) is equivalent to (57) and (59) for all $t > \zeta_1^a$.

While $\sigma_L(t) = u$ and $\sigma_{s_j}^{\Delta t}(t) = a \wedge \sigma_{s_j}^o(t) = a$ for the j th object, substituting the position observer update law in (56) into the time derivative of the position error in (51), using $\hat{q}_{c/k_1}(t) = q_{c/k_1}(t)$, and simplifying yields

$$\begin{aligned} \frac{d}{dt}(\tilde{p}_{c/k_1}(t)) &= -K_p \tilde{p}_{c/k_1}(t) + K_p \tilde{p}_{k_j/c}(t) \\ &+ K_p R(q_{k_j/k_1}) \underline{u}_{c/k_1}(t) \frac{1}{p_{s_j}(t)} \sum_{i=1}^{p_{s_j}(t)} \tilde{d}_{k_j,i/c}(t). \end{aligned} \quad (60)$$

While $\sigma_L(t) = u$ and $\sigma_{s_j}^{\Delta t}(t) = u \vee \sigma_{s_j}^o(t) = u$ for all objects, substituting the position predictor in (58) into the time derivative of the position error in (51), using $\hat{q}_{c/k_1}(t) = q_{c/k_1}(t)$, and simplifying yields

$$\frac{d}{dt}(\tilde{p}_{c/k_1}(t)) = 0_{3 \times 1}. \quad (61)$$

A. Stability of Key Frame Position Observer and Predictor Design

Let $V_{c/k_1}(\tilde{p}_{c/k_1}(t)) : \mathbb{R}^3 \rightarrow \mathbb{R}$ be a candidate Lyapunov function defined as

$$V_{c/k_1}(\tilde{p}_{k_j/c}(t)) \triangleq \frac{1}{2}\tilde{p}_{c/k_1}^T(t)\tilde{p}_{c/k_1}(t), \quad (62)$$

which can be bounded as $\frac{1}{2}\|\tilde{p}_{c/k_1}(t)\|^2 \leq V_{c/k_1}(\tilde{p}_{c/k_1}(t)) \leq \frac{1}{2}\|\tilde{p}_{c/k_1}(t)\|^2$.

Theorem 6. The switching signals $\sigma_L(t)$, $\sigma_{s_j}^{\Delta t}(t)$, and $\sigma_{s_j}^o(t)$ and the observer update law designed in (56) ensure the position error of the camera $\tilde{p}_{c/k_1}(t)$ is GUUB while $t \in [\tau_{s_j}^{a_n}, t_{s_j}^{u_n}]$ in the sense

$$\|\tilde{p}_{c/k_1}(t)\|^2 \leq \|\tilde{p}_{c/k_1}(\tau_{s_j}^{a_n})\|^2 \exp\left(-\beta_p^a(t - \tau_{s_j}^{a_n})\right) + 2\frac{\varepsilon_{k_j}}{\beta_p^a} \quad (63)$$

where $\beta_p^a \triangleq \lambda_{\min}\{K_p\}$, $\varepsilon_{k_j} \triangleq \frac{\left(\lambda_{\max}\{K_p\}\left(\|\tilde{p}_{k_j/c}\| + \tilde{\eta}_{s_j,i}\right)\right)^2}{2\lambda_{\min}\{K_p\}}$, and $\lambda_{\max}\{\cdot\}$ is the maximum eigenvalue of $\{\cdot\}$.

Proof: Taking the time derivative of (62) and substituting (60) yields

$$\frac{d}{dt}\left(V_{c/k_1}(\tilde{p}_{c/k_1}(t))\right) \leq -\beta_p^a V_{c/k_1}(\tilde{p}_{c/k_1}(t)) + \varepsilon_{k_j}. \quad (64)$$



Figure 3. Image of the Kobuki Turtlebot and iDS uEye camera used for experiments.



Figure 4. Image of a section of the wooden hallways in the environment and locations of the motion capture cameras in that section. The motion capture cameras were located throughout the environment attached to the upper portion of each wall.

Invoking the Comparison Lemma [75, Lemma 3.4] on (64) then upper bounding yields (63). ■

Theorem 7. The switching signals $\sigma_L(t)$, $\sigma_{s_j}^{\Delta t}(t)$, and $\sigma_{s_j}^o(t)$ and the predictor update law designed in (58) ensure the position error of the camera $\tilde{p}_{c/k_1}(t)$ is bounded while $t \notin [\tau_{s_j}^{a_n}, t_{s_j}^{u_n}]$ in the sense

$$\|\tilde{p}_{c/k_1}(t)\| \leq \|\tilde{p}_{c/k_1}(t_{s_j}^{u_n})\|. \quad (65)$$

Proof: Taking the time derivative of (62) and substituting (61) yields

$$\frac{d}{dt}\left(V_{c/k_1}(\tilde{p}_{c/k_1}(t))\right) \leq 0. \quad (66)$$

Invoking the Comparison Lemma [75, Lemma 3.4] on the result then upper bounding yields (65). ■

VII. EXPERIMENTS

An experiment is provided to demonstrate the performance of the developed estimator strategy using the observer and predictor design compared to a predictor-only strategy (cf., [1] and [2]). The experiment assumed no return to previous

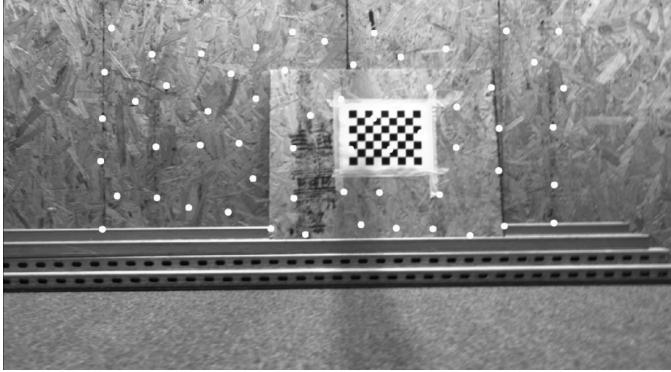


Figure 5. Image of the landmark captured within a key frame (i.e., the checkerboard is leaning on the wooden wall), where the white dots drawn in the image are the extracted corner features.

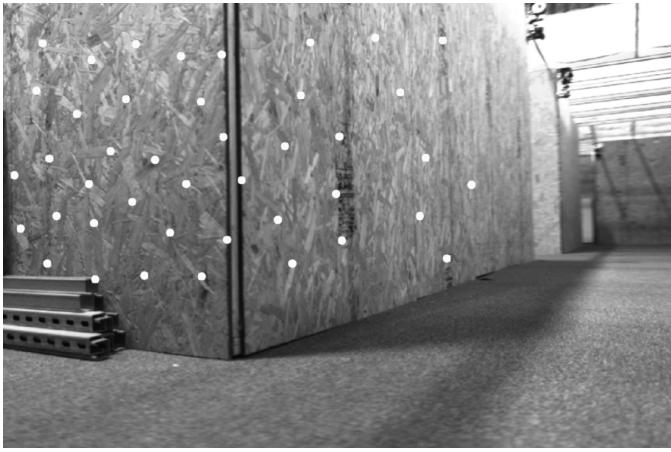


Figure 6. Image of the features extracted from a key frame image of a nonplanar object (i.e., two wooden walls with a 90° angle between them), where the white dots are the extracted corner features.

