

TARGET TRACKING IN UNKNOWN ENVIRONMENTS USING A MONOCULAR
CAMERA SUBJECT TO INTERMITTENT SENSING

By

ZACHARY IAN BELL

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2019

© 2019 Zachary Ian Bell

To my fiancée Michelle Marie Peters, I could not have made it this far without your love
and support

ACKNOWLEDGMENTS

I thank Dr. Warren E. Dixon, who has helped me tremendously improve my capacity as a researcher. I extend my gratitude to my committee members Dr. Carl Crane, Dr. Prabir Barooah, and Dr. John Shea for their oversight and recommendations. I appreciate all the time I have spent working on exciting research with all the people in the Nonlinear Controls and Robotics group at the University of Florida.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS	4
LIST OF TABLES	7
LIST OF FIGURES	8
ABSTRACT	12
CHAPTER	
1 INTRODUCTION	16
1.1 Background	16
1.2 Outline of the Dissertation	22
2 SYSTEM MODEL	25
2.1 Motion Model Using Stationary Features	25
2.2 Extension of Motion Model Using Stationary Features	28
2.3 Motion Model of Camera	32
2.4 Motion Model of Moving Features	35
3 SIMULTANEOUS ESTIMATION OF EUCLIDEAN DISTANCES TO A STA- TIONARY OBJECT'S FEATURES AND THE EUCLIDEAN TRAJECTORY OF A MONOCULAR CAMERA	41
3.1 Integral Concurrent Learning Observer Update Laws for Euclidean Dis- tances	42
3.2 Extended Observer Update Law for Euclidean Distance to Features from Camera	45
3.3 Stability Analysis	46
3.4 Experimental Results	47
3.5 Summary	54
4 POSITION ESTIMATION USING A MONOCULAR CAMERA SUBJECT TO INTERMITTENT FEEDBACK	56
4.1 Learning Feature Structure	57
4.2 Feature Observer Design Without Object Return	60
4.2.1 Feature Observer Design	61
4.2.2 Observer Design Stability Analysis	63
4.3 Feature Observer Design With Object Return	65
4.3.1 Feature Predictor Design	65
4.3.2 Stability Analysis of Feature Predictor Design	67
4.3.3 Ensuring Stability Through Dwell-Time Conditions	68
4.4 Estimator Design for Pose of Camera	71

4.4.1	Stability of Key Frame Position Observer and Predictor Design	75
4.5	Experiments	76
4.6	Summary	100
5	STRUCTURE AND MOTION OF A MOVING TARGET USING A MOVING MONOCULAR CAMERA SUBJECT TO INTERMITTENT FEEDBACK	102
5.1	Learning the First Feature Structure	102
5.2	Learning the Structure of the Remaining Features	107
5.3	Learning the Object Motion Model	110
5.4	Target Estimators	115
5.5	Object Observer and Predictor Analysis	117
5.6	Object Dwell-Time Analysis	120
5.7	Summary	123
6	CONCLUSIONS	124
	REFERENCES	128
	BIOGRAPHICAL SKETCH	138

LIST OF TABLES

<u>Table</u>	<u>page</u>
3-1 RMS Depth Error and Position Error in Meters Over 15 Experiments	51
3-2 RMS Depth Error and Position Error in Meters Over 15 Experiment Before Learning	53
3-3 RMS Depth Error and Position Error in Meters Over 15 Experiment After Learning	53

LIST OF FIGURES

<u>Figure</u>	<u>page</u>
2-1 Example geometry for tracking the position of the i th feature of s where the camera starts at the top left where the key image is taken and is traveling downward from the upper left to the lower left while tracking a stationary object on the right.	26
2-2 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.	29
2-3 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.	33
2-4 Example geometry for tracking the position of the i th feature of m and m^*	36
3-1 Photo courtesy of author. Image shows the checkerboard, Kobuki Turtlebot, and an iDS uEye camera used for experiments.	48
3-2 The camera trajectories for each of the 15 experiments.	50
3-3 The sum of the norm of each depth error across the 48 features (i.e., $\sum_{s_i=1}^{48} \ \tilde{z}_{s_i/c}(t)\ $) for Experiment 11. Estimator 1 (red) refers to (3–10), Estimator 2 (magenta) refers to (3–17), Estimator 3 (green) refers to [1], and Estimator 4 (blue) refers to the EKF. The black vertical line indicates the time when enough information was collected for learning.	51
3-4 The position error of the camera and the distance error using the estimator in (3–11) for Experiment 11.	52
4-1 Photo courtesy of author. Image of the Kobuki Turtlebot and iDS uEye camera used for experiments.	77
4-2 Photo courtesy of author. 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.	78
4-3 Photo courtesy of author. 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.	79
4-4 Photo courtesy of author. 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.	80

4-5	Photo courtesy of author. Image of the features extracted from a key frame image of a planar object, where the white dots are the extracted corner features.	81
4-6	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 are active, respectively.	85
4-7	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 are 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.	86
4-8	Plot of the distance estimator convergence for key frame 33 by showing 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, the flow type term 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%.	87
4-9	Plot of the distance estimator convergence for key frame 173 by showing 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, the flow type term 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%.	88

4-10 Plot of the distance estimator convergence for key frame 215 by showing 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, the flow type term 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%. 89

4-11 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 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. 90

4-12 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. 91

4-13 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 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. 92

4-14 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. 94

4-15	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.	95
4-16	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 are active, respectively.	97
4-17	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 are 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.	98
4-18	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.	99
5-1	Example geometry for a simplified camera with origin at c , angle α_c , and FOV \mathcal{V}_c . The maximum radius of an inscribed sphere in the camera's FOV at a distance \bar{d} is $R_\eta \triangleq \bar{d} \sin(\alpha_c)$	120

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

TARGET TRACKING IN UNKNOWN ENVIRONMENTS USING A MONOCULAR
CAMERA SUBJECT TO INTERMITTENT SENSING

By

Zachary Ian Bell

December 2019

Chair: Warren E. Dixon

Major: Mechanical Engineering

In most applications involving autonomous agents tracking a moving target through uncertain environments, it is necessary to estimate the structure of local features in the environment (e.g., relative positions of objects in the immediate surrounding environment), the pose of an agent (i.e., position and orientation), and the pose and velocity of the target. Many of these applications require traveling over large distances implying the local environment for an agent is always changing, introducing further difficulty. Furthermore, it is often only possible to intermittently sense the target (e.g., environmental obstructions or path constraints of the agent may cause occlusions of the target).

It is often assumed that it is possible to use global sensing to measure the state of an agent. However, state feedback generally requires a sensor that can relate all the states to a common coordinate system (e.g., global positioning system (GPS)). However, GPS may be unavailable (e.g., agents could operate in environments where GPS is restricted or denied). Furthermore, assuming that the entire environment is known and state information from the target is available is a restrictive assumption since targets are not likely to communicate such information and directly sensing the pose and velocity of a target is challenging and not possible in many scenarios. These challenges motivate the development of techniques that rely on local sensing but still allow agents to estimate their own state (i.e., pose) as well as the state of a target (i.e.,

pose and velocity). Additionally, efforts are motivated by the fact that local sensing often has intermittent availability.

Cameras are a potential sensor that can provide local feedback of the environment where coordinates of the target can be related to a common reference frame; however, camera systems don't inherently measure scale, have a limited field-of-view, and are susceptible to intermittent sensing (e.g., due to occlusions). The scale of the Euclidean coordinates of features in an image, (also known as the structure of the features) is not available because images are a two dimensional projection of a Euclidean environment. Monocular camera systems can recover scale by moving the camera (e.g., structure from motion) and tracking features over large periods of time; however, as the distance to a target increases, the distance between views of the feature must also increase making an accurate scale estimate challenging. Additionally, the camera's limited field-of-view can inhibit continuous observation of a specific object. Continuous observation can also be disrupted by occlusions or trajectory constraints that may require an agent to purposefully allow the target to leave its field-of-view periodically. Furthermore, agents may need to track the target over large distances requiring the agent to continuously reconstruct new objects from the global environment when reconstructed objects in the local environment permanently leave their field-of-view.

In Chapter 3, a global exponentially stable observer for feature scale is developed under a finite excitation condition through the use of integral concurrent learning. Since the observer only requires finite excitation to be globally exponentially stable, the observer is more general than previous results. The result indicates that the Euclidean distance to a set of features on a stationary object and the path the camera travels while viewing that object are estimated exponentially fast implying the structure and path are reconstructed exponentially. Furthermore, the developed estimation method does not require the features on the objects to be planar and does not require the positive depth constraint. An experimental study is presented which compares the developed

Euclidean distance observer to previous observers demonstrating the effectiveness of this result.

In Chapter 4, an extension to the learning approaches in Chapter 3 is developed that applies a new learning strategy that maintains a continuous estimate of the position of the camera and estimates the structure of features as they come into the camera's field-of-view. Furthermore, the developed learning strategy allows simulated measurements of features from objects that are no longer in the field-of-view enabling a continuous estimate of the distance to features with respect to the camera. Additionally, this approach shows how the extended observer removes the positive depth constraint required by all previous structure from motion approaches. Using this approach, a camera may travel over large distances without keeping specific features in the field-of-view for all time and allow objects to permanently leave the field-of-view if necessary. A Lyapunov based stability analysis proves that the observers for estimating the path of the camera as well as the structure of each set of objects are globally exponentially stable while features are in the camera's field-of-view. A switched systems analysis is used to develop dwell-time conditions to indicate how long a feature must be tracked to ensure the distance estimation error is below a threshold. After the distance estimates have converged below the threshold, the feature may be used to update the position of the camera. If a feature does not satisfy the dwell-time condition, it is never used to update the position of the agent. Furthermore, the approach does not require a new set of features to be in the camera's FOV when older features leave the camera's FOV.

In Chapter 5, the approach in Chapter 4 is used to provide pose estimates of the camera and an extension of Chapter 3 is developed to exponentially estimate the pose, velocity, and acceleration of the moving target. Specifically, using the pose and velocity of the camera, the estimation error of the Euclidean trajectory of the target as well as the structure of the target, is globally exponentially convergent to an ultimate bound assuming the target velocity and acceleration are bounded and dwell-time conditions

are satisfied. The developed estimator relaxes the requirement to have continuous observation of the target, to know the exact structure, velocity, or acceleration of the target, and does not require the persistence of excitation assumption or positive depth constraint.

Chapter 6 concludes the dissertation with a discussion of the developed estimation algorithms and potential extensions.

CHAPTER 1 INTRODUCTION

1.1 Background

In many applications, the state (e.g., position and orientation) of an autonomous agent and its local environment (e.g., relative positions of objects in the surrounding environment) must be determined from sensor data. This problem is well known as simultaneous localization and mapping (SLAM) (cf., [2–14]). Often, a global positioning system (GPS) is used to estimate the position; however, in many environments GPS is unavailable (e.g., when agents operate in GPS denied or contested environments) motivating the use of only local sensing data (e.g., camera images, inertial measurement units, and wheel encoders) to estimate the position and model the surrounding environment. In applications involving tracking a moving target through uncertain environments, it is additionally necessary to estimate the pose and velocity of the target. Many of these applications require traveling over large distances implying the local environment for an agent is always changing introducing further difficulty. These challenges motivate the development of techniques that rely on local sensing but still allow agents to estimate their own state (i.e., pose) as well as the state of a target (i.e., pose and velocity). Additionally, efforts are motivated by the fact that local sensing often has intermittent availability.

Using cameras to reconstruct the surrounding environment (i.e., determine the Euclidean scale of objects in the environment) requires an assumption that object features are in the camera field-of-view (FOV) and may be extracted and tracked through a sequence of images. However, a significant challenge arises in determining the scale of objects in an image using a camera given the loss of depth information. Specifically, images of objects are 2D projections of the 3D environment. Approaches to reconstruct (i.e., estimate the structure) objects use multiple images of an object along with scale information (cf., [15, 16]) or motion (cf., [1, 17–32]), 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 objects are not known; however, multiple calibrated cameras may be used to recover the scale (cf., [15, 16]). However, this approach is inapplicable in all scenarios 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 may travel to generate enough parallax, which is generally not possible in stereo vision.

The SfM problem may be approached using online iterative methods (cf., [1, 17–32]) and offline batch methods (cf., [15, 16, 33] and the references contained within). These offline approaches perform an optimization over an image sequence, but only show convergence for limited cases (cf., [34, 35]). Most online SfM approaches assume continuous measurements of objects by the camera or only update when a new image is received (cf., [1, 17–27, 29–32]); however, recent results enable objects to temporarily leave the camera’s FOV (cf., [36–38]). Many results apply the extended Kalman filter (EKF) to estimate depth, (cf., [17, 19–21, 28]); however, the EKF generally does not guarantee convergence and may fail in some applications where the system is not sufficiently excited and or the initial error is too large [39, 40]. Compared to the EKF approach, techniques such as [22, 24, 25, 27, 30, 31], show asymptotic convergence of the structure estimation errors. Furthermore, results such as [1, 18, 23, 26, 29] 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, the authors in [26] show exponential convergence assuming the PE condition is met and either the initial estimation error is small or the velocities are limited. Furthermore, the development in [29] yields exponential convergence assuming the observer satisfies the EOJ condition. In [1] 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 [29]

the observer becomes singular. Typically, SfM approaches require the motion (e.g., linear and angular velocities) to be known; however, the design in [31], extending an approach similar to [1], demonstrates a partial solution to the more challenging problem (i.e. compared to SfM) of structure and motion (SaM) 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 are measurable along one axis.

In [32], 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 such as [1, 18, 23, 26, 29] that assume a PE condition, [32] only requires finite excitation. The finite excitation condition results from the use of concurrent learning (CL) (cf., [41–44]). The concept of CL 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., [32, 45–49]). ICL removes the necessity to estimate the highest order derivative of the system required in traditional concurrent learning (cf., [32, 45, 48, 49]).

Although ICL removes the need for measuring the state derivative, it still requires the state to be measurable; yet, a unique challenge in [32] is that the state depends on the unmeasurable distance to the target. Moreover, the traditional state used in results such as [1, 17–27, 29–31] 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 camera's FOV; however, the constraint can be violated for some camera rotations that cause the feature to leave the FOV.

In many visual servoing approaches, the trajectory of the camera is constrained to maintain continuous observation of features on a landmark (cf., [50–61]); however, constraining the trajectory to keep constant observation of a landmark is not possible when traveling over large distances. Recent results develop approaches for relaxing the continuous feedback requirement (cf., [36–38, 62, 63]), enabling the ability to temporarily allow a landmark or target to leave the camera's FOV temporarily given dwell-time conditions are satisfied on the amount of time feedback is unavailable and available (e.g., features are in and out of the FOV). However, [36–38] still require the positive depth constraint while the object is not in the camera's FOV. While these types of approaches may solve issues that arise from occlusions over small periods or for reconstructing smaller environments, it still may not be possible or desirable to return to some regions in large environments. Others approach the problem of temporary loss of feedback by assuming the range to an object is available (cf., [64–66]), which is often not possible, or do not determine the pose of a target (cf., [67]). Other approaches use function approximation methods to learn a motion model for a target (cf., [68–73]); however, these approaches assume models are Gaussian processes and show simulation and experimental results but do not provide a stability analysis.

An approach to allowing features to leave the FOV for an extended period of time, enabling the ability to travel over large distances, is the use of multiple sets of tracked features (cf., [2–5, 7, 9–14, 74–80]); however, most of these approaches assume that each new set of features can be approximated before older sets of features leave the camera's FOV. 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. This drawback motivates the development of techniques that initially learn each set

of features independently. 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. [74–80]). In general, requiring homographic relationships may introduce more error in position estimates or fail to estimate position because features do not always lie on a plane, especially in environments comprised of multiple objects. Additionally, most of these approaches do not take an observer based approach (cf., [2–5, 7, 9–14, 74–80]); instead, known geometry or measurements are used (cf. [74–80]), or an optimization method is used which may not converge without a good initial guess or only converge up to a scale factor (cf., [2–5, 7, 9–14]) and can be computationally expensive. Methods that rely on measurements or geometry reduce robustness to noise because no control of the rate of learning is possible and there is no ability to remove potentially noisy data; however, many of the optimization methods have outlier rejection and consider noise. Stochastic homography based approaches are developed in [79] and [80] 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 or learning must be removed (i.e., it is not possible to instantly reconstruct features if they are nonplanar and so the structure of each set of features must be learned). While all of these approaches enable traveling over large distances, they do not all combine the pose and environment estimation with tracking moving objects (i.e., the structure and motion from motion (SaMfM) problem [81]), which is required for target tracking.