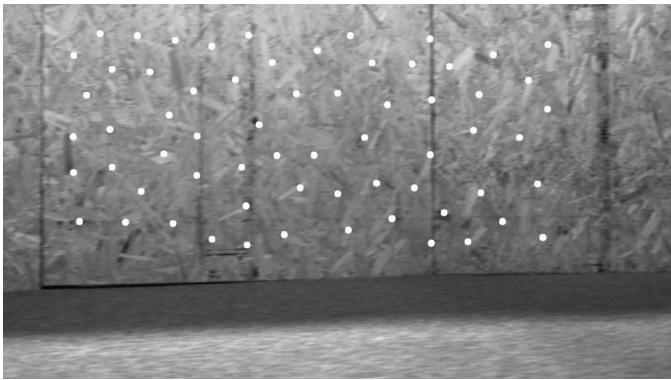


Figure 7. Image of the features extracted from a key frame image of a planar object, where the white dots are the extracted corner features.

objects (i.e., each new set of features was considered a new object) implying the initial minimum dwell-time condition in (44) must always be satisfied before using a new object in the position observer. The performance examined is that of the developed feature observer in (23) and the pose estimation strategy using the reset maps in (53) and (54) when a landmark is in the FOV (i.e., $\sigma_L(t) = a$), the observer update laws in (56) and (57) while no landmark is in the FOV and an object in the FOV has satisfied the minimum dwell-time condition and has enough remaining features in the FOV (i.e., $\sigma_L(t) = u$ and $\sigma_{s_j}^{\Delta t}(t) = a \wedge \sigma_{s_j}^o(t) = a$), and the predictor update laws in (58) and (59) while no landmark is in the FOV and no object has satisfied the dwell-time condition or does not have enough remaining features (i.e., $\sigma_L(t) = u$ and $\sigma_{s_j}^{\Delta t}(t) = u \vee \sigma_{s_j}^o(t) = u$).

A Kobuki Turtlebot with a 1920 × 1080 monochrome iDS uEye camera, shown in Figure 3, provided images and velocity at 30 Hz as it drove through the environment (e.g., Figure 4). The estimator can run real-time (i.e., at 30 Hz) and was implemented using Eigen3, OpenCV, and ROS c++ libraries (cf., [76], [77], and [78], respectively). The landmark was a checkerboard with 8 × 6 corners where each square is 25.4 millimeters × 25.4 millimeters. As the example in Figure 4 shows, the environment was an enclosed series of wooden hallways where a motion capture system provided ground truth. An Optitrack motion capture system operated at 120 Hz and measured the pose of the camera, allowing for the position of each feature relative to the camera to be known for comparison. A computer with an Intel i7 processor running at 3.4 GHz was used to simultaneously perform image processing and estimator updates.

Between approximately 30 and 50 corner features (cf., [70] or [77]) were extracted from each key frame image with an initial spacing of 100 pixels.⁵ A new object and associated key frame was added after the previous object left the FOV. An object and the associated key frame were no longer tracked and considered out of the FOV when less than 20 corner features remained (i.e., $p_{s_j}(t) \geq 20$ implied $\sigma_{s_j}^o(t) = a$ and $p_{s_j}(t) < 20$ implied $\sigma_{s_j}^o(t) = u$).⁶ The environment consisted of both nonplanar (e.g., Figure 6) and planar (e.g., Figure 7) surfaces demonstrating the planar assumption was not always valid. If the landmark was in the FOV while tracking an extracted set of features (e.g., Figure 5), the ground truth was provided to the vehicle and the reset maps in (53) and (54) where used to update the pose estimate.

Extracted corner features were tracked while in the FOV by first predicting the location of a corner feature in a new image using the current estimate of the distance to the feature and the location of the corner feature in the old image. For example, consider tracking the location of the i th feature on the j th object while it is in the FOV (i.e., $s_{j,i} \in \mathcal{P}_{s_j}(t)$), integrating (14) from the previous image at time $t_p \in \mathbb{R}_{>0}$

⁵The extraction method could only find between 30 and 50 corner features in a key frame image depending on what was in the image.

⁶While an absolute minimum of 4 features is required, 4 corner features will typically provide a poor estimate of $q_{k_j/c}(t)$ and $u_{k_j/c}(t)$ and it was experimentally determined that 20 corner features was the lowest number of features that could consistently provide good estimates.

using $\underline{u}_{s_{j,i}/c}(t_p)$, $\hat{d}_{s_{j,i}/c}(t_p)$, $\underline{v}_c(t_p)$, and $\underline{\omega}_c(t_p)$ to the time of the new image t yields

$$\begin{aligned}\hat{u}_{s_{j,i}/c}(t) = & \underline{u}_{s_{j,i}/c}(t_p) - \int_{t_p}^t \underline{\omega}_c^\times(\iota) \underline{u}_{s_{j,i}/c}(\iota) d\iota \\ & + \int_{t_p}^t \frac{1}{\hat{d}_{s_{j,i}/c}(\iota)} \underline{u}_{s_{j,i}/c}(\iota) \underline{u}_{s_{j,i}/c}^T(\iota) \underline{v}_c(\iota) d\iota \\ & - \int_{t_p}^t \frac{1}{\hat{d}_{s_{j,i}/c}(\iota)} \underline{v}_c(\iota) d\iota.\end{aligned}\quad (67)$$

The average of the shift estimated by all the features in $\mathcal{P}_{s_j}(t)$ was then used to estimate an affine transformation between the two images (i.e., the unit vectors were converted back into pixels and the average of the change in pixels was used to determine an affine transformation between the images). The approximated affine transformation was then applied to a 50 pixel \times 50 pixel patch of the previous image around the previous pixel location of each feature $s_{j,i}$. The transformed patch was used as a template to search for a match in a 90 pixel \times 90 pixel patch around the predicted feature location in the new image using normalized cross correlation coefficient template matching (cf., [9] or [77]). The best matches provided by the template matching were then used to determine the set of $\underline{u}_{s_{j,i}/c}(t)$. The average shift after template matching by all the features in $\mathcal{P}_{s_j}(t)$ was then used to determine outliers by calculating the χ^2 value of a features shift compared to the average shift using a standard deviation of 3 pixels. If a feature had a χ^2 value greater than 6.63 it was considered an outlier. The value of $\frac{d}{dt}(\underline{u}_{s_{j,i}/c}(t))$ was then estimated using a filtered backwards difference on $\underline{u}_{s_{j,i}/c}(t)$ and used to update (15) and subsequently update (23), where $K_\xi = 40I_{3 \times 3}$ was determined to consistently improve performance with low noise sensitivity. After finding the features, multiple methods (i.e., essential, homography, and perspective-n-point decompositions, [77]) were used to approximate $q_{k_j/c}(t)$ and $\underline{u}_{k_j/c}(t)$ from the set of features $\mathcal{P}_{s_j}(t)$, and any solution that had a norm difference between expected solution and approximated solution less than 0.1 was averaged together and passed into a low pass filter. If no solution had a small enough error, the expected solution was used instead, where the expected $q_{k_j/c}(t)$ and $\underline{u}_{k_j/c}(t)$ were determined using the current pose estimate (i.e., $\hat{p}_{c/k_1}(t)$ and $\hat{q}_{c/k_1}(t)$).

An update to the estimator in (23) was then processed for each feature in $\mathcal{P}_{s_j}(t)$. Using a value of $\lambda_a = 0.3$ for each feature, the state of $\sigma_{s_{j,i}}$ was determined and if $\sigma_{s_{j,i}} = a$, $\mathcal{Y}_{s_{j,i}}(t)$ and $\mathcal{U}_{s_{j,i}}(t)$ from (9) were calculated where the maximum value for ς was 1.0 second. The pose estimate of $\hat{p}_{c/k_1}(t)$ and $\hat{q}_{c/k_1}(t)$ was used to determine $\hat{p}_{k_j/c}(t)$ expressed in the camera frame and $\hat{q}_{k_j/c}(t)$. Using a standard deviation of 3 pixels, the reprojection error between the pixel coordinates of the feature determined from $\underline{u}_{s_{j,i}/c}(t)$ and the pixel coordinates determined using the estimate of $d_{s_{j,i}/k_j}$ from $\mathcal{Y}_{s_{j,i}}(t)$ and $\mathcal{U}_{s_{j,i}}(t)$ was used to calculate the χ^2 value if the estimate of $d_{s_{j,i}/k_j}$ from $\mathcal{Y}_{s_{j,i}}(t)$ and $\mathcal{U}_{s_{j,i}}(t)$ fell within the known bounds, $\underline{d}_2 \leq d_{s_{j,i}/k_j} \leq \bar{d}$, and the value of $\|\mathcal{Y}_{s_{j,i}}(t)\| > 0.1$ and $\|\mathcal{U}_{s_{j,i}}(t)\| > 0.15$, where the thresh-

olds on $\|\mathcal{Y}_{s_{j,i}}(t)\|$ and $\|\mathcal{U}_{s_{j,i}}(t)\|$ and the distance bounds $\underline{d}_2 = 0.5$ meters and $\bar{d} = 5.0$ meters were selected based on the environment. If the χ^2 value was smaller than 6.63, the value of $d_{s_{j,i}/k_j}$ from $\mathcal{Y}_{s_{j,i}}(t)$ and $\mathcal{U}_{s_{j,i}}(t)$ was considered an inlier and the $\mathcal{Y}_{s_{j,i}}(t)$ and $\mathcal{U}_{s_{j,i}}(t)$ pair were added to the history stack in (11) where $N = 50$. Using $\lambda_\tau = 0.0001$, when $\lambda_{min}\{\Sigma_{\mathcal{Y}_{s_{j,i}}}\} > \lambda_\tau$, the value of $\mathcal{X}_{s_{j,i}}$ in (12) was determined and used to update (23), where $K_X = 30I_{3 \times 3}$ was determined to consistently perform well with low noise sensitivity. Additionally, after $\lambda_{min}\{\Sigma_{\mathcal{Y}_{s_{j,i}}}\} > \lambda_\tau$, the value of $\mathcal{X}_{s_{j,i}}$ was also used to estimate the reprojection error and if the χ^2 value was too large when comparing the measured pixel coordinates from $\underline{u}_{s_{j,i}/c}(t)$ to the pixel coordinates determined from $\mathcal{X}_{s_{j,i}}$, the new $\mathcal{Y}_{s_{j,i}}(t)$ and $\mathcal{U}_{s_{j,i}}(t)$ were considered outliers, enabling another method of rejecting noisy measurements. Additionally, the dwell-time for feature $s_{j,i}$ started accumulating time while the switching signal $\sigma_{s_{j,i}} = a$. Once the dwell-time exceeded the initial minimum dwell-time condition in (44), the switching signal $\sigma_{s_{j,i}}^{\Delta t}(t)$ activated (i.e., $\sigma_{s_{j,i}}^{\Delta t}(t) = a$), where $\tilde{\eta}_{s_{j,i}}$ was selected to be 1 centimeter and given $\lambda_{min}\{K_X\} = 30$, $\lambda_a = 0.3$, and $\bar{d} = 5.0$ meters, $\Delta t_{s_j}^{a_1} \geq 0.75$ seconds. When all of the remaining features on the j th object were activated $\sigma_{s_j}^{\Delta t}(t) = a$ and while $\sigma_{s_j}^o(t) = a$ (i.e., at least 20 features on the j th object remained in the FOV), the observer update laws for the camera pose in (56) and (57) where used to estimate the pose. After $\sigma_{s_j}^o(t) = u$, the features on the j th object were no longer tracked and a new set of features were extracted establishing the next object and key frame. After the object was no longer tracked, the predictor update laws in (58) and (59) were used to estimate the pose until the next object satisfied the minimum dwell-time condition. If an object never satisfied the minimum dwell-time condition, that object was never used to update the pose.

The experiment was approximately 560 seconds and consisted of driving the ground vehicle shown in Figure 3 over a path of approximately 250 meters through the series of wooden hallways shown in Figure 4. The true and estimated path over the experiment is shown in Figure 8 where the estimator state estimate is marked in red, green, or blue depending if the landmark is in the FOV, the predictor is activated, or the observer is activated, respectively. The norm of the position error for the experiment is shown in Figure 9 where the error is marked in red, green, or blue depending if the landmark is in the FOV, the predictor is activated, or the observer is activated, respectively. As shown in Figure 9, the RMS error over the entire experiment was 0.58 meters. The maximum error before a reset was 1.18 meters implying the error over path length before reset was approximately 2.8%. The average of the maximums was 1.0 meters implying the average maximum error over path length was approximately 2.4%. Over the experiment there were 246 key frames; however, only 83 key frames satisfied the minimum dwell-time condition and were used in the position observer. The distance to the features was approximately 1.5 meters on average and as shown in the example convergence in Figures 10-12, and the histograms of