The primary difficulty that arises in estimating the trajectory of a moving object using a moving camera is that the velocity of the moving object is not typically measurable. Some approaches assume a known length on the object and estimate the pose directly (cf., [82–89]). One of the first results to estimate the path of moving objects

was discussed [90] where a batch method of trajectory triangulation was developed by restricting the object motion to be along lines and conics. The result in [90] was generalized in [91] where the object trajectory could be along any curve. The work in [92] and [93] discuss Euclidean path estimation of objects moving along lines. However, even though these approaches discuss the number of images required to calculate trajectories of moving objects and are shown to work in select environments, stability is not guaranteed. Additionally, these batch methods are restrictive in the sense they require the velocities of the camera and objects to be small and are difficult to implement online because repeated optimization over image sequences is required.

The first SaMfM result that included a stability analysis to show asymptotic estimation is in [81] where a robust integral sign of the error (RISE) observer is developed under the assumptions of constant velocities for moving features, known bounds on distances, measurable optical flows, differentiable camera accelerations, and persistence of excitation (PE). The result in [94] asymptotically estimated an object's velocity where some length on the object is assumed known and the objects initial orientation are used to develop a RISE estimator for a single stationary camera. Results in [95] and [96] use a passivity-based approach to show the estimation error asymptotically converges under the assumption of known lengths on the moving object. The assumption that the moving features have a constant velocity in [81] is relaxed in [97] and [98] by restricting the motion of the object to a plane and requiring PE, using an extended unknown input observer to show the observer error converges exponentially to zero. In [99], a recursive least squares observer is developed that assumes a known bound on the distance to moving and stationary features, known camera angular velocities, and PE to develop a RISE-based method that asymptotically estimates the distance to moving features and the linear velocities of the features and the camera. The result in [99] is extended in [100] to remove the PE condition by assuming the velocity of the moving features and the angular velocity of the camera are known to develop a RISE-based approach

to asymptotically estimate the camera linear velocities and distances to the features. In [45], an approach similar to [32] is developed that uses velocity measurements of the camera and tracked stationary features to learn the Euclidean positions of the stationary features and the trajectory of the camera. The velocity of the camera and the estimates of the positions of the stationary features are then used to show the estimation error of the Euclidean path and velocity of a moving object, as well as the structure of the moving object, is globally exponentially convergent to an ultimate bound assuming the object velocity and acceleration are bounded. Additionally, if the velocity of the moving object is constant, the ultimate bound of the estimation error will be zero. The developed observer for the moving object features relaxed the requirement to measure optical flow and the PE assumption. In [101], estimators are developed for a network of stationary cameras to exponentially estimate the Euclidean distance to a moving object's features and the object's velocity, with respect to each camera. The objective is similar to the inverse of the daisy-chaining type problem (cf., [74–80]). However, [45] and most other SaMfM approaches assume continuous feedback of the moving object. Unlike the camera pose estimation problem, the moving object will typically only leave the camera's FOV temporarily implying approaches that consider temporary loss of feedback (cf., [36–38, 62, 63]) may be developed to determine how long a target may leave the camera's FOV.

1.2 Outline of the Dissertation

Chapter 2 describes the dynamics for a moving monocular camera tracking stationary features and a moving target's features. The dynamics present a unique approach to the SfM and SaMfM where relationships are developed showing how the Euclidean distance to stationary features relates to the pose of the agent and the Euclidean distance to moving features relates to the pose and velocity of the moving target.

A global exponentially stable observer for feature scale is developed in Chapter 3, under a finite excitation condition through the use of ICL. Since the observer only

requires finite excitation to be globally exponentially stable, the observer is more general than previous results. The result indicates that the Euclidean distance to a set of features on a stationary object and the path the camera travels while viewing that object are estimated exponentially fast implying the structure (i.e., Euclidean coordinates of the tracked features) and path are reconstructed exponentially. Furthermore, the developed estimation method does not require the features on the objects to be planar and does not require the positive depth constraint. An experimental study is presented which compares the developed Euclidean distance observer to previous observers demonstrating the effectiveness of this result.

In Chapter 4, an extension to the learning approaches in Chapter 3 is developed that applies a new learning strategy that maintains a continuous estimate of the position of the camera and estimates the structure of features as they come into the camera's field-of-view. Furthermore, the developed learning strategy allows simulated measurements of features from objects that are no longer in the field-of-view enabling a continuous estimate of the distance to features with respect to the camera. Additionally, this approach shows how the extended observer removes the positive depth constraint required by all previous structure from motion approaches. Using this approach, a camera may travel over large distances without keeping specific features in the field-of-view for all time and allow objects to permanently leave the field-of-view if necessary. A Lyapunov based stability analysis proves that the observers for estimating the path of the camera as well as the structure of each set of objects are globally exponentially stable while features are in the camera's field-of-view. A switched systems analysis is used to develop dwell-time conditions to indicate how long a feature must be tracked to ensure the distance estimation error is below a threshold. After the distance estimates have converged below the threshold, the feature may be used to update the position of the camera. If a feature does not satisfy the dwell-time condition, it is never used to update the position of the agent. Furthermore, the approach does not require a new set of

features to be in the camera's FOV when older features leave the camera's FOV. Finally, if a recognized landmark enters the camera's FOV, the feedback is used to compensate for drift error.

In Chapter 5, the approach in Chapter 4 is used to provide pose estimates of the camera and an extension of Chapter 3 is developed to exponentially estimate the pose, velocity, and acceleration of the moving target. Specifically, using the pose and velocity of the camera, the estimation error of the Euclidean trajectory of the target as well as the structure of the target, is globally exponentially convergent to an ultimate bound assuming the target velocity and acceleration are bounded and dwell-time conditions are satisfied. The developed estimator relaxes the requirement to have continuous observation of the target, to know structure, velocity, or acceleration of the target, and does not require the persistence of excitation assumption or positive depth constraint.

Chapter 6 concludes the dissertation with a discussion of the developed estimation algorithms and potential extensions.

CHAPTER 2 SYSTEM MODEL

The development in this chapter presents the dynamics and assumptions used in Chapters 3-5. Section 2.1 develops the dynamics of a single stationary object's features relative to a moving camera, Section 2.2 extends Section 2.1 to a sequence of objects, and Section 2.4 develops the dynamics of a moving object's features relative to a moving camera.

2.1 Motion Model Using Stationary Features

This section describes the dynamic relationship between a single object's features and a moving camera. To facilitate the subsequent development, a key frame is defined as the camera frame when features are first extracted from an image of an object. Furthermore, a key frame, denoted by \mathcal{F}_k , has its origin at the principal point of that image, denoted by k , and basis $\{\underline{x}_k, \underline{y}_k, \underline{z}_k\}$. The frame when 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).. This implies that \mathcal{F}_k is established to coincide with \mathcal{F}_c at time $t = 0$, where $t \in \mathbb{R}_{\geq 0}$ represent time.

Assumption 2.1. A stationary object, s , has at least $p \in \mathbb{Z}_{\geq 4}$ features that can be detected and tracked provided they are within the FOV of the camera.

Remark 2.1. Features on an object in the camera's FOV can be tracked using descriptor and matching techniques such as [102–108] or feature extraction and tracking techniques such as [109–112].

Assumption 2.2. The camera intrinsic matrix $A \in \mathbb{R}^{3 \times 3}$ is known and invertable [16].

Assumption 2.3. 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 $\|\underline{v}_c(t)\| \leq \overline{v}_c$ and $\|\underline{\omega}_c(t)\| \leq \overline{\omega}_c$, where $\overline{v}_c, \overline{\omega}_c \in \mathbb{R}_{>0}$ are known constants.

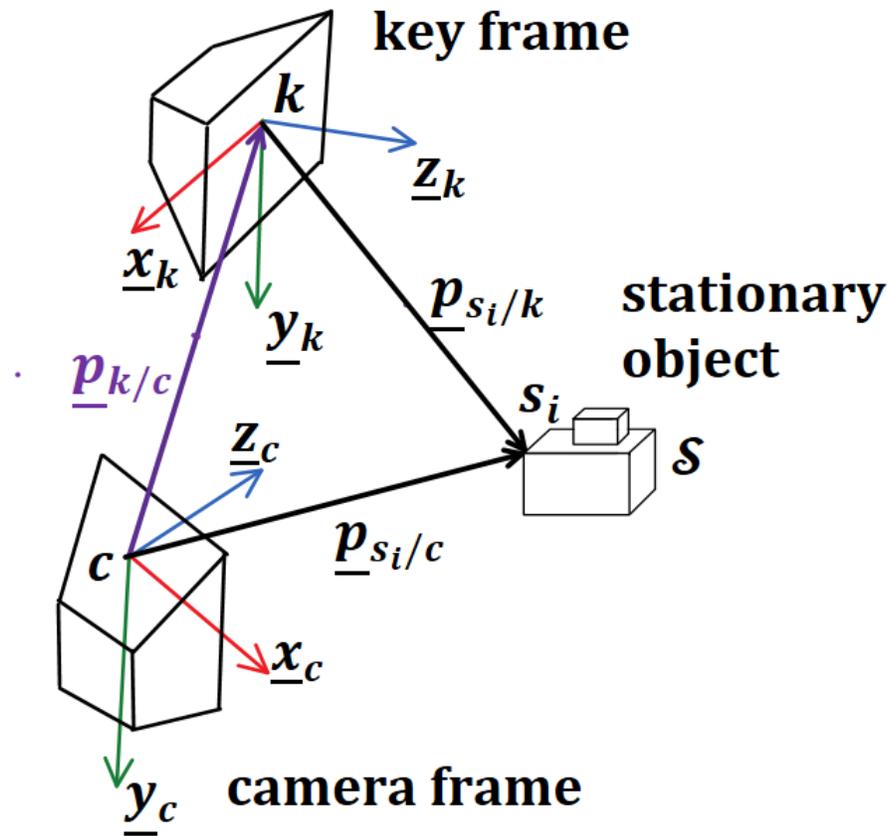


Figure 2-1. Example geometry for tracking the position of the i th feature of s where the camera starts at the top left where the key image is taken and is traveling downward from the upper left to the lower left while tracking a stationary object on the right.

As shown in Figure 2-1, the position of the i th feature on s , $s_i \in \mathbb{Z}_{>0}$, $\forall i = \{1, \dots, p\}$, can be described as

$$\underline{p}_{s_i/c}(t) = \underline{p}_{k/c}(t) + R_{k/c}(t) \underline{p}_{s_i/k}, \quad (2-1)$$

where $\underline{p}_{k/c}(t) \in \mathbb{R}^3$ is the position of the k with respect to c expressed in \mathcal{F}_c , $\underline{p}_{s_i/k}(t) \in \mathbb{R}^3$ is the position of feature s_i with respect to k expressed in \mathcal{F}_k , $R_{k/c}(t) \in \mathbb{R}^{3 \times 3}$ is the rotation matrix describing the orientation of \mathcal{F}_k with respect to \mathcal{F}_c , and $\underline{p}_{s_i/c}(t) \in \mathbb{R}^3$ is the position of feature s_i with respect to c expressed in \mathcal{F}_c . Rearranging (2-1) gives

$$Y_{s_i}(t) \begin{bmatrix} d_{s_i/c}(t) \\ d_{k/c}(t) \end{bmatrix} = R_{k/c}(t) \underline{u}_{s_i/k} d_{s_i/k}, \quad (2-2)$$

where $Y_{s_i}(t) \triangleq \begin{bmatrix} \underline{u}_{s_i/c}(t) & -\underline{u}_{k/c}(t) \end{bmatrix}$, $d_{s_i/c}(t) \in \mathbb{R}_{>0}$ and $\underline{u}_{s_i/c}(t) \in \mathbb{R}^3$ are the distance and unit vector of feature s_i with respect to c expressed in \mathcal{F}_c , $d_{k/c}(t) \in \mathbb{R}_{>0}$ and $\underline{u}_{k/c}(t) \in \mathbb{R}^3$ are the distance and unit vector of k with respect to c expressed in \mathcal{F}_c , and $d_{s_i/k} \in \mathbb{R}_{>0}$ and $\underline{u}_{s_i/k} \in \mathbb{R}^3$ are the distance and unit vector of feature s_i with respect to k expressed in \mathcal{F}_k .

While a set of features, $\{s_i\}_{i=1}^p$, are in the camera's FOV and $d_{k/c}(t) > 0$, the rotation $R_{k/c}(t)$ and unit vector $\underline{u}_{k/c}(t)$ can be determined from a general set of stationary features, using existing techniques such as planar homography decomposition or essential decomposition.¹ Additionally, $\underline{u}_{s_i/k}$ and $\underline{u}_{s_i/c}(t)$ can always be determined from $\underline{u}_{s_i/k} = \frac{A^{-1}p_{s_i/k}}{\|A^{-1}p_{s_i/k}\|}$ and $\underline{u}_{s_i/c}(t) = \frac{A^{-1}p_{s_i/c}(t)}{\|A^{-1}p_{s_i/c}(t)\|}$ where $p_{s_i/k}$, $p_{s_i/c}(t) \in \mathbb{R}^3$ are the homogeneous pixel coordinates of feature s_i in \mathcal{F}_k and \mathcal{F}_c , respectively. When the motion of the camera is not parallel to the direction to a feature, $\left(1 - \|\underline{u}_{k/c}^T(t) \underline{u}_{s_i/c}(t)\|\right) > \lambda_a$, where

¹ See [15], [16], and [113] for examples on calculating the rotation and normalized translation from planar and nonplanar features.

$\lambda_a \in (0, 1)$ is a selected constant, and $d_{k/c}(t) > 0$, (2-2) can be written as

$$\begin{bmatrix} d_{s_i/c}(t) \\ d_{k/c}(t) \end{bmatrix} = \psi_{s_i}(t) d_{s_i/k}, \quad (2-3)$$

where $\psi_{s_i}(t) \triangleq (Y_{s_i}^T(t) Y_{s_i}(t))^{-1} Y_{s_i}^T(t) R_{k/c}(t) \underline{u}_{s_i/k}$ is invertable and measurable while $(1 - \|\underline{u}_{k/c}^T(t) \underline{u}_{s_i/c}(t)\|) > \lambda_a$. Furthermore, given \mathcal{F}_k and s are stationary, the time derivatives of the unknown distances are measurable as

$$\frac{d}{dt}(d_{s_i/c}(t)) = -\underline{u}_{s_i/c}^T(t) \underline{v}_c(t), \quad (2-4)$$

$$\frac{d}{dt}(d_{k/c}(t)) = -\underline{u}_{k/c}^T(t) \underline{v}_c(t), \quad (2-5)$$

and

$$\frac{d}{dt}(d_{s_i/k}) = 0. \quad (2-6)$$

2.2 Extension of Motion Model Using Stationary Features

Considering the objective is to travel over large distances, this section extends Section 2.1 to a sequence of stationary objects.

Let $\{\mathcal{F}_{k_j}\}_{j=1}^{\bar{p}_s(t)}$ and $\{\zeta_j\}_{j=1}^{\bar{p}_s(t)}$ 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 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 $t = \zeta_j^a$. Let $\mathcal{V}_c \subset \mathbb{R}^3$ represent the Euclidean space of the camera's FOV expressed in \mathcal{F}_c .

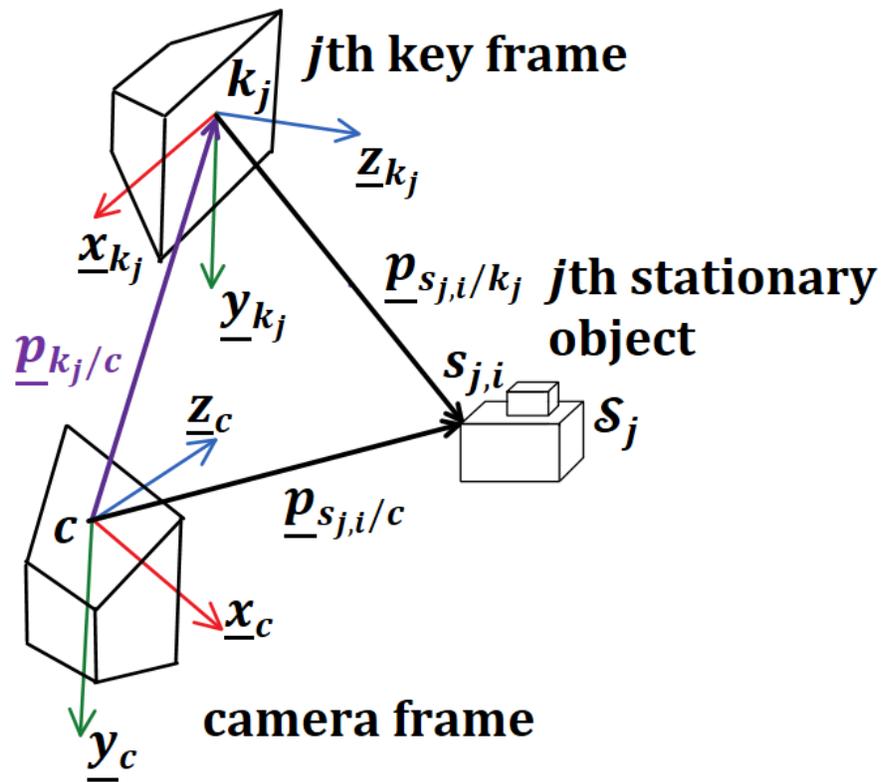


Figure 2-2. Example geometry for tracking the position of the i th feature in S_j where the key image taken from the top left and the camera traveling downward from the upper left to the lower left while tracking S_j on the right.