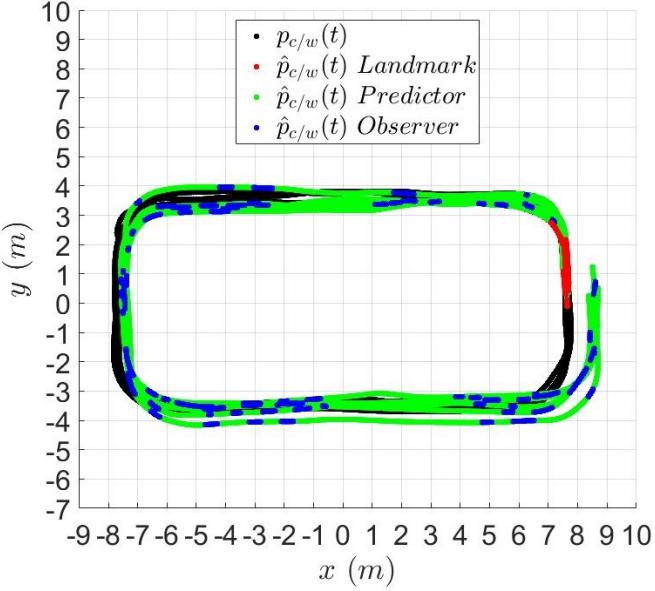


Figure 8. Plot of the path of the camera during the experiment and the estimated path of the camera using a standard deviation of 3 pixels for the history stack rejection algorithm. The true path is marked in black. The estimated path shows which estimator is activated over the experiment. While the landmark is in the FOV, the estimated path is shown using the red marker. Similarly, the estimated path is marked in green or blue when the predictor or observer is active, respectively.

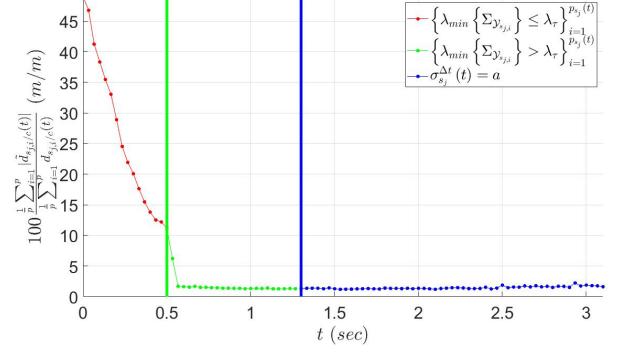


Figure 10. Plot of the distance estimator convergence for key frame 33 that shows the average percentage error of the distance to the features relative to the true distance to the features using a standard deviation of 3 pixels for accepting data onto the history stack. The plot shows while the set of estimators have not satisfied the eigenvalue condition, marked in red, $\Psi_{\xi_{s_{j,i}}}(t)$ enables convergence. When enough data has been collected and the eigenvalue condition is satisfied, marked by the green vertical line and green markers, the error demonstrates exponential decay. After the initial minimum dwell-time condition is satisfied, marked by the vertical blue line and blue markers, the set of features was used in the position observer. As shown, the error percentage relative to the distance is approximately 1.6%.

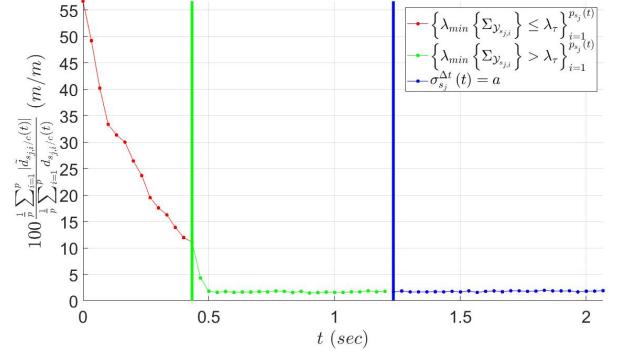


Figure 11. Plot of the distance estimator convergence for key frame 173 that shows the average percentage error of the distance to the features relative to the true distance to the features using a standard deviation of 3 pixels for accepting data onto the history stack. The plot shows while the set of estimators have not satisfied the eigenvalue condition, marked in red, $\Psi_{\xi_{s_{j,i}}}(t)$ enables convergence. When enough data has been collected and the eigenvalue condition is satisfied, marked by the green vertical line and green markers, the error demonstrates exponential decay. After the initial minimum dwell-time condition is satisfied, marked by the vertical blue line and blue markers, the set of features was used in the position observer. As shown, the error percentage relative to the distance is approximately 2.3%.

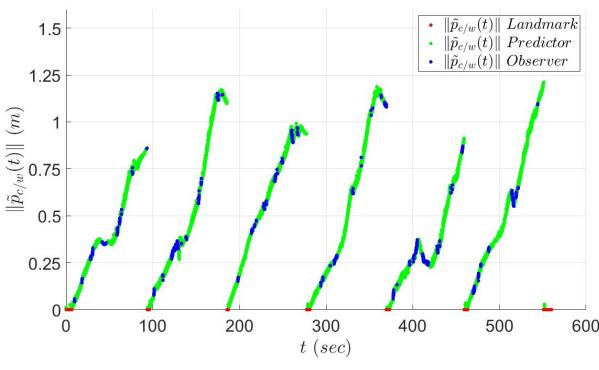


Figure 9. Plot of the norm of the camera position error during the experiment using a standard deviation of 3 pixels for accepting data onto the history stack. The camera position error shows which estimator is activated over the experiment. While the landmark is in the FOV, the estimated path is shown using the red marker. Similarly, the estimated path is marked in green or blue when the predictor or observer is active, respectively. As shown, the error resets to zero each time the landmark enters the FOV. The maximum position error was approximately 1.18 meters while the average of the maximums was 1.0 meters and the RMS of the position error was 0.58 meters.

the RMS error (i.e., $\left\{ \text{RMS} \left(100 \frac{\frac{1}{p} \sum_{i=1}^p |\tilde{d}_{s_{j,i}/c}(t)|}{\frac{1}{p} \sum_{i=1}^p d_{s_{j,i}/c}(t)} \right) \right\}_{j=1}^{p_{s_j}(t)}$) in

Figures 13-15, the average feature distance error of the 83 key frames was at its lowest when used by the position estimator as was predicted by the minimum dwell-time. Specifically, the histograms showed that before the dwell-time condition is satisfied the percentage error was 19.4% on average as shown in Figure 14; however, after the minimum dwell-time condition was satisfied the error was 4.2 percent on average but had a median error of 3.4%.

Similar to our previous work in [1] and [2], the developed position estimator strategy is to ensure the error in the position

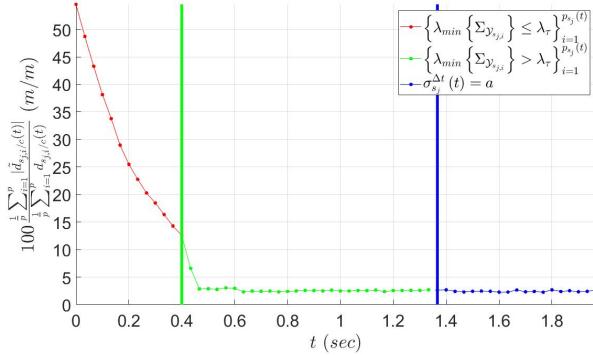


Figure 12. Plot of the distance estimator convergence for key frame 215 that shows the average percentage error of the distance to the features relative to the true distance to the features using a standard deviation of 3 pixels for accepting data onto the history stack. The plot shows while the set of estimators have not satisfied the eigenvalue condition, marked in red, $\Psi_{\xi_{s_{j,i}}}(t)$ enables convergence. When enough data has been collected and the eigenvalue condition is satisfied, marked by the green vertical line and green markers, the error demonstrates exponential decay. After the initial minimum dwell-time condition is satisfied, marked by the vertical blue line and blue markers, the set of features was used in the position observer. As shown, the error percentage relative to the distance is approximately 1.9%.

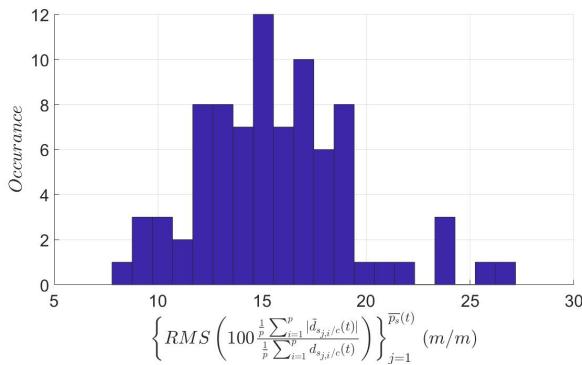


Figure 13. Histogram plot of the RMS of the average percentage error of the distance to the features relative to the true distance to the features across the entire experiment using a standard deviation of 3 pixels for accepting data onto the history stack. The histogram shows the RMS errors over the entire time a key frame tracked. The RMS error was on average of 15.6% with a standard deviation of 3.7% and a median error of 15.4% over the entire time a key frame was tracked.

does not exceed a desired threshold through the development of dwell-time conditions. However, [1] and [2] use a predictor-only strategy when feedback on the position is not directly available. The purpose of the developed estimator in this paper is to improve upon a predictor-only estimator and guarantee the error in the position estimate grows at a slower rate when a vehicle is operating in an environment with no feedback from landmarks or a positioning system. This is achieved through the use of the position observer strategy when features satisfy the minimum dwell-time condition. As shown in Figures 16 and 17, using a predictor-only strategy, similar to [1] and [2], results in larger maximum position error compared to using the predictor and observer strategy. Specifically, the position error using a predictor-only strategy had an overall RMS position error that was 8% larger (i.e., 0.63 meters for the predictor-

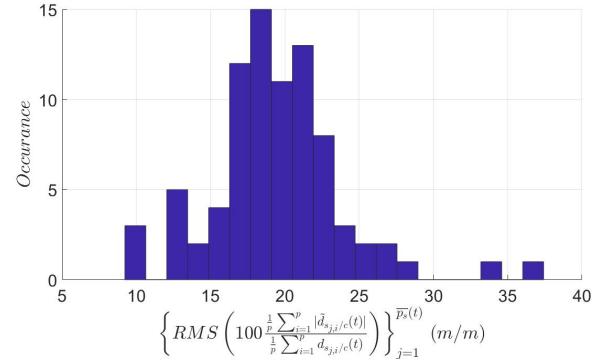


Figure 14. Histogram plot of the RMS of the average percentage error of the distance to the features relative to the true distance to the features across before the minimum dwell-time condition is satisfied using a standard deviation of 3 pixels for accepting data onto the history stack. The histogram shows the RMS errors over the time from extracting the features from a key frame to the time just before the minimum dwell-time condition is satisfied. The RMS error was on average of 19.4% with a standard deviation of 4.5% and a median error of 19.1% before the minimum dwell-time condition was satisfied.

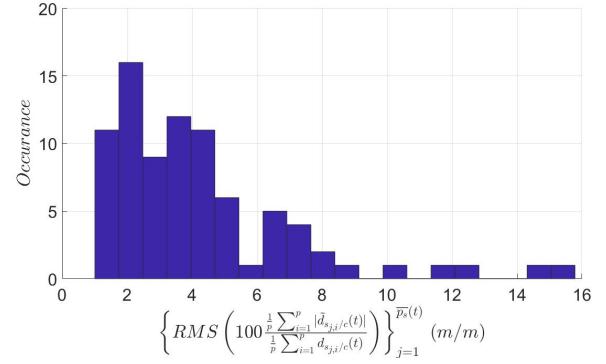


Figure 15. Histogram plot of the RMS of the average percentage error of the distance to the features relative to the true distance to the features across after the minimum dwell-time condition is satisfied using a standard deviation of 3 pixels for accepting data onto the history stack. The histogram shows the RMS errors over the time the minimum dwell-time condition is satisfied to the time a key frame was no longer tracked. The RMS error was on average of 4.2% with a standard deviation of 2.9% and a median error of 3.4% after the minimum dwell-time condition was satisfied.

only strategy while the predictor and observer strategy was 0.58 meters) and an average maximum that was 8.3% larger (i.e., 1.09 meters for the predictor-only strategy compared to 1.0 meters using the predictor and observer strategy). These experimental results demonstrate that the estimator strategy using the minimum dwell-time condition ensures that only features with low error are used in the position observer. The result of this paper enables the position estimation error to remain smaller compared to using a predictor-only strategy which enables a vehicle to operate in an environment with no feedback from landmarks or a positioning system for longer periods of time.

Remark 3. Reducing the standard deviation for accepting data onto the history stack improves the overall performance of both the distance estimators and the position estimator as demonstrated by comparing Figures 8, 9, and 15 to Figures 18-20. Specifically, the maximum position error using a 3

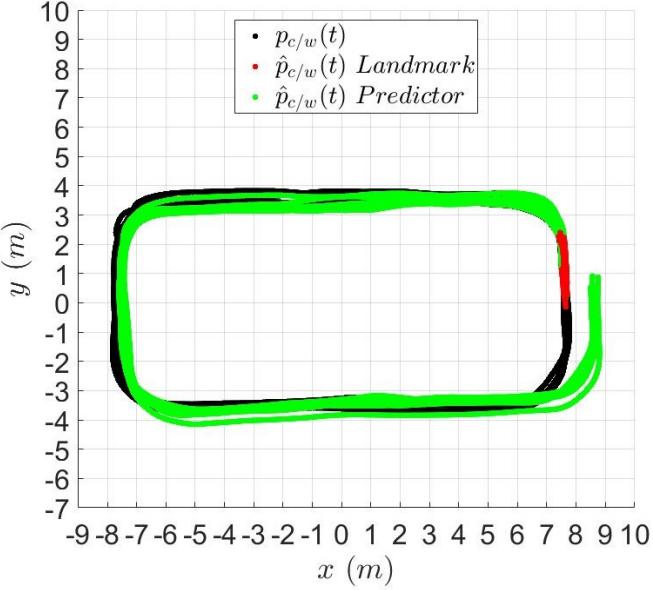


Figure 16. Plot of the path of the camera during the experiment and the estimated path of the camera using the predictor-only strategy when no landmark is in the FOV. The true path is marked in black. The estimated path shows if the landmark is in the FOV or the predictor is activated. While the landmark is in the FOV, the estimated path is shown using the red marker. Similarly, the estimated path is marked in green when the predictor is active.