Assumption 2.4. 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 the known constant $\bar{d} \in \mathbb{R}_{>0}$.

Remark 2.2. Assumption 2.4 is necessary for the subsequent development of the maximum and minimum dwell-time conditions and is inherently required when estimating distances to an unknown object's features that may leave the camera's FOV.

Assumption 2.5. 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 $\mathcal{O}_{s_j} \triangleq \{s_{j,i}\}_{i=1}^{\bar{p}_{s_j}}$ 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.

As illustrated in Figure 2-2, the position feature $s_{j,i} \in \mathcal{O}_{s_j}$, can be described as

$$\underline{p}_{s_{j,i}/c}(t) = \underline{p}_{k_j/c}(t) + R_{k_j/c}(t) \underline{p}_{s_{j,i}/k_j}, \quad (2-7)$$

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 , $\underline{p}_{s_{j,i}/k_j}(t) \in \mathbb{R}^3$ 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 (2-7) gives

$$Y_{s_{j,i}}(t) \begin{bmatrix} d_{s_{j,i}/c}(t) \\ d_{k_j/c}(t) \end{bmatrix} = R_{k_j/c}(t) \underline{u}_{s_{j,i}/k_j} d_{s_{j,i}/k_j}, \quad (2-8)$$

where $Y_{s_{j,i}}(t) \triangleq \begin{bmatrix} \underline{u}_{s_{j,i}/c}(t) & -\underline{u}_{k_j/c}(t) \end{bmatrix}$, $d_{s_{j,i}/c}(t) \in \mathbb{R}_{>0}$ and $\underline{u}_{s_{j,i}/c}(t) \in \mathbb{R}^3$ are the distance and unit vector of feature $s_{j,i}$ with respect to c expressed in \mathcal{F}_c , $d_{k_j/c}(t) \in \mathbb{R}_{>0}$ and $\underline{u}_{k_j/c}(t) \in \mathbb{R}^3$ are the 2-8 distance and unit vector of k_j with respect to c expressed in \mathcal{F}_c , and $d_{s_{j,i}/k_j} \in \mathbb{R}_{>0}$ and $\underline{u}_{s_{j,i}/k_j} \in \mathbb{R}^3$ 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's 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, $\mathcal{P}_{s_j}(t) \triangleq \{s_{j,i} \in \mathcal{O}_{s_j} : \underline{p}_{s_{j,i}/c}(t) \in \mathcal{S}_j \cap \mathcal{V}_c\}$, 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 $\mathcal{P}_{s_j}(t)$. 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 $\zeta_j^u \triangleq \max \left\{ \zeta_{s_{j,i}}^u \right\}_{i=1}^{p_{s_j}(t)}$ (i.e., the time the last feature leaves).

While the origins k_j and c are not coincident (i.e., $d_{k_j/c}(t) > \underline{d}_1$ where $\underline{d}_1 \in \mathbb{R}_{>0}$ is a constant), Assumptions 2.2 and 2.5 ensure the rotation $R_{k_j/c}(t)$ 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 $p_{s_j}(t) \geq 4$. 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.

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 $\left(1 - \|\underline{u}_{k_j/c}^T(t) \underline{u}_{s_{j,i}/c}(t)\|\right) < \lambda_a$ or $\left(1 - \|\underline{u}_{k_j/c}^T(t) \underline{u}_{s_{j,i}/c}(t)\|\right) > \lambda_a$, respectively. While the origins k_j and c are not coincident (i.e., $d_{k_j/c}(t) > \underline{d}_1$) and the motion of the camera is not parallel to the direction to a feature (i.e., $\left(1 - \|\underline{u}_{k_j/c}^T(t) \underline{u}_{s_{j,i}/c}(t)\|\right) > \lambda_a$), $\sigma_{s_{j,i}}(t) = a$ (i.e., not parallel motion), (2–8) is invertable in the sense that $Y_{s_{j,i}}(t)$ is full column rank and (2–8) may be written

$$\begin{bmatrix} d_{s_{j,i}/c}(t) \\ d_{k_j/c}(t) \end{bmatrix} = \psi_{s_{j,i}}^a(t) d_{s_{j,i}/k_j}, \quad (2-9)$$

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 2.2 and 2.5. When $\sigma_{s_{j,i}}(t) = u$ (i.e., parallel motion), (2–8) is not invertable. However, (2–8) can always be written $d_{s_{j,i}/c}(t) =$

$\underline{u}_{s_j,i/c}^T(t) \underline{u}_{k_j/c}(t) d_{k_j/c}(t) + \underline{u}_{s_j,i/c}^T(t) R_{k_j/c}(t) \underline{u}_{s_j,i/k_j} d_{s_j,i/k_j}$ yielding

$$d_{s_j,i/c}(t) = \psi_{s_j,i}^u(t) \begin{bmatrix} d_{k_j/c}(t) \\ d_{s_j,i/k_j} \end{bmatrix}, \quad (2-10)$$

where

$$\psi_{s_j,i}^u(t) \triangleq \begin{bmatrix} \underline{u}_{s_j,i/c}^T(t) \underline{u}_{k_j/c}(t) & \underline{u}_{s_j,i/c}^T(t) R_{k_j/c}(t) \underline{u}_{s_j,i/k_j} \end{bmatrix}.$$

Using Assumption 2.3 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 $d_{s_j,i/k_j}$ are measurable for $s_{j,i} \in \mathcal{P}_{s_j}(t)$ as

$$\frac{d}{dt}(d_{s_j,i/c}(t)) = -\underline{u}_{s_j,i/c}^T(t) \underline{v}_c(t), \quad (2-11)$$

$$\frac{d}{dt}(d_{k_j/c}(t)) = -\underline{u}_{k_j/c}^T(t) \underline{v}_c(t), \quad (2-12)$$

and

$$\frac{d}{dt}(d_{s_j,i/k_j}) = 0. \quad (2-13)$$

Taking the time derivative of $\underline{u}_{s_j,i/c}(t)$ yields

$$\frac{d}{dt}(\underline{u}_{s_j,i/c}(t)) = -\underline{\omega}_c^\times(t) \underline{u}_{s_j,i/c}(t) + \frac{1}{d_{s_j,i/c}(t)} (\underline{u}_{s_j,i/c}(t) \underline{u}_{s_j,i/c}^T(t) - I_{3 \times 3}) \underline{v}_c(t), \quad (2-14)$$

where $\underline{\omega}_c^\times(t) \triangleq \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}$, and $I_{3 \times 3} \triangleq \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

2.3 Motion Model of Camera

As previously discussed, it is necessary to have a continuous estimate of the position of the camera over time regardless of the visibility of objects (i.e., there will time periods where no objects remain in the camera's FOV). As shown in Figure 2-3, the position of the camera may be expressed through the sequence of objects. Since the starting location of the camera may be unknown, the position of the camera over time

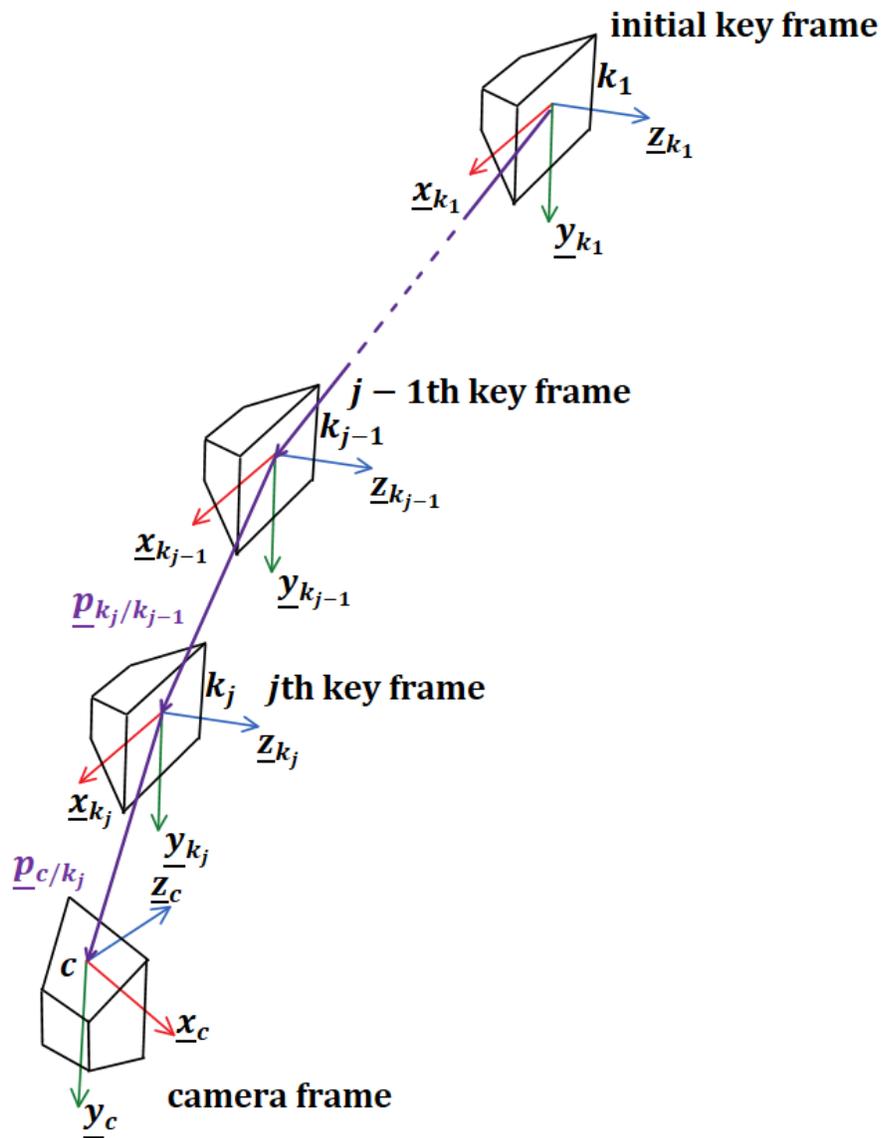


Figure 2-3. Example geometry for pose of the camera over time where the camera starts at the top where the first key frame is located and is traveling downward to the lower left.

can be expressed relative to the first key frame as

$$\underline{p}_{c/k_1}(t) = R_{k_{\overline{p_s}(t)}/k_1} \underline{p}_{c/k_{\overline{p_s}(t)}}(t) + \sum_{j=2}^{\overline{p_s}(t)} R_{k_{j-1}/k_1} \underline{p}_{k_j/k_{j-1}}. \quad (2-15)$$

The derivative of the position with respect to time can be expressed similarly

$$\frac{d}{dt} \left(\underline{p}_{c/k_1}(t) \right) = R_{c/k_1} \underline{v}_c(t). \quad (2-16)$$

The orientation of $R_{c/k_1}(t)$ is derived using a unit quaternion, $q_{c/k_1}(t) \in \mathbb{H}$, which can be represented as $q_{c/k_1}(t) \in \mathbb{R}^4$, where $q_{c/k_1}^T(t) q_{c/k_1}(t) = 1$. The derivative with respect to time for $q_{c/k_1}(t)$ is

$$\frac{d}{dt} (q_{c/k_1}(t)) = \frac{1}{2} B(q_{c/k_1}(t)) \underline{\omega}_c(t), \quad (2-17)$$

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 $B^T(q(t)) B(q(t)) = I_{3 \times 3}$.² The rotation matrix representation of a unit quaternion $q(t)$ is

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

Remark 2.3. The orientation $q_{c/k_1}(t)$ is often measurable using local sensors (e.g., inertial measurement units and magnetometers). So measurements of $q_{c/k_1}(t)$ may be

² Time dependence is suppressed except when needed for clarity or introducing terms.

available at all times in many applications. In those applications, no estimate of $q_{c/k_1}(t)$ would be required.

Given $\underline{u}_{k_j/c}(t)$ is only measurable locally, an estimate of subsequently $\underline{p}_{k_j/c}(t)$ is developed. Taking the derivative of the position of the camera expressed in the camera frame with respect to time yields

$$\frac{d}{dt} \left(\underline{p}_{k_j/c}(t) \right) = -\underline{v}_c(t) - \underline{\omega}_c^\times(t) \underline{p}_{k_j/c}(t). \quad (2-18)$$

Similarly, the orientation used for each key frame is $R_{k_j/c}(t)$ as shown in (2-7). The unit quaternion form of the orientation is $\underline{q}_{k_j/c}(t) \in \mathbb{H}$, 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 derivative with respect to time for $q_{k_j/c}(t)$ is

$$\frac{d}{dt} (q_{k_j/c}(t)) = -\frac{1}{2} B(q_{k_j/c}(t)) \underline{\omega}_c(t). \quad (2-19)$$

2.4 Motion Model of Moving Features

Given the objective is to track a moving object that intermittently leaves the camera's FOV, this section develops the dynamic relationships between the a moving object and moving camera. Tracking features on a moving object that intermittently leaves the camera's FOV can be a significantly more challenging problem than tracking stationary features. Descriptor and matching techniques such as [102–108] may have better tracking performance for moving objects.

Assumption 2.6. A moving object represented by $\mathcal{M} \subset \mathbb{R}^3$, where \mathcal{M} represents the minimum Euclidean sphere enclosing the object expressed in \mathcal{F}_c , has $n \in \mathbb{Z}_{\geq 4}$ features that can be detected and tracked while $\mathcal{M} \subset \mathcal{V}_c$ (i.e., the object is in the camera's FOV).

While \mathcal{M} is in the FOV of the camera, the position of each feature in \mathcal{M} can be related to the camera. Let $\mathcal{O}_m \triangleq \{m_i\}_{i=1}^n$ represent the set of features on the object, $\underline{p}_{m_i/c}(t) \in \mathbb{R}^3$ represent the position of feature m_i with respect to c expressed in \mathcal{F}_c , and let $\sigma_{\mathcal{O}}(t) \in \{u, a\}$ be a switching signal indicating whether $\mathcal{M} \subset \mathcal{V}_c$ implying when

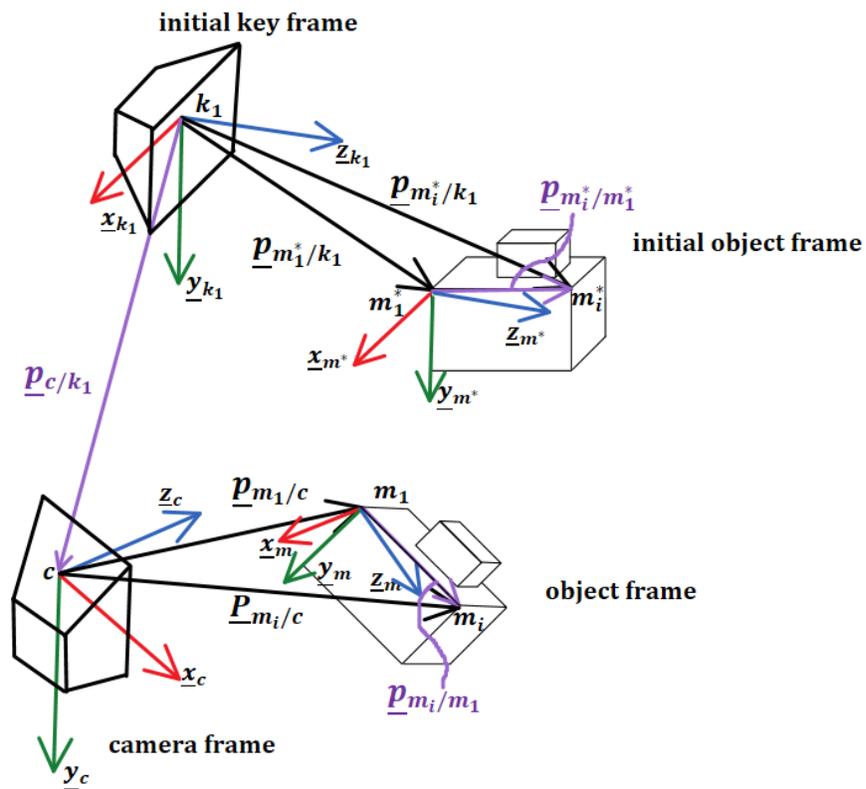


Figure 2-4. Example geometry for tracking the position of the i th feature of m and m^* .