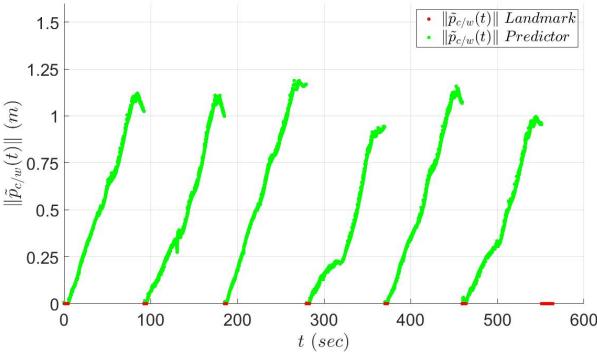


Figure 17. Plot of the norm of the camera position error during the experiment using the predictor-only strategy when no landmark is in the FOV. The camera position error shows which estimator is activated over the experiment. While the landmark is in the FOV, the estimated path is shown using the red marker. Similarly, the estimated path is marked in green when the predictor is active. As shown, the error resets to zero each time the landmark enters the FOV. The maximum position error was approximately 1.18 meters while the average of the maximums was 1.09 meters and the RMS of the position error was 0.63 meters.

pixel standard deviation for accepting data was on average 1.0 meters compared to 1.21 meters when using a 10 pixel standard deviation for accepting data. This shows that relaxing the threshold to accept data causes more error to be injected into the system reducing performance compared to the predictor-only strategy which had an average maximum error of 1.1 meters. Similarly, the RMS error of the distances after satisfying the minimum dwell-time was 4.2% on average with a standard deviation of 2.9% and a median error of 3.4% for the 3 pixel standard deviation for accepting data compared

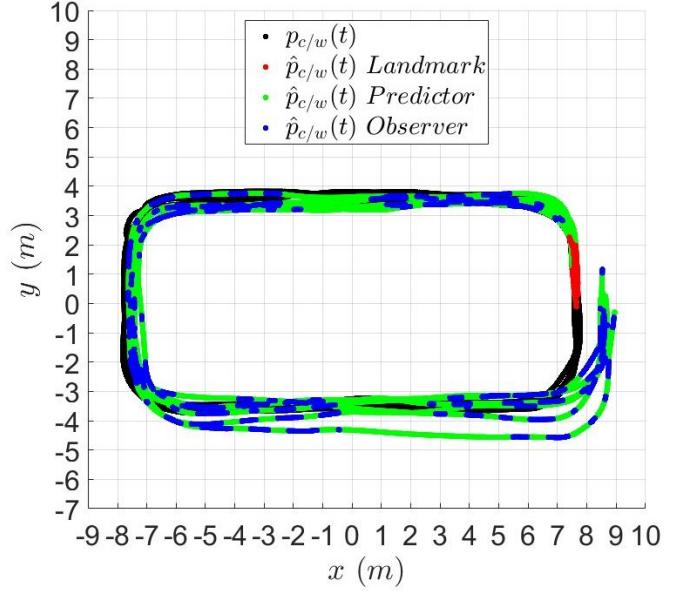


Figure 18. Plot of the path of the camera during the experiment and the estimated path of the camera using a standard deviation of 10 pixels for the history stack rejection algorithm. The true path is marked in black. The estimated path shows which estimator is activated over the experiment. While the landmark is in the FOV, the estimated path is shown using the red marker. Similarly, the estimated path is marked in green or blue when the predictor or observer is active, respectively.

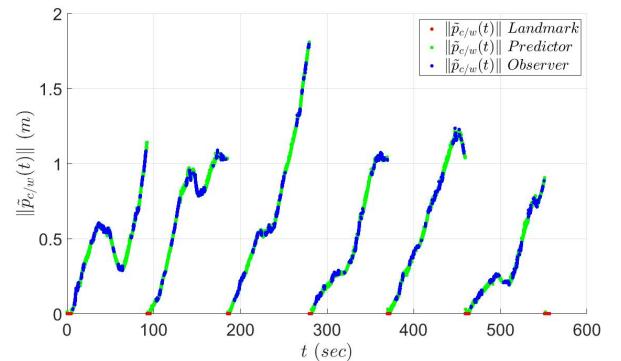


Figure 19. Plot of the norm of the camera position error during the experiment using a standard deviation of 10 pixels for accepting data onto the history stack. The camera position error shows which estimator is activated over the experiment. While the landmark is in the FOV, the estimated path is shown using the red marker. Similarly, the estimated path is marked in green or blue when the predictor or observer is active, respectively. As shown, the error resets to zero each time the landmark enters the FOV. The maximum position error was approximately 1.81 meters while the average of the maximums was 1.21 meters and the RMS of the position error was 0.65 meters.

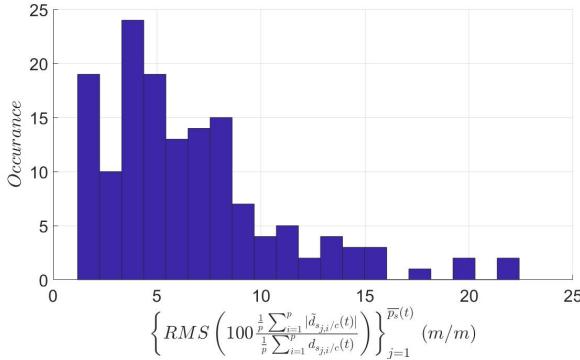


Figure 20. Histogram plot of the RMS of the average percentage error of the distance to the features relative to the true distance to the features across after the minimum dwell-time condition is satisfied using a standard deviation of 10 pixels for accepting data onto the history stack. The histogram shows the histogram of the RMS errors over the time the minimum dwell-time condition is satisfied to the time a key frame was no longer tracked. The RMS error was on average of 6.6% with a standard deviation of 4.3% and a median error of 5.5% after the minimum dwell-time condition was satisfied.

to using a 10 pixel standard deviation for accepting data which had a 6.6% average RMS error with a standard deviation of 4.3% and a median of 5.5%. However, reducing the threshold results in less features satisfying the minimum dwell-time given 147 key frames out of 246 satisfied the minimum dwell-time condition when using the 10 pixel standard deviation for accepting data compared to 83 key frames out of 246 when using the 3 pixel standard deviation for accepting data. Since the objective is to provide better position estimates when the landmark is not in the camera's FOV, this trade-off is acceptable; however, if the goal was to estimate more of the environment then allowing for a higher standard deviation for accepting data would be acceptable. Additionally, if the resulting structure of all the objects and the resulting path of the camera were passed into an optimization algorithm implementing bundle adjustments, the result may enable a richer estimate of the environment. An optimization could be applied regardless; however, having more features will result in a more dense estimate of the environment.