$\sigma_{\mathcal{O}}(t) = a$, $\{p_{m_i/c}(t)\}_{i=1}^n \subset \mathcal{M} \cap \mathcal{V}_c$. Furthermore, let $\pi_j^a \in \mathbb{R}_{\geq 0}$ be the j th time t when $\sigma_{\mathcal{O}}(t) = a$ occurs, and let $\pi_j^u \in \mathbb{R}_{\geq \pi_j^a}$ represent the j th time when $\sigma_{\mathcal{O}}(t) = u$.

Assumption 2.7. The features on the moving object are initially contained in the camera's FOV (i.e., $\{p_{m_i/c}(\pi_1^a)\}_{i=1}^n \subset \mathcal{M} \cap \mathcal{V}_c$) and the moving object and first stationary object are detected at the same time (i.e., $\pi_1^a = \zeta_1^a$).

As illustrated in Figure 2-4, the object frame, denoted by \mathcal{F}_m , has its origin at feature m_1 , with basis $\{\underline{x}_m, \underline{y}_m, \underline{z}_m\}$. The initial object frame (i.e., the object at the first key frame), denoted by \mathcal{F}_{m^*} , has its origin at m_1^* and basis $\{\underline{x}_{m^*}, \underline{y}_{m^*}, \underline{z}_{m^*}\}$, where m^* represents m at time $t = \pi_1^a$.

Remark 2.4. The initial orientation of the object may be selected arbitrarily given any coordinate frame may be attached to the moving object body. To aid in the subsequent development, the object initial basis is selected to align with the first key frame (i.e.,

$$\{\underline{x}_{m^*}, \underline{y}_{m^*}, \underline{z}_{m^*}\} = \{\underline{x}_{k_1}, \underline{y}_{k_1}, \underline{z}_{k_1}\} \text{ and } R_{m^*/c}(\pi_1^a) = R_{m^*/k_1} = I_{3 \times 3}.$$

As also shown in Figure 2-4, the position of the i th feature on \mathcal{M} may be described in \mathcal{F}_c as

$$\underline{p}_{m_i/c}(t) = \underline{p}_{m_1/c}(t) + R_{m/c}(t) \underline{p}_{m_i/m_1}, \quad (2-20)$$

where $R_{m/c}(t) \in \mathbb{R}^{3 \times 3}$ is the rotation matrix describing the orientation of \mathcal{F}_m with respect to \mathcal{F}_c , and $\underline{p}_{m_i/m_1} \in \mathbb{R}^3$ is the constant position of feature m_i with respect to m_1 expressed in \mathcal{F}_m . The same expression is true when features in \mathcal{M} are first extracted (i.e., when $\sigma_{\mathcal{O}}(t) = a$ at time $t = \pi_1^a$); specifically, $\underline{p}_{m_i^*/c}(\pi_1^a) \triangleq \underline{p}_{m_i/c}(\pi_1^a) = \underline{p}_{m_i^*/k_1}$ and as described in Remark 2.4, $R_{m^*/c}(\pi_1^a) = R_{m^*/k_1} = I_{3 \times 3}$ implying

$$\underline{p}_{m_i/m_1} = \underline{p}_{m_i^*/k_1} - \underline{p}_{m_1^*/k_1}. \quad (2-21)$$

Substituting (2-21) into (2-20), the position of the i th feature in \mathcal{V}_c can be described in \mathcal{F}_c by

$$\underline{p}_{m_i/c}(t) = \underline{p}_{m/m^*}(t) + R_{m/c}(t) \underline{p}_{m_i^*/k_1}, \quad (2-22)$$

where $\underline{p}_{m/m^*}(t) \triangleq \underline{p}_{m_1/c}(t) - R_{m/c}(t) \underline{p}_{m_1^*/k_1}$. Rearranging (2–22) gives

$$Y_{m_i}(t) \begin{bmatrix} d_{m_i/c}(t) \\ d_{m/m^*}(t) \end{bmatrix} = R_{m/c}(t) \underline{u}_{m_i^*/k_1} d_{m_i^*/k_1}, \quad (2–23)$$

where $Y_{m_i}(t) \triangleq \begin{bmatrix} \underline{u}_{m_i/c}(t) & -\underline{u}_{m/m^*}(t) \end{bmatrix}$, $d_{m_i/c}(t) \in \mathbb{R}_{>0}$ and $\underline{u}_{m_i/c}(t) \in \mathbb{R}^3$ are the distance and unit vector of feature m_i with respect to c expressed in \mathcal{F}_c , $d_{m/m^*}(t) \in \mathbb{R}_{>0}$ and $\underline{u}_{m/m^*}(t) \in \mathbb{R}^3$ are the distance and unit vector of $\underline{p}_{m/m^*}(t)$ from (2–22), and $d_{m_i^*/k_1} \in \mathbb{R}_{>0}$ and $\underline{u}_{m_i^*/k_1} \in \mathbb{R}^3$ are the distance and unit vector of feature m_i^* with respect to k_1 expressed in \mathcal{F}_{k_1} .

Under Assumptions 2.2 and 2.6, while $\sigma_{\mathcal{O}}(t) = a$, the rotation matrix $R_{m/c}(t)$ and unit vector $\underline{u}_{m/m^*}(t)$ can be determined from a general set of features, using existing techniques. Additionally, $\underline{u}_{m_i^*/k_1}$ and $\underline{u}_{m_i/c}(t)$ can be determined from $\underline{u}_{m_i^*/k_1} = \frac{A^{-1} p_{m_i^*/k_1}}{\|A^{-1} p_{m_i^*/k_1}\|}$ and $\underline{u}_{m_i/c}(t) = \frac{A^{-1} p_{m_i/c}(t)}{\|A^{-1} p_{m_i/c}(t)\|}$ where $p_{m_i^*/k_1}, p_{m_i/c}(t) \in \mathbb{R}^3$ are the homogeneous pixel coordinates of feature m_i^* and m_i in \mathcal{F}_{k_1} and \mathcal{F}_c , respectively.

Let $\{\sigma_{m_i}(t)\}_{m_i \in \mathcal{O}_m}$ be the set of switching signals for the features in \mathcal{O}_m , where $\sigma_{m_i}(t) \in \{u, a\}$ indicates whether $(1 - \|\underline{u}_{m/m^*}^T(t) \underline{u}_{m_i/c}(t)\|) \leq \lambda_a$ or $(1 - \|\underline{u}_{m/m^*}^T(t) \underline{u}_{m_i/c}(t)\|) > \lambda_a$, respectively. Furthermore, let $\pi_{j,m_i}^{a_l} \in [\pi_j^a, \pi_j^u)$ represent the l th instance the i th moving feature satisfies the eigenvalue condition (i.e., $\sigma_{m_i}(t) = a$) during the j th instance the object enters the camera's FOV (i.e., $t \in [\pi_j^a, \pi_j^u)$) and let $\pi_{j,m_i}^{u_l} \in [\pi_j^a, \pi_j^u)$ represent the l th instance it doesn't satisfy the eigenvalue condition (i.e., $\sigma_{m_i}(t) = u$).

Remark 2.5. The set of features excluding the origin (i.e., $\{\sigma_{m_i}(t)\}_{m_i \in \mathcal{O}_m \setminus \{m_1\}}$) are $\sigma_{m_i}(t) = u$ if the origin does not satisfy the eigenvalue condition (i.e., $\sigma_{m_1}(t) = u$). This relationship results because the features are all dependent on the origin in the subsequent development. The set of features (i.e., $\{\sigma_{m_i}(t)\}_{m_i \in \mathcal{O}_m}$) are set as $\sigma_{m_i}(t) = u$ if the object leaves the camera's FOV (i.e., $\sigma_{\mathcal{O}}(t) = u$).

While the relative direction of motion is not parallel to the direction of a feature (i.e., $(1 - \|\underline{u}_{m/m^*}^T(t) \underline{u}_{m_i/c}(t)\|) > \lambda_a$), $\sigma_{m_i}(t) = a$ and (2-23) may be rearranged as

$$\begin{bmatrix} d_{m_i/c}(t) \\ d_{m/m^*}(t) \end{bmatrix} = \psi_{m_i}(t) d_{m_i^*/k_1}, \quad (2-24)$$

where

$$\psi_{m_i}(t) \triangleq (Y_{m_i}^T(t) Y_{m_i}(t))^{-1} Y_{m_i}^T(t) R_{m/c}(t) \underline{u}_{m_i^*/k_1}$$

and $\psi_{m_i}(t)$ is measurable.

Based on the definitions of $\underline{p}_{m_i/c}(t)$ and $\underline{p}_{m/m^*}(t)$, their derivatives with respect to time are

$$\frac{d}{dt} (\underline{p}_{m_i/c}(t)) = \underline{v}_m(t) - \underline{v}_c(t) - \underline{\omega}_c^\times(t) \underline{p}_{m_i/c}(t) + \underline{\omega}_m^\times(t) R_{m/c}(t) \underline{p}_{m_i/m_1} \quad (2-25)$$

and

$$\frac{d}{dt} (\underline{p}_{m/m^*}(t)) = \underline{v}_m(t) - \underline{v}_c(t) - \underline{\omega}_m^\times(t) R_{m/c}(t) \underline{p}_{m_1^*/k_1} - \underline{\omega}_c^\times(t) \underline{p}_{m/m^*}(t). \quad (2-26)$$

Taking the time derivatives of the unknown distances and using (2-25) and (2-26) yields

$$\begin{aligned} \frac{d}{dt} (d_{m_i/c}(t)) &= \underline{u}_{m_i/c}^T(t) \underline{v}_m(t) - \underline{u}_{m_i/c}^T(t) \underline{v}_c(t) \\ &\quad + \underline{u}_{m_i/c}^T(t) \underline{\omega}_m^\times(t) R_{m/c}(t) \underline{u}_{m_i^*/k_1} d_{m_i^*/k_1} \\ &\quad - \underline{u}_{m_i/c}^T(t) \underline{\omega}_c^\times(t) R_{m/c}(t) \underline{u}_{m_1^*/k_1} d_{m_1^*/k_1}, \end{aligned} \quad (2-27)$$

$$\begin{aligned} \frac{d}{dt} (d_{m/m^*}(t)) &= \underline{u}_{m/m^*}^T(t) \underline{v}_m(t) - \underline{u}_{m/m^*}^T(t) \underline{v}_c(t) \\ &\quad - \underline{u}_{m/m^*}^T(t) \underline{\omega}_m^\times(t) R_{m/c}(t) \underline{u}_{m_1^*/k_1} d_{m_1^*/k_1}, \end{aligned} \quad (2-28)$$

and

$$\frac{d}{dt}(d_{m_i^*/c_j}) = 0. \quad (2-29)$$

Taking the time derivatives of the directions yields

$$\begin{aligned} \frac{d}{dt}(\underline{u}_{m_i/c}(t)) &= \frac{1}{d_{m_i/c}(t)} \Psi_{m_i}(t) (\underline{v}_m(t) - \underline{v}_c(t)) - \underline{\omega}_c^\times(t) \underline{u}_{m_i/c}(t) \\ &\quad + \frac{1}{d_{m_i/c}(t)} \Psi_{m_i}(t) \underline{\omega}_m^\times(t) R_{m/c}(t) \underline{u}_{m_i^*/k_1} d_{m_i^*/k_1} \\ &\quad - \frac{1}{d_{m_i/c}(t)} \Psi_{m_i}(t) \underline{\omega}_m^\times(t) R_{m/c}(t) \underline{u}_{m_1^*/k_1} d_{m_1^*/k_1} \end{aligned} \quad (2-30)$$

and

$$\begin{aligned} \frac{d}{dt}(\underline{u}_{m/m^*}(t)) &= \frac{1}{d_{m/m^*}(t)} \Psi_{m/m^*}(t) (\underline{v}_m(t) - \underline{v}_c(t)) - \underline{\omega}_c^\times(t) \underline{u}_{m/m^*}(t) \\ &\quad - \frac{1}{d_{m/m^*}(t)} \Psi_{m/m^*}(t) \underline{\omega}_m^\times(t) R_{m/c}(t) \underline{u}_{m_1^*/k_1} d_{m_1^*/k_1}, \end{aligned} \quad (2-31)$$

where $\Psi_{m_i}(t) = \left(I_{3 \times 3} - \underline{u}_{m_i/c}(t) \underline{u}_{m_i/c}^T(t) \right)$ and $\Psi_{m/m^*}(t) \triangleq \left(I_{3 \times 3} - \underline{u}_{m/m^*}(t) \underline{u}_{m/m^*}^T(t) \right)$.

CHAPTER 3

SIMULTANEOUS ESTIMATION OF EUCLIDEAN DISTANCES TO A STATIONARY OBJECT'S FEATURES AND THE EUCLIDEAN TRAJECTORY OF A MONOCULAR CAMERA

In this chapter, image geometry insights are exploited to express the error system with a more general distance measure that only becomes zero when the target and camera are coincident; thereby, avoiding the positive depth constraint. While this result also requires the features to remain in the FOV, eliminating the positive depth constraint eliminates a barrier for future development that would allow intermittent viewing of the features. Although, the new image-geometry based error system avoids the potential depth singularity, the resulting error system still contains the unmeasurable distance to the target. However, the development in Section 3.1 illustrates how the unmeasurable state can be related to an unknown constant to enable the use of ICL. Regardless of the system identification method used, there is a delay before sufficient excitation occurs to identify the parameters. Therefore, the preliminary result in [32] and the development in Section 3.1 exhibit an arbitrarily long delay before determining the feature Euclidean coordinates. In Section 3.2, we modify the developed learning strategy to include gradient terms that enable transient learning until sufficient data has been collected for the ICL terms.

To illustrate the performance of the developed observers, multiple experiments are presented, including a comparison of the observers in Section 3.1 and Section 3.2 with the results in [1] and an EKF. These results indicate that the EKF and result in [1] have improved transient performance over the result in Section 3.1, before the ICL-based estimates converge. The EKF and result in [1] have similar transient response as the observer in Section 3.2 before the ICL-based estimates converge. After the ICL-based estimates converge, the observers in Sections 3.1 and 3.2 converge to steady-state with improved performance over the EKF and observer in [1].

3.1 Integral Concurrent Learning Observer Update Laws for Euclidean Distances

Motivated by the developments in [32, 45], an ICL update law is implemented to estimate the constant unknown distances, $d_{s_i/k}$, by integrating (2-4) and (2-5) over a time window $\varsigma \in \mathbb{R}_{>0}$ yielding

$$\begin{bmatrix} d_{s_i/c}(t) \\ d_{k/c}(t) \end{bmatrix} - \begin{bmatrix} d_{s_i/c}(t-\varsigma) \\ d_{k/c}(t-\varsigma) \end{bmatrix} = \int_{t-\varsigma}^t \eta_{s_i}(\iota) d\iota, \quad t > \varsigma,$$

where ς may be constant in size or change over time. While $\int_{t-\varsigma}^t \eta_{s_i}(\iota) d\iota$ is a known quantity, $\begin{bmatrix} d_{s_i/c}(t) \\ d_{k/c}(t) \end{bmatrix}$ and $\begin{bmatrix} d_{s_i/c}(t-\varsigma) \\ d_{k/c}(t-\varsigma) \end{bmatrix}$ are unknown; however, the relationship in (2-3) may be utilized at the current time t and the previous time $t-\varsigma$ yielding

$$\mathcal{Y}_{s_i}(t) d_{s_i/k} = \mathcal{U}_{s_i}(t), \quad (3-1)$$

where

$$\mathcal{Y}_{s_i}(t) \triangleq \begin{cases} 0_{2 \times 1}, & t \leq \varsigma, \\ (\psi_{s_i}(t) - \psi_{s_i}(t-\varsigma)), & t > \varsigma, \end{cases}$$

and

$$\mathcal{U}_{s_i}(t) \triangleq \begin{cases} 0_{2 \times 1}, & t \leq \varsigma, \\ \int_{t-\varsigma}^t \eta_{s_i}(\iota) d\iota, & t > \varsigma. \end{cases}$$

The dynamics in (3-1) demonstrate that concurrent learning can estimate the constant distances, $d_{s_i/k}$, to the features on s . Specifically, multiplying both sides of (3-1) by $\mathcal{Y}_{s_i}^T(t)$ yields

$$\mathcal{Y}_{s_i}^T(t) \mathcal{Y}_{s_i}(t) d_{s_i/k} = \mathcal{Y}_{s_i}^T(t) \mathcal{U}_{s_i}(t). \quad (3-2)$$

In general, $\mathcal{Y}_{s_i}(t)$ will not have full column rank (e.g., when the camera is stationary) implying $\mathcal{Y}_{s_i}^T(t) \mathcal{Y}_{s_i}(t) \geq 0$. However, the equality in (3-2) may be evaluated at instances

in time and summed together (i.e., history stacks) as

$$\Sigma_{\mathcal{Y}_{s_i}} d_{s_i/k} = \Sigma_{\mathcal{U}_{s_i}}, \quad (3-3)$$

where $\Sigma_{\mathcal{Y}_{s_i}} \triangleq \sum_{h_i=1}^N \mathcal{Y}_{s_i}^T(t_{h_i}) \mathcal{Y}_{s_i}(t_{h_i})$, $\Sigma_{\mathcal{U}_{s_i}} \triangleq \sum_{h_i=1}^N \mathcal{Y}_{s_i}^T(t_{h_i}) \mathcal{U}_{s_i}(t_{h_i})$, $t_{h_i} \in (\varsigma, t)$, and $N \in \mathbb{Z}_{>1}$.