VIII. CONCLUSION

This paper develops and tests a novel estimator strategy using a single monocular camera and structure from motion theory to estimate the position of a vehicle through a feedback-denied region by estimating the Euclidean distance to features on stationary objects in the immediate environment. The developed estimation strategy demonstrates the position estimation error is GUUB through the use of minimum dwell-times determined using a Lyapunov-based stability analysis. Specifically, a minimum dwell-time condition is developed that theoretically ensures only features having small structure estimation error are used to estimate the position of the camera. Furthermore, the developed strategy doesn't require an object to remain in the FOV enabling the camera to travel over large distances, doesn't require new features to be in the FOV when old features leave the FOV, and does not require the positive depth constraint relaxing constraints on

motion of the vehicle. An experimental analysis is presented that demonstrates the effectiveness of this estimation strategy compared to a predictor-only strategy in a feedback-denied region by maintaining lower RMS position estimation error. Furthermore, the RMS error in feature structure estimates is presented and shown to be at its smallest after satisfying the minimum dwell-time condition at which time it is used by the position estimator.

REFERENCES

- [1] H.-Y. Chen, Z. I. Bell, P. Deptula, and W. E. Dixon, "A switched systems framework for path following with intermittent state feedback," *IEEE Control Syst. Lett.*, vol. 2, no. 4, pp. 749–754, Oct. 2018.
- [2] H.-Y. Chen, Z. Bell, P. Deptula, and W. E. Dixon, "A switched systems approach to path following with intermittent state feedback," *IEEE Trans. Robot.*, vol. 35, no. 3, pp. 725–733, 2019.
- [3] 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.
- [4] J. Engel, T. Schöps, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," in *Computer Vision – ECCV 2014*, 2014, pp. 834–849.
- [5] G. Dubbelman and B. Browning, "Cop-slam: Closed-form online pose-chain optimization for visual slam," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1194–1213, Oct 2015.
- [6] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "Orb-slam: A versatile and accurate monocular slam system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, Oct 2015.
- [7] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgbd cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, Oct 2017.
- [8] M. Karrer, P. Schmuck, and M. Chli, "Cvi-slam collaborative visual-inertial slam," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 2762–2769, Oct 2018.
- [9] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.
- [10] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry, *An Invitation to 3-D Vision*. Springer, 2004.
- [11] L. Matthies, T. Kanade, and R. Szeliski, "Kalman filter-based algorithm for estimating depth from image sequences," *Int. J. Comput. Vision*, vol. 3, pp. 209–236, 1989.
- [12] M. Jankovic and B. 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.
- [13] S. Soatto, R. Frezza, and P. Perona, "Motion estimation via dynamic vision," *IEEE Trans. Autom. Control*, vol. 41, no. 3, pp. 393–413, 1996.
- [14] H. Kano, B. K. Ghosh, and H. Kanai, "Single camera based motion and shape estimation using extended Kalman filtering," *Math. Comput. Model.*, vol. 34, pp. 511–525, 2001.
- [15] 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.
- [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, pp. 2232–2238, 2003.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] G. Hu, D. Aiken, S. Gupta, and W. Dixon, "Lyapunov-based range identification for a paracatadioptric system," *IEEE Trans. Autom. Control*, vol. 53, no. 7, pp. 1775–1781, 2008.