Assumption 3.1. Motion of the camera occurs so there exists finite constants $\tau_{s_i} \in \mathbb{R}_{>\varsigma}$, $\lambda_\tau \in \mathbb{R}_{>0}$ such that for all time $t \geq \tau_{s_i}$, $\lambda_{\min} \{ \Sigma_{\mathcal{Y}_{s_i}} \} > \lambda_\tau$, where $\lambda_{\min} \{ \cdot \}$ and $\lambda_{\max} \{ \cdot \}$ are the minimum and maximum eigenvalues of $\{ \cdot \}$.¹

Assumption 3.1 can be verified online and is heuristically easy to satisfy because it only requires a finite collection of sufficiently exciting $\mathcal{Y}_{s_i}(t)$ and $\mathcal{U}_{s_i}(t)$ to yield $\lambda_{\min} \{ \Sigma_{\mathcal{Y}_{s_i}} \} > \lambda_\tau$. The time τ_{s_i} is unknown; however, it can be determined online by checking the minimum eigenvalue of $\Sigma_{\mathcal{Y}_{s_i}}$. After τ_{s_i} , $\lambda_{\min} \{ \Sigma_{\mathcal{Y}_{s_i}} \} > \lambda_\tau$ implies that a constant unknown distance, $d_{s_i/k}$, can be determined from (3-3) as

$$d_{s_i/k} = \mathcal{X}_{s_i}, \quad t \geq \tau_{s_i}, \quad (3-4)$$

where

$$\mathcal{X}_{s_i} \triangleq \begin{cases} 0, & t < \tau_{s_i}, \\ \Sigma_{\mathcal{Y}_{s_i}}^{-1} \Sigma_{\mathcal{U}_{s_i}}, & t \geq \tau_{s_i}. \end{cases}$$

When $t \geq \tau_{s_i}$, (3-4) can be substituted into (2-3) to yield

$$d_{s_i/c}(t) = \nu_{s_i,1}(t), \quad t \geq \tau_{s_i}, \quad (3-5)$$

and

$$d_{k/c}(t) = \nu_{s_i,2}(t), \quad t \geq \tau_{s_i}, \quad (3-6)$$

where $\nu_{s_i,1}(t)$, $\nu_{s_i,2}(t)$ are the first and second elements of $\nu_{s_i}(t) \triangleq \psi_{s_i}(t) \mathcal{X}_{s_i}$.

¹ See [114] or [115] for some examples of methods for selecting data to satisfy the assumption.

The estimation errors, $\tilde{d}_{s_i/c}(t)$, $\tilde{d}_{k/c}(t)$, $\tilde{d}_{s_i/k}(t) \in \mathbb{R}$, are defined as

$$\tilde{d}_{s_i/c}(t) \triangleq d_{s_i/c}(t) - \hat{d}_{s_i/c}(t), \quad (3-7)$$

$$\tilde{d}_{k/c}(t) \triangleq d_{k/c}(t) - \hat{d}_{k/c}(t), \quad (3-8)$$

and

$$\tilde{d}_{s_i/k}(t) \triangleq d_{s_i/k} - \hat{d}_{s_i/k}(t), \quad (3-9)$$

where $\hat{d}_{s_i/c}(t)$, $\hat{d}_{k/c}(t)$, $\hat{d}_{s_i/k}(t) \in \mathbb{R}$ are the estimates. Motivated by the subsequent stability analysis, the implementable observer update laws for the estimates are designed using (3-4)-(3-6) as

$$\frac{d}{dt}(\hat{d}_{s_i/c}(t)) \triangleq \begin{cases} \eta_{s_i,1}(t), & t < \tau_{s_i}, \\ \eta_{s_i,1}(t) + k_1(\nu_{s_i,1}(t) - \hat{d}_{s_i/c}(t)), & t \geq \tau_{s_i}, \end{cases} \quad (3-10)$$

$$\frac{d}{dt}(\hat{d}_{k/c}(t)) \triangleq \begin{cases} \eta_{s_i,2}(t), & t < \tau_{s_i}, \\ \eta_{s_i,2}(t) + k_2(\nu_{s_i,2}(t) - \hat{d}_{k/c}(t)), & t \geq \tau_{s_i}, \end{cases} \quad (3-11)$$

and

$$\frac{d}{dt}(\hat{d}_{s_i/k}(t)) \triangleq \begin{cases} 0, & t < \tau_{s_i}, \\ k_3(\mathcal{X}_{s_i} - \hat{d}_{s_i/k}(t)), & t \geq \tau_{s_i}, \end{cases} \quad (3-12)$$

where $k_1, k_2, k_3 \in \mathbb{R}_{>0}$ are constants. Taking the time derivative of (3-7)-(3-9), and substituting (3-4)-(3-9), (2-4)-(2-6), and (3-10)-(3-12) yields

$$\frac{d}{dt}(\tilde{d}_{s_i/c}(t)) = \begin{cases} 0, & t < \tau_{s_i}, \\ -k_1\tilde{d}_{s_i/c}(t), & t \geq \tau_{s_i}, \end{cases} \quad (3-13)$$

$$\frac{d}{dt}(\tilde{d}_{k/c}(t)) = \begin{cases} 0, & t < \tau_{s_i}, \\ -k_2\tilde{d}_{k/c}(t), & t \geq \tau_{s_i}, \end{cases} \quad (3-14)$$

and

$$\frac{d}{dt} \left(\tilde{d}_{s_i/k}(t) \right) = \begin{cases} 0, & t < \tau_{s_i}, \\ -k_3 \tilde{d}_{s_i/k}(t), & t \geq \tau_{s_i}, \end{cases} \quad (3-15)$$

implying for all time $t \geq \tau_{s_i}$, the estimation error derivatives are negative definite functions of the estimation errors. The form of the update laws in (3-10)-(3-12) are implementable and used in practice, while the form of the time-derivative of the estimation errors in (3-13)-(3-15) are analytical and provided to facilitate the subsequent analysis.

3.2 Extended Observer Update Law for Euclidean Distance to Features from Camera

The subsequent analysis demonstrates that (3-7) and (3-13) will remain bounded while $t < \tau_{s_i}$. However, after sufficient data is gathered, for all $t \geq \tau_{s_i}$, (3-7) is bounded by an exponential envelope. The delay required to get sufficient excitation may reduce transient performance (i.e., the error is not guaranteed to reduce until after time $t \geq \tau_{s_i}$) which is a disadvantage compared to previous approaches such as [1], which improve estimation errors by estimating optical flow. Motivated by the optical flow estimator form of the inverse depth estimator in [1], the time rate of change of $\underline{u}_{s_i/c}(t)$ is approximated and used to provide additional information to the estimator in (3-10) which will improve transient performance until sufficient excitation occurs.

The time derivative of $\underline{u}_{s_i/c}(t)$ is

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

and

$$\xi_{s_i}^T(t) \xi_{s_i}(t) d_{s_i/c}(t) = \xi_{s_i}^T(t) \rho_{s_i}(t), \quad (3-16)$$

where $\xi_{s_i}(t) \triangleq \left(\frac{d}{dt} (\underline{u}_{s_i/c}(t)) + \underline{\omega}_c^\times(t) \underline{u}_{s_i/c}(t) \right)$, $\rho_{s_i}(t) \triangleq \left(\underline{u}_{s_i/c}(t) \underline{u}_{s_i/c}^T(t) - I_{3 \times 3} \right) \underline{v}_c(t)$,

$$\underline{\omega}_c^\times(t) \triangleq \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}, \text{ and } I_{3 \times 3} \triangleq \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \text{ To aid in the subsequent analysis let}$$

$\mu_{s_i}(t) \triangleq \eta_{s_i,1}(t) + k_\xi \left(\xi_{s_i}^T(t) \rho_{s_i}(t) - \xi_{s_i}^T(t) \xi_{s_i}(t) \hat{d}_{s_i/c}(t) \right)$, then an extended version of the estimator in (3–10) is designed as

$$\frac{d}{dt} \left(\hat{d}_{s_i/c}(t) \right) \triangleq \begin{cases} \mu_{s_i}(t), & t < \tau_{s_i}, \\ \mu_{s_i}(t) + k_1 \left(\nu_{s_i,1}(t) - \hat{d}_{s_i/c}(t) \right), & t \geq \tau_{s_i}, \end{cases} \quad (3-17)$$

where $k_\xi \in \mathbb{R}_{>0}$. Using (3–7) and (3–16) in (3–17) then simplifying yields

$$\frac{d}{dt} \left(\hat{d}_{s_i/c}(t) \right) = \begin{cases} \eta_{s_i,1}(t) + k_\xi \Xi_{s_i}(t) \tilde{d}_{s_i/c}(t), & t < \tau_{s_i}, \\ \eta_{s_i,1}(t) + (k_1 + k_\xi \Xi_{s_i}(t)) \tilde{d}_{s_i/c}(t), & t \geq \tau_{s_i}, \end{cases} \quad (3-18)$$

where $\Xi_{s_i}(t) \triangleq \xi_{s_i}^T(t) \xi_{s_i}(t)$. Substituting (3–18) into the time derivative of (3–7) yields

$$\frac{d}{dt} \left(\tilde{d}_{s_i/c}(t) \right) = \begin{cases} -k_\xi \Xi_{s_i}(t) \tilde{d}_{s_i/c}(t), & t < \tau_{s_i}, \\ -(k_1 + k_\xi \Xi_{s_i}(t)) \tilde{d}_{s_i/c}(t), & t \geq \tau_{s_i}. \end{cases} \quad (3-19)$$

Remark 3.1. Under Assumption 3.1, $\Xi_{s_i}(t) \geq 0$ since $\|\underline{v}_c(t)\|$ may be zero for any period of time; however, for Assumption 3.1 to be satisfied, there will be times where $\Xi_{s_i}(t) > 0$. Specifically, there will exist a set of times $\mathcal{T}_{s_i} \subset \bigcup_{h_i=1}^N (t_{h_i} - \varsigma, t_{h_i})$ such that $\Xi_{s_i}(t) > 0, \forall t \in \mathcal{T}_{s_i}$, where h_i, t_{h_i} are from (3–3), implying the design in (3–17) may improve transient performance under Assumption 3.1.

3.3 Stability Analysis

Since the observer in (3–17) is an extension of (3–10), the resulting stability analysis of (3–10) is identical to Theorem 3.1 and is excluded. Let $\eta(t) \triangleq$

$\begin{bmatrix} \tilde{d}_{s_i/c}(t) & \tilde{d}_{k/c}(t) & \tilde{d}_{s_i/k}(t) \end{bmatrix}^T$ and $V(\eta(t)) : \mathbb{R}^3 \rightarrow \mathbb{R}$ be a candidate Lyapunov function defined as

$$V(\eta(t)) \triangleq \frac{1}{2} \eta^T(t) \eta(t), \quad (3-20)$$

which can be bounded as $\frac{1}{2} \|\eta(t)\|^2 \leq V(\eta(t)) \leq \frac{1}{2} \|\eta(t)\|^2$.

Theorem 3.1. *The observer update laws defined in (3-11), (3-12), and (3-17) ensure the estimation errors in $\eta(t)$ are bounded and globally exponentially stable in the sense that*

$$\|\eta(t)\| \leq \|\eta(0)\| \exp(\beta \tau_{s_i}) \exp(-\beta t). \quad (3-21)$$

Proof. Taking the time derivative of (3-20) then substituting the error derivatives in (3-14), (3-15), and (3-19), results in the upper bound

$$\frac{d}{dt} (V(\eta(t))) \leq \begin{cases} 0, & t < \tau_{s_i}, \\ -2\beta V(\eta(t)), & t \geq \tau_{s_i}, \end{cases} \quad (3-22)$$

where $\beta = \min\{k_1, k_2, k_3\}$. From (3-20) and (3-22), [116, Theorem 8.4] can be invoked to conclude that $\|\eta(t)\|^2 \leq \|\eta(0)\|^2, \forall t \leq \tau_{s_i}$. From [116, Theorem 4.10], $\|\eta(t)\|^2 \leq \|\eta(\tau_{s_i})\|^2 \exp(2\beta \tau_{s_i}) \exp(-2\beta t), \forall t \geq \tau_{s_i}$. Evaluating the first bound on $\|\eta(t)\|^2$ at $t = \tau_{s_i}$ then substituting into the second bound on $\|\eta(t)\|^2$ and taking the square root yields (3-21). □

3.4 Experimental Results

Fifteen experiments are provided to demonstrate the performance of the developed observers. The performance of the developed observers in (3-10)-(3-12) and (3-17) were tested using the Eigen3, OpenCV, and ROS c++ libraries (cf., [117], [118], and [119], respectively). A Kobuki Turtlebot with a 1920 × 1080 monochrome iDS uEye camera, shown in Figure 3-1, provides images at 30 Hz. Features were extracted from images of a checkerboard, shown in Figure 3-1, with 8 × 6 corners (48 total features)



Figure 3-1. Photo courtesy of author. Image shows the checkerboard, Kobuki Turtlebot, and an iDS uEye camera used for experiments.

where each square is 0.06 meters \times 0.06 meters. The linear and angular velocity of the camera were calculated using the Turtlebot wheel encoders and a gyroscope at 50 Hz. An Optitrack motion capture system operating at 120 Hz measured the pose of the camera and checkerboard, allowing for the position of each feature relative to the camera to be known for comparison. Image processing and estimators executed simultaneously on a computer with an Intel i7 processor running at 3.4 GHz. The error of the distance estimators in (3-10) and (3-17) are compared to the estimator in [1] and an EKF. Since the estimator in [1] and the EKF estimate the inverse depth (i.e., $\frac{1}{z_{s_i/c}(t)}$, where $z_{s_i/c}(t)$ is the depth to feature s_i from c expressed in \mathcal{F}_c), while the estimators in (3-10) and (3-17) estimate the distance, $z_{s_i/c}(t)$ (the third element of $u_{s_i/c}(t)$ $d_{s_i/c}(t)$) is used to compare the four methods.

For each experiment, the Turtlebot started approximately 3 meters away from the checkerboard, and various trajectories were taken, shown in Figure 3-2, while maintaining the checkerboard in the FOV. In each experiment, the Turtlebot initially

started at rest, and after traveling 2.5 meters the estimators were stopped to provide a large baseline. After the Turtlebot started its motion, the Turtlebot traveled without stopping until after the estimators were stopped in an effort to have the ideal conditions for estimation (i.e., continuous motion of the features in the camera FOV and continuous linear motion of the camera as is required for [1] and the EKF). The initial distances for the estimators in (3–10), (3–12), (3–17), the estimator in [1] and the EKF were initialized with a depth of 0.5 meters. The estimator in (3–11) was initialized to 0.0 meters. The gains for (3–10)-(3–12) and (3–17) were selected as $k_1 = k_2 = k_3 = 25.0$ and $k_\xi = 25.0k_1$, respectively. The maximum value for ς was 5 seconds. The 48 feature estimates were combined using a mean at each instance to update (3–11). The gain for the method in [1] was selected to be 100.0. The covariance matrices for the EKF were determined through experimentation and values found to have low steady state error and fast convergence were $R = r \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ for the measurement covariance,

$$Q = r \begin{bmatrix} 100 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 100000 \end{bmatrix} \text{ for the process covariance, and } P(0) = r \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 150000 \end{bmatrix} \text{ for the initial state covariance, respectively, where } r = 0.00001.$$

Remark 3.2. For a general system, the optimal method to select good data and remove bad data for (3–3) (e.g., due to noise or parameter changes) remains an open problem and is often left to intuition. In these experiments, the selection of data was based off of knowledge about approximate noise magnitudes in feature tracking and velocity measurements. Specifically, data was only selected if $\|\mathcal{Y}_{s_i}(t)\| \geq \epsilon_y$ and $\|\mathcal{U}_{s_i}(t)\| \geq \epsilon_u$ where $\epsilon_y, \epsilon_u \in \mathbb{R}_{>0}$ are values selected based on trial and error. Because $\mathcal{Y}_{s_i}(t)$ is full column rank when $\|\mathcal{Y}_{s_i}(t)\| \geq \epsilon_y$ and $\|\mathcal{U}_{s_i}(t)\| \geq \epsilon_u$, the value of $d_{s_i/k}$ approximated by $\mathcal{Y}_{s_i}(t)$ and $\mathcal{U}_{s_i}(t)$ can be determined. Given $d_{s_i/k} > 0$, and some knowledge about reasonable values for the distances, values of $d_{s_i/k}$ can be determined and only $\mathcal{Y}_{s_i}(t)$

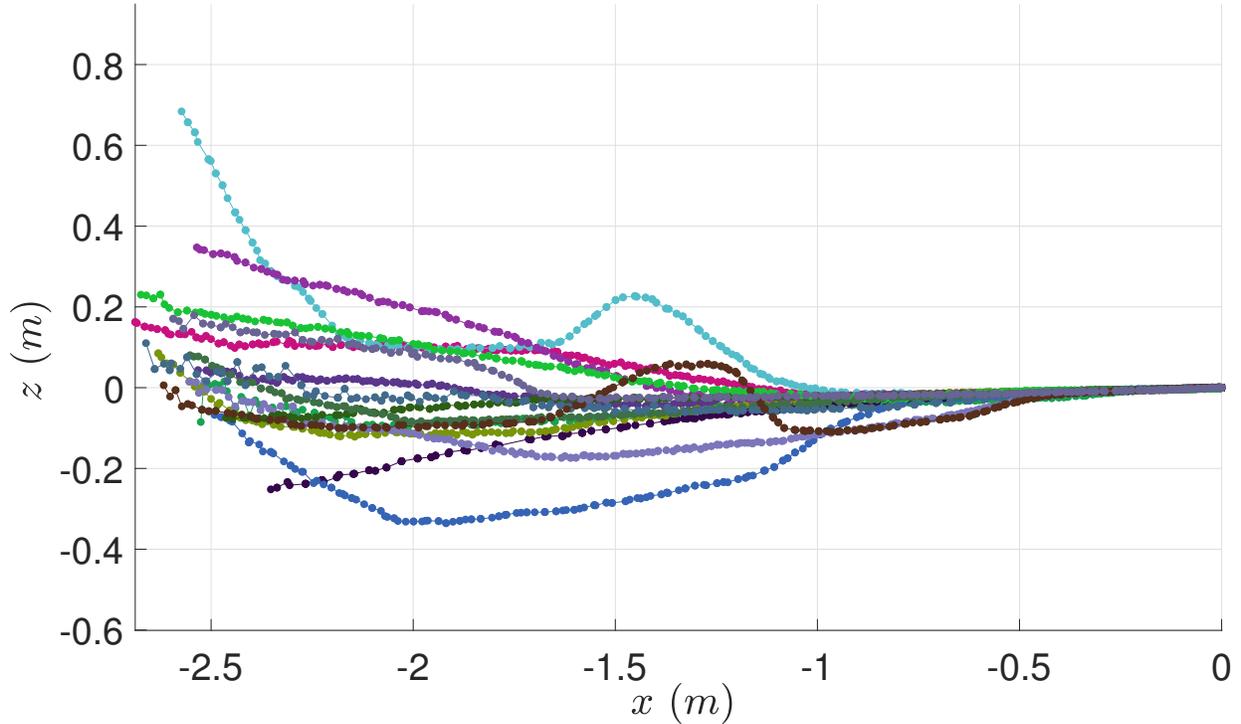


Figure 3-2. The camera trajectories for each of the 15 experiments.

and $U_{s_i}(t)$ values that had $d_{s_i/k}$ estimates falling in these bounds are saved to $\Sigma_{y_{s_i}}$ and $\Sigma_{u_{s_i}}$. The values for ϵ_y and ϵ_u were $\epsilon_y = \epsilon_u = 0.1$ and the bounds on the distance were selected as 0.5 meters and 6.0 meters.

A comparison of the example performance over time of the estimators is shown in Figure 3-3 and Tables 3-1-3-3, where before learning refers to $t < \max\{\tau_{s_i}\}$ and after learning refers to $t \geq \max\{\tau_{s_i}\}$. Figure 3-3 shows a comparison of the sum of the norm of each depth error across the 48 features (i.e., $\sum_{s_i=1}^{48} \|\tilde{z}_{s_i/c}(t)\|$) on the checkerboard for the estimators in (3-10), (3-17), [1], and the EKF, respectively. As shown in Figure 3-3, the EKF estimator starts converging the fastest, but reaches steady state slower than the estimators in (3-10), (3-17), and [1]. However, after converging, the EKF has a similar error to the estimators in (3-10) and (3-17). Figure 3-3 also shows that the estimator in (3-10) does not converge until sufficient learning occurs (at $t = 3.6$ seconds for experiment 11). The extension in Section 3.2 shows an advantage of using current input-output data in the estimator, as shown by the mean RMS errors in Table 3-1 and

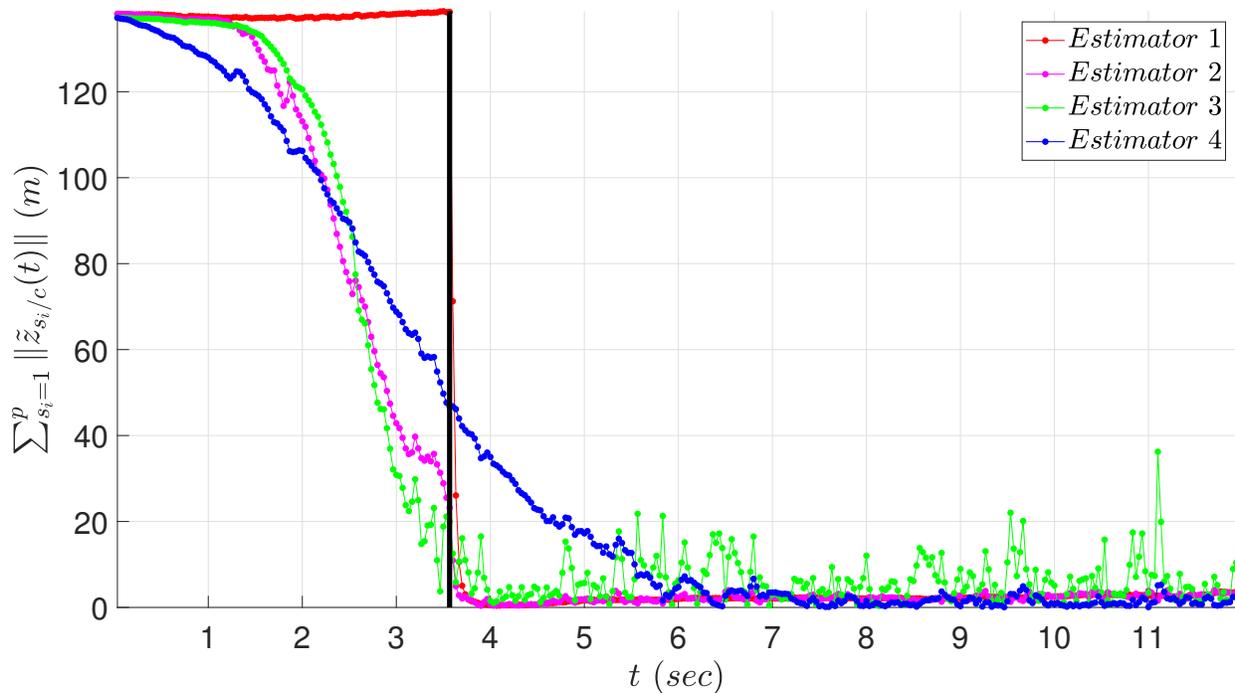


Figure 3-3. The sum of the norm of each depth error across the 48 features (i.e., $\sum_{s_i=1}^{48} \|\tilde{z}_{s_i/c}(t)\|$) for Experiment 11. Estimator 1 (red) refers to (3–10), Estimator 2 (magenta) refers to (3–17), Estimator 3 (green) refers to [1], and Estimator 4 (blue) refers to the EKF. The black vertical line indicates the time when enough information was collected for learning.

Table 3-1. RMS Depth Error and Position Error in Meters Over 15 Experiments

Experiment	Estimator 1	Estimator 2	Estimator 3	Estimator 4	Trajectory
1	70.499	55.525	58.325	55.435	0.017
2	66.358	42.065	52.827	42.838	0.026
3	80.466	64.701	67.616	63.706	0.037
4	76.250	59.583	65.593	59.565	0.032
5	79.419	63.244	70.798	65.674	0.032
6	82.611	69.300	71.996	68.558	0.018
7	65.971	48.192	57.287	46.392	0.020
8	73.864	60.002	64.290	59.858	0.027
9	77.699	61.468	69.227	60.708	0.020
10	72.053	55.916	63.679	55.771	0.023
11	75.472	59.757	59.916	60.798	0.016
12	77.571	63.623	66.451	63.686	0.022
13	83.663	67.817	72.995	68.274	0.036
14	74.030	58.764	66.941	59.245	0.030
15	77.742	56.067	64.983	56.612	0.029
Mean	75.578	59.068	64.862	59.141	0.026
Standard Deviation	5.086	6.819	5.522	6.938	0.007

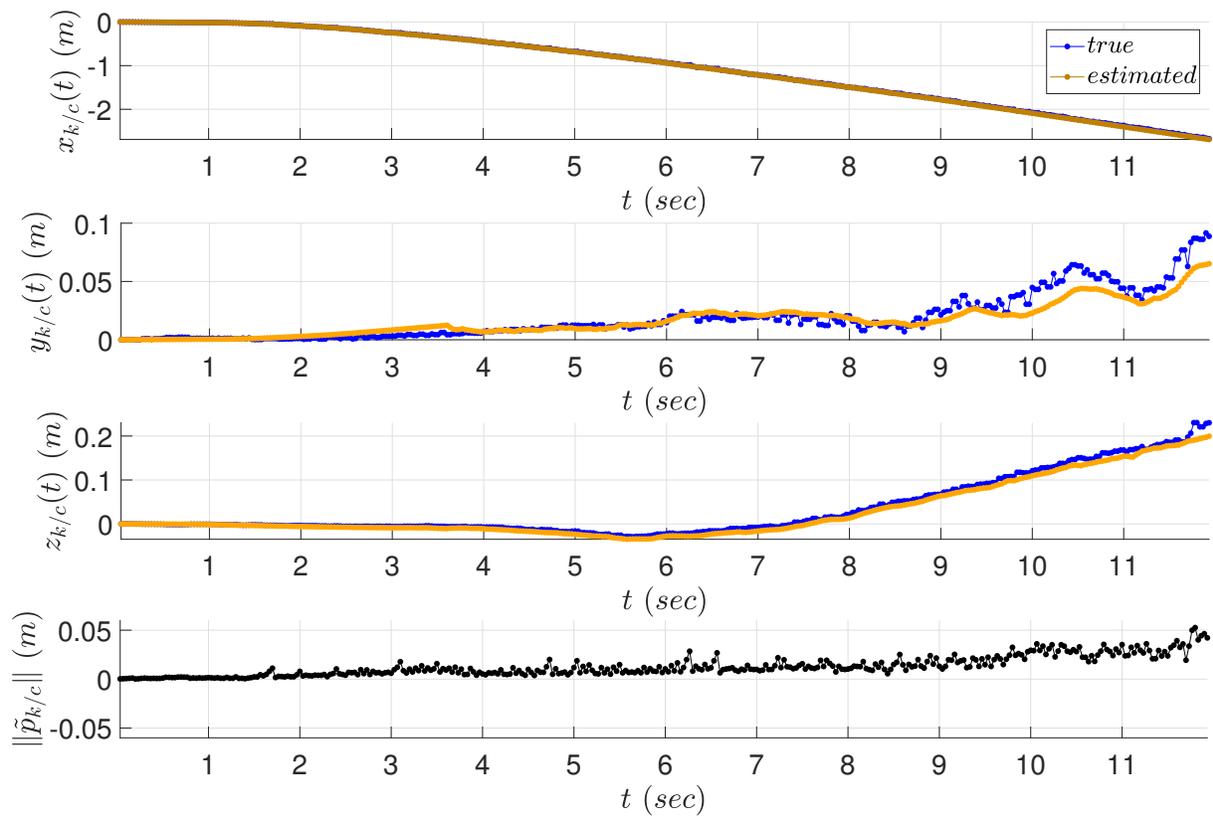


Figure 3-4. The position error of the camera and the distance error using the estimator in (3-11) for Experiment 11.

Table 3-2. RMS Depth Error and Position Error in Meters Over 15 Experiment Before Learning

Experiment	Estimator 1	Estimator 2	Estimator 3	Estimator 4	Trajectory
1	133.393	105.884	109.492	104.984	0.007
2	137.855	87.811	107.410	88.252	0.005
3	137.133	110.694	114.093	108.139	0.005
4	134.047	105.321	113.742	104.732	0.005
5	138.285	110.761	120.448	112.189	0.005
6	137.122	115.571	118.032	113.951	0.005
7	126.254	92.370	103.992	87.503	0.004
8	128.662	104.859	109.866	103.421	0.004
9	136.862	108.817	119.151	106.569	0.005
10	128.454	100.033	110.136	98.836	0.004
11	137.677	109.201	107.660	110.506	0.006
12	132.034	108.494	111.989	108.105	0.005
13	139.927	113.816	119.680	112.477	0.009
14	131.922	104.987	114.095	101.154	0.005
15	142.222	103.213	116.301	103.489	0.006
Mean	134.790	105.455	113.073	104.287	0.005
Standard Deviation	4.440	7.212	4.825	7.645	0.001

Table 3-3. RMS Depth Error and Position Error in Meters Over 15 Experiment After Learning

Experiment	Estimator 1	Estimator 2	Estimator 3	Estimator 4	Trajectory
1	13.374	3.979	12.838	8.567	0.017
2	10.655	1.868	13.894	8.166	0.029
3	12.513	2.657	14.246	10.123	0.045
4	12.517	1.180	15.656	7.629	0.038
5	13.431	1.903	20.777	17.759	0.038
6	13.359	2.414	16.889	7.562	0.022
7	10.403	3.502	22.177	10.390	0.024
8	11.632	3.804	17.099	11.610	0.033
9	12.661	1.683	19.959	9.847	0.024
10	11.162	1.861	20.035	9.457	0.028
11	10.239	2.828	13.483	8.119	0.019
12	9.996	1.576	12.845	7.833	0.027
13	13.093	3.261	19.967	16.608	0.044
14	11.044	3.090	24.677	21.283	0.036
15	13.885	2.599	18.542	8.494	0.034
Mean	11.998	2.547	17.539	10.897	0.031
Standard Deviation	1.284	0.831	3.567	4.080	0.008

3-2. Specifically, the estimator in (3-10) is at a disadvantage to the other estimators before sufficient excitation has occurred, while the estimator in (3-17) starts converging to the true depths at a similar rate as the estimator in [1]. The average RMS error of (3-10) is more than 10 meters greater than the other estimators over the entirety of each experiment, and more than 20 meters greater before learning. However, Table 3-3 shows that the average error of (3-10) after learning is only 1 meter greater than the EKF on average.

The extension in Section 3.2, specifically the design in (3-17), improves the error convergence of (3-10) such that the RMS error is lower than the EKF on average. As shown in Table 3-1, the average error over the entire experiment runtime was 59.068 meters for (3-17) compared to 59.141 meters for the EKF. After learning, the average RMS error for the estimator in (3-17) was smaller (2.547 meters) compared to the EKF (10.897 meters). However, as shown in Table 3-2, the RMS error before learning was smaller for the EKF compared to (3-17), where the errors were 104.287 meters for the EKF and 105.455 meters for the estimator in (3-17). Additionally, Tables 3-1-3-3 show that the design in (3-17) has a smaller RMS error than the design in [1] on average. Figure 3-4 and Tables 3-1-3-3 show the position error using (3-11) is small with an average RMS error of 0.026 meters over the entire run; 0.005 meters before learning and 0.031 after learning, which is approximately 1.2% error relative to trajectory length. The error increase after learning is a result of noise, which as shown in Figure 3-3 and Table 3-3 causes the depth error to remain small but bounded at approximately 1.8% of the initial error. These experimental results demonstrate the ability of the observer in (3-10) and (3-17) to leverage both immediate information and learning to both converge quickly with low RMS error and maintain a low RMS error after converging.

3.5 Summary

Novel observers using a single camera and structure from motion theory are developed to estimate the Euclidean distance to features on a stationary object and

the Euclidean trajectory the camera takes while observing the object. Unlike previous results that estimate the inverse depth to features, the developed observer for estimating the Euclidean distance to features does not require the positive depth constraint. A Lyapunov-based stability analysis shows the observer error exponentially converges where persistence of excitation is replaced by finite excitation through the use of ICL. An experimental comparison of the developed estimator to existing estimators shows that it achieves lower RMS error when comparing feature depth estimates on average, and the RMS error of the position also remains within 1.8%.