- [22] A. Mourikis, N. Trawny, S. Roumeliotis, A. Johnson, A. Ansar, and L. Matthies, "Vision-aided inertial navigation for spacecraft entry, descent, and landing," *IEEE Trans. Robot.*, vol. 25, no. 2, pp. 264–280, Apr. 2009.
- [23] F. Morbidi and D. Prattichizzo, "Range estimation from a moving camera: an immersion and invariance approach," in *Proc. IEEE Int. Conf. Robot. Autom.*, Kobe, Japan, May 2009, pp. 2810–2815.
- [24] 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.
- [25] A. Dani, N. Fischer, Z. Kan, and W. E. Dixon, "Globally exponentially stable observer for vision-based range estimation," *Mechatron.*, vol. 22, no. 4, pp. 381–389, Special Issue on Visual Servoing 2012.
- [26] A. Dani, N. Fischer, and W. E. Dixon, "Single camera structure and motion," *IEEE Trans. Autom. Control*, vol. 57, no. 1, pp. 241–246, Jan. 2012.
- [27] 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*, 2017, pp. 3670–3675.
- [28] J. Oliensis, "A critique of structure-from-motion algorithms," *Comput. Vis. Image Understand.*, vol. 80, pp. 172–214, 2000.
- [29] J. Oliensis and R. Hartley, "Iterative extensions of the strum/triggs algorithm: convergence and nonconvergence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2217–2233, 2007.
- [30] F. Kahl and R. Hartley, "Multiple-view geometry under the L_∞ -norm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1603–1617, Sep. 2008.
- [31] A. Parikh, T.-H. Cheng, H.-Y. Chen, and W. E. Dixon, "A switched systems framework for guaranteed convergence of image-based observers with intermittent measurements," *IEEE Trans. Robot.*, vol. 33, no. 2, pp. 266–280, April 2017.
- [32] A. Parikh, T.-H. Cheng, R. Licitra, and W. E. Dixon, "A switched systems approach to image-based localization of targets that temporarily leave the camera field of view," *IEEE Trans. Control Syst. Technol.*, vol. 26, no. 6, pp. 2149–2156, 2018.
- [33] 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, 2018.
- [34] 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. on Autom. Control*, vol. 42, no. 4, pp. 581–586, 1997.
- [35] 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.
- [36] Z. Bell, P. Deptula, H.-Y. Chen, E. Doucette, and W. E. Dixon, "Velocity and path reconstruction of a moving object using a moving camera," in *Proc. Am. Control Conf.*, 2018, pp. 5256–5261.
- [37] 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.
- [38] G. Chowdhary, T. Yucelen, M. Mühllegg, 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.
- [39] G. Chowdhary, M. Mühllegg, J. How, and F. Holzapfel, "Concurrent learning adaptive model predictive control," in *Advances in Aerospace Guidance, Navigation and Control*, Q. Chu, B. Mulder, D. Choukroun, E.-J. van Kampen, C. de Visser, and G. Looye, Eds. Springer Berlin Heidelberg, 2013, pp. 29–47.
- [40] R. Kamalapurkar, P. Walters, and W. E. Dixon, "Model-based reinforcement learning for approximate optimal regulation," *Automatica*, vol. 64, pp. 94–104, 2016.
- [41] R. Licitra, Z. I. Bell, E. Doucette, and W. E. Dixon, "Single agent indirect herding of multiple targets: A switched adaptive control approach," *IEEE Control Syst. Lett.*, vol. 2, no. 1, pp. 127–132, January 2018.
- [42] R. Licitra, Z. Bell, and W. Dixon, "Single agent indirect herding of multiple targets with unknown dynamics," *IEEE Trans. Robotics*, vol. 35, no. 4, pp. 847–860, 2019.
- [43] A. Parikh, R. Kamalapurkar, and W. E. Dixon, "Integral concurrent learning: Adaptive control with parameter convergence using finite excitation," *Int J Adapt Control Signal Process.*, to appear.
- [44] Z. Bell, J. Nezadovitz, A. Parikh, E. Schwartz, and W. Dixon, "Global exponential tracking control for an autonomous surface vessel: An integral concurrent learning approach," *IEEE J. Ocean Eng.*, to appear.
- [45] G. Hu, S. Mehta, N. Gans, and W. E. Dixon, "Daisy chaining based visual servo control part I: Adaptive quaternion-based tracking control," in *IEEE Multi-Conf. Syst. and Contr.*, Suntec City, Singapore, Oct. 2007, pp. 1474–1479.
- [46] G. Hu, N. Gans, S. Mehta, and W. E. Dixon, "Daisy chaining based visual servo control part II: Extensions, applications and open problems," in *Proc. IEEE Conf. Control Appl.*, Suntec City, Singapore, Oct. 2007, pp. 729–734.
- [47] R. Sim, P. Elinas, and J. J. Little, "A study of the rao-blackwellised particle filter for efficient and accurate vision-based slam," *Int. J. Comput. Vision*, vol. 74, no. 3, pp. 303–318, 2007.
- [48] T. Lemaire, C. Berger, I.-K. Jung, and S. Lacroix, "Vision-based slam: Stereo and monocular approaches," *Int. J. Comput. Vision*, vol. 74, no. 3, pp. 343–364, Sep. 2007.
- [49] S. S. Mehta, T. Burks, and W. E. Dixon, "A theoretical model for vision-based localization of a wheeled mobile robot in greenhouse applications: A daisy chaining approach," *Computers Electron. Agric.*, vol. 63, pp. 28–37, 2008.
- [50] K. Dupree, N. R. Gans, W. MacKunis, , and W. E. Dixon, "Euclidean calculation of feature points of a rotating satellite: A daisy chaining approach," *AIAA J. Guid. Control Dyn.*, vol. 31, pp. 954–961, 2008.
- [51] F. Bonin-Font, A. Ortiz, and G. Oliver, "Visual navigation for mobile robots: A survey," *J. Intell. Rob. Syst.*, vol. 53, no. 3, pp. 263–296, Nov. 2008.
- [52] M. Kaiser, N. Gans, and W. Dixon, "Vision-based estimation for guidance, navigation, and control of an aerial vehicle," *IEEE Transactions on aerospace and electronic systems*, vol. 46, no. 3, 2010.
- [53] S. Y. Chen, "Kalman filter for robot vision: A survey," *IEEE Trans. Ind. Electron.*, vol. 59, no. 11, pp. 4409–4420, Aug. 2011.
- [54] T. Wang, C. Wang, J. Liang, Y. Chen, and Y. Zhang, "Vision-aided inertial navigation for small unmanned aerial vehicles in gps-denied environments," *Int. J. Adv. Robot. Syst.*, vol. 10, no. 6, 2013.
- [55] D. Lee, Y. Kim, and H. Bang, "Vision-based terrain referenced navigation for unmanned aerial vehicles using homography relationship," *J. Intell. Robot. Syst.*, vol. 69, no. 1, pp. 489–497, Jan 2013.
- [56] T. Taketomi, H. Uchiyama, and S. Ikeda, "Visual slam algorithms: A survey from 2010 to 2016," *IPSJ Transactions on Computer Vision and Applications*, vol. 9, no. 1, p. 16, 2017.
- [57] B. M. Haralick, C.-N. Lee, K. Ottenberg, and M. Nölle, "Review and analysis of solutions of the three point perspective pose estimation problem," *Int. J. Comput. Vision*, vol. 13, no. 3, pp. 331–356, Dec. 1994.
- [58] D. DeMenthon and L. Davis, "Model-based object pose in 25 lines of code," *Int. J. Comput. Vision*, vol. 15, pp. 123–141, 1995.
- [59] L. Quan and Z. Lan, "Linear N-point camera pose determination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 8, pp. 774–780, 1999.
- [60] X.-S. Gao, X.-R. Hou, J. Tang, and H.-F. Cheng, "Complete solution classification for the perspective-three-point problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 930–943, 2003.
- [61] J. Lewis, "Fast template matching," *Vis. Interface*, pp. 120–123, 1995.
- [62] D. Lowe, "Distinctive image features from scale-invariant keypoints," in *IJCV*, 2004.
- [63] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [64] T. Tuytelaars and K. Mikolajczyk, *Local Invariant Feature Detectors: A Survey*, ser. Foundations and Trends in Computer Graphics and Vision. Now Publishers Inc, 2008, vol. 10.
- [65] E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski, "Orb: An efficient alternative to sift or surf," in *Proc. IEEE Int. Conf. Comput. Vision*, 2011, pp. 2564–2571.
- [66] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1409–1422, Jul. 2012.
- [67] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vision*, pp. 1–42, 2015.
- [68] 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.
- [69] C. Harris and M. Stephens, "A combined corner and edge detector," in *Alvey Vis. Conf.*, vol. 15, 1988, pp. 147–151.
- [70] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1994, pp. 593–600.

- [71] 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.
- [72] K. Fathian, J. P. Ramirez-Paredes, E. A. Doucette, J. W. Curtis, and N. R. Gans, “Quest: A quaternion-based approach for camera motion estimation from minimal feature points,” *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 857–864, April 2018.
- [73] G. Chowdhary and E. Johnson, “A singular value maximizing data recording algorithm for concurrent learning,” in *Proc. Am. Control Conf.*, 2011, pp. 3547–3552.
- [74] 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, July 2017.
- [75] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2002.
- [76] G. Guennebaud, B. Jacob *et al.*, “Eigen v3,” <http://eigen.tuxfamily.org>, 2010.
- [77] G. Bradski, “The OpenCV Library,” *Dr. Dobb's Journal of Software Tools*, 2000.
- [78] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, “Ros: an open-source robot operating system,” in *ICRA workshop on open source software*, vol. 3, no. 3.2. Kobe, Japan, 2009, p. 5.