CHAPTER 4 POSITION ESTIMATION USING A MONOCULAR CAMERA SUBJECT TO INTERMITTENT FEEDBACK

In this chapter, an extension to the learning approaches in [32], [45], and Chapter 3, is developed that applies a new learning strategy that maintains a continuous estimate of the position of the camera and estimates the structure of features as they come into the camera's FOV. Furthermore, 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. Additionally, this approach shows how the extended observer removes the positive depth constraint required by all previous structure from motion approaches. Using this approach, a camera may travel over large distances without keeping specific features in the FOV for all time and allow objects to permanently leave the FOV if necessary. A Lyapunov based stability analysis proves that the observers for estimating the path of the camera as well as the structure of each set of objects are globally exponentially stable while features are in the FOV. A switched systems analysis is used to develop dwell-time conditions to indicate how long a feature must be tracked to ensure the distance estimation error is below a threshold. After the distance estimates have converged below the threshold, the feature may be used to update the position of the camera. If a feature does not satisfy the dwell-time condition, it is never used to update the position of the agent. Furthermore, the approach does not require a new set of features to be in the FOV when older features leave the FOV. Finally, if a recognized landmark enters the FOV, the feedback is used to compensate for drift error. The results in this chapter demonstrate that the observer and predictor strategy outperforms a predictor-only strategy (cf., [62] and [63]) when feedback is unavailable, provided structure estimation error is less than the thresholds used to develop the developed dwell-times.

4.1 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 [32] and [45] is used to learn a constant unknown distance, $d_{s_{j,i}/k_j}$. Specifically, (2-11) and (2-12) are integrated over a time window $\varsigma \in \mathbb{R}_{>0}$ yielding

$$\begin{bmatrix} d_{s_{j,i}/c}(t) \\ d_{k_j/c}(t) \end{bmatrix} - \begin{bmatrix} d_{s_{j,i}/c}(t-\varsigma) \\ d_{k_j/c}(t-\varsigma) \end{bmatrix} = - \int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{s_{j,i}/c}^T(l) \\ \underline{u}_{k_j/c}^T(l) \end{bmatrix} \underline{v}_c(l) dl, t > \varsigma, \quad (4-1)$$

where ς may be constant in size or change over time. As described in (2-11) and

(2-12), while $s_{j,i} \in \mathcal{P}_{s_j}(t)$, $-\int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{s_{j,i}/c}^T(l) \\ \underline{u}_{k_j/c}^T(l) \end{bmatrix} \underline{v}_c(l) dl$ 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 (4-1) is unknown.

However, when $\sigma_{s_{j,i}}(t) = a$, the relationship in (2-9) may be utilized in (4-1) yielding

$$\mathcal{Y}_{s_{j,i}}(t) d_{s_{j,i}/k_j} = \mathcal{U}_{s_{j,i}}(t), t > \zeta_j^a, \quad (4-2)$$

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

$$\mathcal{Y}_{s_{j,i}}(t) \triangleq \begin{cases} \left(\psi_{s_{j,i}}^a(t) - \psi_{s_{j,i}}^a(\zeta_{s_{j,i}}^{a_l}) \right) t \in \left(\zeta_{s_{j,i}}^{a_l}, \zeta_{s_{j,i}}^{a_l} + \varsigma \right], \\ \left(\psi_{s_{j,i}}^a(t) - \psi_{s_{j,i}}^a(t-\varsigma) \right), t \in \left(\zeta_{s_{j,i}}^{a_l} + \varsigma, \zeta_{s_{j,i}}^{u_l} \right], \\ 0_{2 \times 1}, t \in \left(\zeta_{s_{j,i}}^{u_l}, \zeta_{s_{j,i}}^{a_{l+1}} \right], \end{cases}$$

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

and $\zeta_{s_{j,i}}^{a_l}, \zeta_{s_{j,i}}^{u_l} \in \left[\zeta_j^a, \zeta_{s_{j,i}}^u \right]$ 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 (4-2) by $\mathcal{Y}_{s_{j,i}}^T(t)$ yields

$$\mathcal{Y}_{s_{j,i}}^T(t) \mathcal{Y}_{s_{j,i}}(t) d_{s_{j,i}/k_j} = \mathcal{Y}_{s_{j,i}}^T(t) \mathcal{U}_{s_{j,i}}(t). \quad (4-3)$$

In general, $\mathcal{Y}_{s_{j,i}}(t)$ will is 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 (4-3) may be evaluated at any instance in time and summed together (i.e., history stacks) yielding

$$\Sigma_{\mathcal{Y}_{s_{j,i}}} d_{s_{j,i}/k_j} = \Sigma_{\mathcal{U}_{s_{j,i}}}, \quad (4-4)$$

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

Assumption 4.1. 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 $t > \tau_{s_{j,i}}$, $\lambda_{\min} \left\{ \Sigma_{\mathcal{Y}_{s_{j,i}}} \right\} > \lambda_\tau$, where $\tau_{s_{j,i}} \in \left(\zeta_j^a, \zeta_{s_{j,i}}^u \right)$ 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 $\mathcal{A}_{s_j}(t)$.¹

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 no guarantee that the motion of the camera will be sufficient before every feature leaves the camera's FOV permanently. Camera motion in Assumption 4.1 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 $\lambda_{\min} \left\{ \Sigma_{\mathcal{Y}_{s_{j,i}}} \right\} > \lambda_{\tau}$. 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 each feature.

If motion occurs as discussed in Assumption 4.1, 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 (4-4) yielding

$$d_{s_{j,i}/k_j} = \mathcal{X}_{s_{j,i}}, \quad s_{j,i} \in \mathcal{A}_{s_j}(t), \quad (4-5)$$

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 (4-5) into (2-8) yields

$$Y_{s_{j,i}}(t) \begin{bmatrix} d_{s_{j,i}/c}(t) \\ d_{k_j/c}(t) \end{bmatrix} = R_{k_j/c}(t) \underline{u}_{s_{j,i}/k_j} \mathcal{X}_{s_{j,i}}, \quad s_{j,i} \in \mathcal{A}_{s_j}(t). \quad (4-6)$$

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 [1]. 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

$$\frac{d}{dt} \left(\underline{u}_{s_{j,i}/c}(t) \right) = -\underline{\omega}_c^\times(t) \underline{u}_{s_{j,i}/c}(t)$$

¹ See [114] or [115] for some examples of methods for selecting data to satisfy the assumption.

$$+\frac{1}{d_{s_{j,i}/c}(t)}\left(\underline{u}_{s_{j,i}/c}(t)\underline{u}_{s_{j,i}/c}^T(t)-I_{3\times 3}\right)\underline{v}_c(t), \quad (4-7)$$

implying

$$\xi_{s_{j,i}}(t)d_{s_{j,i}/c}(t) = \rho_{s_{j,i}}(t), \quad (4-8)$$

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

Substituting (2-10) into (4-8) yields

$$\xi_{s_{j,i}}(t)\psi_{s_{j,i}}^u(t)\begin{bmatrix} d_{k_j/c}(t) \\ d_{s_{j,i}/k_j} \end{bmatrix} = \rho_{s_{j,i}}(t). \quad (4-9)$$

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 (4-5)

and (4-6) yields

$$Y_{\mathcal{X}_{s_{j,i}}}(t)\eta_{s_{j,i}}(t) = u_{\mathcal{X}_{s_{j,i}}}(t), \quad (4-10)$$

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

(4-8) and (4-9) yields

$$Y_{\xi_{s_{j,i}}}(t)\eta_{s_{j,i}}(t) = u_{\xi_{s_{j,i}}}(t), \quad (4-11)$$

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

4.2 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

$$\tilde{d}_{s_{j,i}/c}(t) \triangleq d_{s_{j,i}/c}(t) - \hat{d}_{s_{j,i}/c}(t), \quad (4-12)$$

$$\tilde{d}_{k_{j,i}/c}(t) \triangleq d_{k_{j,i}/c}(t) - \hat{d}_{k_{j,i}/c}(t), \quad (4-13)$$

and

$$\tilde{d}_{s_{j,i}/k_j}(t) \triangleq d_{s_{j,i}/k_j}(t) - \hat{d}_{s_{j,i}/k_j}(t), \quad (4-14)$$

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 $d_{s_{j,i}/k_j}$, 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 (4-12)-(4-14) as

$$\tilde{\eta}_{s_{j,i}}(t) \triangleq \eta_{s_{j,i}}(t) - \hat{\eta}_{s_{j,i}}(t), \quad (4-15)$$

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 $\tilde{\eta}_{s_{j,i}}(t) \triangleq \begin{bmatrix} \tilde{d}_{s_{j,i}/c}(t) \\ \tilde{d}_{k_{j,i}/c}(t) \\ \tilde{d}_{s_{j,i}/k_j}(t) \end{bmatrix}$.

4.2.1 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

$$\frac{d}{dt}(\hat{\eta}_{s_{j,i}}(t)) \triangleq \begin{cases} 0_{3 \times 1}, & s_{j,i} \in \mathcal{P}_{s_j}^c(t), \\ \text{proj}(\mu_{\xi_{s_{j,i}}}(t)), & s_{j,i} \in \mathcal{P}_{s_j}(t), \\ \text{proj}(\mu_{\xi_{s_{j,i}}}(t) + \mu_{\chi_{s_{j,i}}}(t)), & s_{j,i} \in \mathcal{A}_{s_j} \cap \mathcal{P}_{s_j}(t), \end{cases} \quad (4-16)$$

where $\mu_{\xi_{s_{j,i}}}(t) \triangleq \begin{bmatrix} -\underline{u}_{s_{j,i}/c}^T(t) \underline{v}_c(t) \\ -\underline{u}_{k_j/c}^T(t) \underline{v}_c(t) \\ 0 \end{bmatrix} + K_\xi Y_{\xi_{s_{j,i}}}^T(t) u_{\xi_{s_{j,i}}}(t) - K_\xi Y_{\xi_{s_{j,i}}}(t) Y_{\xi_{s_{j,i}}}(t) \hat{\eta}_{s_{j,i}}(t)$,
 $\mu_{\mathcal{X}_{s_{j,i}}}(t) \triangleq K_{\mathcal{X}} Y_{\mathcal{X}_{s_{j,i}}}^T(t) u_{\mathcal{X}_{s_{j,i}}}(t) - K_{\mathcal{X}} Y_{\mathcal{X}_{s_{j,i}}}(t) Y_{\mathcal{X}_{s_{j,i}}}(t) \hat{\eta}_{s_{j,i}}(t)$, and $K_\xi, K_{\mathcal{X}} \in \mathbb{R}^{3 \times 3}$
are positive definite gain matrices, and $proj(\cdot)$ 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 $\underline{d}_1 \leq \hat{d}_{k_j/c}(t)$.²

Taking the time derivative of (4–15), substituting (2–11)–(2–13), (4–10), (4–11), (4–15), and (4–16), and simplifying yields

$$\frac{d}{dt}(\tilde{\eta}_{s_{j,i}}(t)) = \begin{cases} \begin{bmatrix} -\underline{u}_{s_{j,i}/c}^T(t) \underline{v}_c(t) \\ -\underline{u}_{k_j/c}^T(t) \underline{v}_c(t) \\ 0 \end{bmatrix}, & s_{j,i} \in \mathcal{P}_{s_j}^c(t), \\ -\Psi_{\xi_{s_{j,i}}}(t) \tilde{\eta}_{s_{j,i}}(t), & s_{j,i} \in \mathcal{P}_{s_j}(t), \\ -\Psi_{s_{j,i}}(t) \tilde{\eta}_{s_{j,i}}(t), & s_{j,i} \in \mathcal{A}_{s_j}(t) \cap \mathcal{P}_{s_j}(t), \end{cases} \quad (4-17)$$

where $\Psi_{s_{j,i}}(t) \triangleq \Psi_{\xi_{s_{j,i}}}(t) + \Psi_{\mathcal{X}_{s_{j,i}}}(t)$, $\Psi_{\xi_{s_{j,i}}}(t) \triangleq K_\xi Y_{\xi_{s_{j,i}}}^T(t) Y_{\xi_{s_{j,i}}}(t)$, 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.

² See [32, Appendix E] or [33, Remark 3.7] for examples on implementing a smooth projection operator

4.2.2 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

$$V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t)) \triangleq \frac{1}{2} \tilde{\eta}_{s_{j,i}}^T(t) \tilde{\eta}_{s_{j,i}}(t), \quad (4-18)$$

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 4.1. *The observer update law defined in (4-16) 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*

$$\|\tilde{\eta}_{s_{j,i}}(t)\| \leq \|\tilde{\eta}_{s_{j,i}}(\zeta_j^a)\|, \quad s_{j,i} \in \mathcal{P}_{s_j}(t). \quad (4-19)$$

Proof. Taking the time derivative of (4-18), substituting (4-17) 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

$$\frac{d}{dt}(V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t))) \leq 0. \quad (4-20)$$

Invoking [116, Theorem 8.4] on (4-20) yields $\|\tilde{\eta}_{s_{j,i}}(t)\|^2 \leq \|\tilde{\eta}_{s_{j,i}}(\zeta_j^a)\|^2$ and taking the square root yields (4-19). \square

Lemma 4.2. *The observer update law defined in (4-16) 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*

$$\|\tilde{\eta}_{s_{j,i}}(t)\| \leq \|\tilde{\eta}_{s_{j,i}}(\zeta_{s_{j,i}}^{u_l})\|, \quad s_{j,i} \in \mathcal{A}_{s_j}(t) \cap \mathcal{P}_{s_j}(t) \cap \mathcal{B}_{s_j}^c(t). \quad (4-21)$$

Proof. Taking the time derivative of (4-18), substituting (4-17) 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

$$\frac{d}{dt}(V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t))) \leq 0. \quad (4-22)$$

Invoking [116, Theorem 8.4] on (4-22) 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 (4-21). \square

Lemma 4.3. *The observer update law defined in (4–16) 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*

$$\|\tilde{\eta}_{s_{j,i}}(t)\| \leq \|\tilde{\eta}_{s_{j,i}}(\zeta_{s_{j,i}}^{a_l})\| \exp\left(-\beta(t - \zeta_{s_{j,i}}^{a_l})\right), s_{j,i} \in \mathcal{B}_{s_j}(t). \quad (4-23)$$

Proof. Taking the time derivative of (4–18), substituting (4–17) 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

$$\frac{d}{dt}(V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t))) \leq -2\beta V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t)), \quad (4-24)$$

where $\beta \triangleq \lambda_a \lambda_{\min}\{K_{\mathcal{X}}\}$. Invoking [116, Theorem 4.10] on (4–24) 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 (4–23). \square

Theorem 4.1. *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 (4–16) is globally uniformly ultimately bounded (GUUB) as*

$$\|\tilde{\eta}_{s_{j,i}}(\zeta_{s_{j,i}}^u)\| \leq \|\tilde{\eta}_{s_{j,i}}(\zeta_{s_{j,i}}^a)\| \exp\left(-\beta \sum_{l \in \mathcal{L}_{s_{j,i}}} \Delta \zeta_{s_{j,i}}^{a_l}\right). \quad (4-25)$$

Proof. Using the bounds in (4–19), (4–21), and (4–23) 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 (4–25). \square

As shown in (4–25), 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 (4–15) and (4–17) for the case when feature $s_{j,i} \in \mathcal{P}_{s_j}^c(t)$ implies $\frac{d}{dt}(V_{s_{j,i}}(\tilde{\eta}_{s_{j,i}}(t))) \leq \|\tilde{d}_{s_{j,i}/c}(t)\| \|\underline{v}_c(t)\| + \|\tilde{d}_{k_{j,i}/c}(t)\| \|\underline{v}_c(t)\|$, 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 $s_{j,i} \in \mathcal{P}_{s_j}^c(t)$.

4.3 Feature Observer Design With Object Return

As discussed in recent work [such as](#) [62] and [63], 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 (4–25) and compensate for the error growth through the development of dwell-time conditions.

4.3.1 Feature Predictor Design

As shown in (4–17), 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

$$\frac{d}{dt} \left(\hat{\underline{u}}_{s_{j,i}/c}(t) \right) \triangleq \mu_{\underline{u}_{s_{j,i}/c}}(t), \quad s_{j,i} \in \mathcal{P}_{s_j}^c(t), \quad (4-26)$$

where $\mu_{\underline{u}_{s_{j,i}/c}}(t) \triangleq -\underline{\omega}_c^\times(t) \hat{\underline{u}}_{s_{j,i}/c}(t) + \frac{1}{\hat{d}_{s_{j,i}/c}(t)} \left(\hat{\underline{u}}_{s_{j,i}/c}(t) \hat{\underline{u}}_{s_{j,i}/c}^T(t) - I_{3 \times 3} \right) \underline{v}_c(t)$. While $s_{j,i} \in \mathcal{P}_{s_j}(t)$, a reset map (cf., [62] and [63]) is used to set $\hat{\underline{u}}_{s_{j,i}/c}(t) \rightarrow \underline{u}_{s_{j,i}/c}(t)$. Also, [since](#) $\underline{u}_{s_{j,i}/c}(t)$ is a unit vector, $\|\hat{\underline{u}}_{s_{j,i}/c}(t)\| = 1$. Let the predictor error for $\underline{u}_{s_{j,i}/c}(t)$ be quantified as

$$\tilde{\underline{u}}_{s_{j,i}/c}(t) \triangleq \underline{u}_{s_{j,i}/c}(t) - \hat{\underline{u}}_{s_{j,i}/c}(t), \quad (4-27)$$

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

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

$$\begin{aligned} \frac{d}{dt} \left(\underline{u}_{k_j/c}(t) \right) &= -\underline{\omega}_c^\times(t) \underline{u}_{k_j/c}(t) \\ &\quad + \frac{1}{d_{k_j/c}(t)} \left(\underline{u}_{k_j/c}(t) \underline{u}_{k_j/c}^T(t) - I_{3 \times 3} \right) \underline{v}_c(t), \end{aligned} \quad (4-28)$$

a predictor similar to (4–26) is designed to estimate $\underline{u}_{k_j/c}(t)$ as

$$\frac{d}{dt} \left(\hat{\underline{u}}_{k_j,i/c}(t) \right) \triangleq \mu_{\underline{u}_{k_j,i/c}}(t), \quad s_{j,i} \in \mathcal{P}_{s_j}^c(t), \quad (4-29)$$

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

$$\tilde{\underline{u}}_{k_j,i/c}(t) \triangleq \underline{u}_{k_j/c}(t) - \hat{\underline{u}}_{k_j,i/c}(t) \quad (4-30)$$

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

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 derivative with respect to time for $q_{k_j/c}(t)$ is

$$\frac{d}{dt} (q_{k_j/c}(t)) = -\frac{1}{2} B(q_{k_j/c}(t)) \underline{\omega}_c(t), \quad (4-31)$$

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 $B^T(q(t)) B(q(t)) = I_{3 \times 3}$.³ The rotation matrix representation of a unit quaternion $q(t)$ is

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

³ Time dependence is suppressed except when needed for clarity or introducing terms.

Similar to (4-31), a predictor is designed for $q_{k_j/c}(t)$ as

$$\frac{d}{dt} (\hat{q}_{k_j/c}(t)) = -\frac{1}{2} B (\hat{q}_{k_j/c}(t)) \underline{\omega}_c(t), \quad (4-32)$$

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 (4-31) and (4-32) are equivalent for all $t > \zeta_j^u$.

Using the estimates from the predictors in (4-26), (4-29), and (4-32), a predictor is designed for $\eta_{s_{j,i}}(t)$ when $s_{j,i} \in \mathcal{P}_{s_j}^c(t)$ as

$$\frac{d}{dt} (\hat{\eta}_{s_{j,i}}(t)) \triangleq \text{proj} \left(\hat{\mu}_{\xi_{s_{j,i}}}(t) \right), \quad s_{j,i} \in \mathcal{P}_{s_j}^c(t), \quad (4-33)$$

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 (4-15), substituting (2-11)-(2-13), (4-27), (4-30), and (4-33), simplifying yields

$$\frac{d}{dt} (\tilde{\eta}_{s_{j,i}}(t)) = - \begin{bmatrix} \tilde{u}_{s_{j,i}/c}^T(t) \underline{v}_c(t) \\ \tilde{u}_{k_j,i/c}^T(t) \underline{v}_c(t) \\ 0 \end{bmatrix}. \quad (4-34)$$

4.3.2 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., $s_{j,i} \in \mathcal{P}_{s_j} \left(t_{s_{j,i}}^{a_n} \right)$) with $\sigma_{s_{j,i}}^o \left(t_{s_{j,i}}^{a_n} \right) = a \wedge \sigma_{s_{j,i}} \left(t_{s_{j,i}}^{a_n} \right) = a$, 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., $s_{j,i} \in \mathcal{P}_{s_j}^c \left(t_{s_{j,i}}^{u_n} \right)$).

Theorem 4.2. *The predictor design in (4–33) 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*

$$\|\tilde{\eta}_{s_{j,i}}(t)\| \leq \|\tilde{\eta}_{s_{j,i}} \left(t_{s_{j,i}}^{u_n} \right)\| + 4\bar{v}_c \left(t - t_{s_{j,i}}^{u_n} \right). \quad (4-35)$$

Proof. Taking the time derivative of (4–18), substituting (4–34), and using the bounds

$\|\tilde{u}_{k_{j,i}/c}(t)\| \leq 2$ and $\|\tilde{u}_{s_{j,i}/c}(t)\| \leq 2$ yields

$$\frac{d}{dt} \left(V_{s_{j,i}} \left(\tilde{\eta}_{s_{j,i}}(t) \right) \right) \leq 4\sqrt{2\bar{v}_c} \sqrt{V_{s_{j,i}} \left(\tilde{\eta}_{s_{j,i}}(t) \right)}. \quad (4-36)$$

Invoking the Comparison Lemma [116, Lemma 3.4] on (4–36) yields the result in (4–35). □

4.3.3 Ensuring Stability Through Dwell-Time Conditions

To facilitate the subsequent development, let $\mathcal{L}_{s_{j,i}}^n \triangleq \left\{ l \in \mathbb{Z}_{>0} : \zeta_{s_{j,i}}^{a_l} \in \left[t_{s_{j,i}}^{a_n}, t_{s_{j,i}}^{u_n} \right] \cap t > \tau_{s_{j,i}} \right\}$, 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 $\bar{l}_n \triangleq \bar{l}_n - \underline{l}_n$. Additionally, let $l_n \in \{1, 2, \dots, \bar{l}_n + 1\}$ represent the index of $\mathcal{L}_{s_{j,i}}^n$, and let $\zeta_{s_{j,i}}^{a_{l_n,n}} \in \left[t_{s_{j,i}}^{a_n}, t_{s_{j,i}}^{u_n} \right] \cap t > \tau_{s_{j,i}}$ and $\zeta_{s_{j,i}}^{u_{l_n,n}} \in \left[t_{s_{j,i}}^{a_n}, t_{s_{j,i}}^{u_n} \right] \cap t > \tau_{s_{j,i}}$ 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 $\Delta \zeta_{s_{j,i}}^{u_n} \triangleq t_{s_{j,i}}^{u_n} - \zeta_{s_{j,i}}^{u_{\bar{l}_n,n}}$.

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 [62] and [63] will not be possible for switched systems like those defined by $\sigma_{s_{j,i}}(t)$ and $\sigma_{s_{j,i}}^o(t)$ given

$\tilde{\eta}_{s_j,i} \left(t_{s_j,i}^{a_n} \right)$ and $\tilde{\eta}_{s_j,i} \left(t_{s_j,i}^{u_n} \right)$ are unknown and cannot be reset. However, $\tilde{d}_{s_j,i/c} \left(t_{s_j,i}^{a_1} \right) \leq \bar{d}$, $\tilde{d}_{k_j,i/c} \left(t_{s_j,i}^{a_1} \right) \leq \bar{d}$, and $\tilde{d}_{s_j,i/k_j} \left(t_{s_j,i}^{a_1} \right) \leq \bar{d}$ implying $\|\tilde{\eta}_{s_j,i} \left(t_{s_j,i}^{a_1} \right)\| \leq \sqrt{3}\bar{d}$.

Let $\underline{\tilde{\eta}}_{s_j,i} \in (0, \bar{d})$ be a user-defined threshold such that $\|\tilde{\eta}_{s_j,i} \left(\zeta_{s_j,i}^{u_{\bar{t}_n,n}} \right)\| \leq \underline{\tilde{\eta}}_{s_j,i}$ (i.e., the user will need the error below some threshold before allowing the object to leave the FOV). Furthermore, let $\overline{\tilde{\eta}}_{s_j,i} \in \left(\underline{\tilde{\eta}}_{s_j,i}, \bar{d} \right]$ be a user-defined threshold such that $\underline{\tilde{\eta}}_{s_j,i} < \|\tilde{\eta}_{s_j,i} \left(\zeta_{s_j,i}^{a_{1,n}} \right)\| \leq \overline{\tilde{\eta}}_{s_j,i}$ 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 4.2. 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 2.1 and 2.2, 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 $p_{s_j} (t) \geq 4$. This implies that it is not sufficient to only consider the dwell-times for individual features given the observers are used on the assumption $\underline{u}_{k_j/c} (t)$ 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 $\underline{u}_{k_j/c} (t)$ and $R_{k_j/c} (t)$; 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, $\sigma_{s_j}^o (t) = a \wedge \bigcap_{s_j,i \in \mathcal{A}_{s_j} (t)} \left(\sigma_{s_j,i}^o (t) = a \wedge \sigma_{s_j,i} (t) = a \right)$ 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 $\overline{t}_{s_j}^{a_n} \triangleq \max \{ \zeta_{s_j,i}^{a_{1,n}} \}$, $\underline{t}_{s_j}^{u_n} \triangleq \min \{ \zeta_{s_j,i}^{u_{\bar{t}_n,n}} \}$, and $\underline{\Delta} t_{s_j}^{a_n} \triangleq \underline{t}_{s_j}^{u_n} - \overline{t}_{s_j}^{a_n}$ for the n th

switching cycle across all features in $\mathcal{A}_{s_j}(t)$. Furthermore, let $\overline{t_{s_j}^{a_{n+1}}} \triangleq \max \{ \zeta_{s_j,i}^{a_{1,n+1}} \}$ and $\underline{\Delta t_{s_j}^{u_n}} \triangleq \overline{t_{s_j}^{a_{n+1}}} - \underline{t_{s_j}^{u_n}}$.

Theorem 4.3. 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 (4–16) ensure the estimation error in $\tilde{\eta}_{s_j,i}(t)$ at time $t = t_{s_j,i}^{u_1}$ is GUUB as $\|\tilde{\eta}_{s_j,i}(t_{s_j,i}^{u_1})\| \leq \underline{\tilde{\eta}_{s_j,i}}$ provided the switching signals satisfy the initial minimum feedback availability dwell-time condition

$$\underline{\Delta t_{s_j}^{a_1}} \geq -\frac{1}{\beta} \ln \left(\frac{\tilde{\eta}_{s_j,i}}{\sqrt{3}\bar{d}} \right) > -\frac{1}{\beta} \ln \left(\frac{1}{\sqrt{3}} \right). \quad (4-37)$$

Proof. Using (4–25) 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 $\|\tilde{\eta}_{s_j,i}(t_{s_j,i}^{u_1})\| \leq \underline{\tilde{\eta}_{s_j,i}}$ 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 $\underline{\tilde{\eta}_{s_j,i}} < \bar{d}$, $\frac{\tilde{\eta}_{s_j,i}}{\sqrt{3}\bar{d}} < \frac{1}{\sqrt{3}}$ yielding the bound in (4–37). \square

Theorem 4.4. 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 (4–16) ensure the estimation error in $\tilde{\eta}_{s_j,i}(t)$ is GUUB as $\|\tilde{\eta}_{s_j,i}(t_{s_j,i}^{u_n})\| \leq \underline{\tilde{\eta}_{s_j,i}}$ provided the switching signals satisfy the minimum feedback availability dwell-time condition

$$\underline{\Delta t_{s_j}^{a_n}} \geq -\frac{1}{\beta} \ln \left(\frac{\tilde{\eta}_{s_j,i}}{\underline{\tilde{\eta}_{s_j,i}}} \right) > 0, n > 1 \quad (4-38)$$

Proof. The proof follows Theorem 4.3 using the upper $\|\tilde{\eta}_{s_j,i}(t_{s_j,i}^{a_n})\| \leq \overline{\tilde{\eta}_{s_j,i}}$. \square

Theorem 4.5. 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 (4–33) ensure the estimation error in $\tilde{\eta}_{s_j,i}(t)$ is GUUB as

$\|\tilde{\eta}_{s_j,i} \left(t_{s_j,i}^{a_{n+1}} \right)\| \leq \overline{\tilde{\eta}_{s_j,i}}$ provided the switching signals satisfy the maximum loss of feedback dwell-time condition

$$0 < \overline{\Delta t_{s_j}^{u_n}} \leq \frac{\overline{\tilde{\eta}_{s_j,i}} - \underline{\tilde{\eta}_{s_j,i}}}{4\underline{v_c}} \quad (4-39)$$

Proof. Using (4-35) for the n th instance implies $\|\tilde{\eta}_{s_j,i} \left(t_{s_j,i}^{a_{n+1}} \right)\| \leq \|\tilde{\eta}_{s_j,i} \left(t_{s_j,i}^{u_n} \right)\| + 4\underline{v_c} \overline{\Delta t_{s_j}^{u_n}}$. It is desired to have $\|\tilde{\eta}_{s_j,i} \left(t_{s_j,i}^{a_{n+1}} \right)\| \leq \overline{\tilde{\eta}_{s_j,i}}$ and $\|\tilde{\eta}_{s_j,i} \left(t_{s_j,i}^{u_n} \right)\| \leq \underline{\tilde{\eta}_{s_j,i}}$. Substituting the second bounds into the first and solving for $\overline{\Delta t_{s_j}^{u_n}}$ yields $\overline{\Delta t_{s_j}^{u_n}} \leq \frac{\overline{\tilde{\eta}_{s_j,i}} - \underline{\tilde{\eta}_{s_j,i}}}{4\underline{v_c}}$. Because $\overline{\Delta t_{s_j}^{u_n}}$ must upper bound the dwell-times to ensure all of the feature observers are implementable, $\overline{\Delta t_{s_j}^{u_n}} \leq \frac{\overline{\tilde{\eta}_{s_j,i}} - \underline{\tilde{\eta}_{s_j,i}}}{4\underline{v_c}}$. Given $\overline{\tilde{\eta}_{s_j,i}} > \underline{\tilde{\eta}_{s_j,i}}$, $\overline{\tilde{\eta}_{s_j,i}} - \underline{\tilde{\eta}_{s_j,i}} > 0$, yielding the bound in (4-39). \square

Ensuring that (4-37)-(4-39) are satisfied guarantees the error remains bounded as $\underline{\tilde{\eta}_{s_j,i}} < \|\tilde{\eta}_{s_j,i}(t)\| \leq \sqrt{3} \bar{d}$ for all time $t < t_{s_j,i}^{u_1}$, $\underline{\tilde{\eta}_{s_j,i}} < \|\tilde{\eta}_{s_j,i}(t)\| \leq \overline{\tilde{\eta}_{s_j,i}}$ 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.

4.4 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 2-3, 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., [62] and [63]). Let $p_{c/k_1}(t) \in \mathbb{R}^3$ 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 4.3. 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 4.3, 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 $\tau_{s_j}^{a n} \triangleq \min \left\{ t > t_{s_{j,i}}^{a n} : \sigma_{s_j}^{\Delta t} = a \right\}$, 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., $\|\tilde{\eta}_{s_{j,i}}(t)\| \leq \underline{\tilde{\eta}}_{s_{j,i}}$), implying $\tilde{d}_{k_j,i/c}(t) \leq \underline{\tilde{\eta}}_{s_{j,i}}$.

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

$$\tilde{\underline{p}}_{c/k_1}(t) \triangleq \underline{p}_{c/k_1}(t) - \hat{\underline{p}}_{c/k_1}(t), \quad (4-40)$$

$$\tilde{q}_{c/k_1}(t) \triangleq q_{c/k_1}(t) - \hat{q}_{c/k_1}(t), \quad (4-41)$$

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 $p_{c/k_1}(t)$ and $q_{c/k_1}(t)$, respectively.

Taking the **time** derivative of $p_{c/k_1}(t)$ yields

$$\frac{d}{dt} \left(p_{c/k_1}(t) \right) = R(q_{c/k_1}(t)) \underline{v}_c(t). \quad (4-42)$$

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

$$\frac{d}{dt} (q_{c/k_1}(t)) = \frac{1}{2} B(q_{c/k_1}(t)) \underline{\omega}_c(t). \quad (4-43)$$

Taking the **time** derivative of $\tilde{p}_{c/k_1}(t)$ and $\tilde{q}_{c/k_1}(t)$, and substituting (4-42) and (4-43) yields

$$\frac{d}{dt} \left(\tilde{p}_{c/k_1}(t) \right) = R(q_{c/k_1}(t)) \underline{v}_c(t) - \frac{d}{dt} \left(\hat{p}_{c/k_1}(t) \right) \quad (4-44)$$

and

$$\frac{d}{dt} (\tilde{q}_{c/k_1}(t)) = \frac{1}{2} B(q_{c/k_1}(t)) \underline{\omega}_c(t) - \frac{d}{dt} (\hat{q}_{c/k_1}(t)). \quad (4-45)$$

When the landmark is in the FOV, $\sigma_L(t) = a$ and feedback of the pose of the camera is directly available under Assumption 4.3. While $\sigma_L(t) = a$, a reset map is used to reset both position and orientation as

$$\hat{p}_{c/k_1}(t) \rightarrow p_{c/k_1}(t) \quad (4-46)$$

and

$$\hat{q}_{c/k_1}(t) \rightarrow q_{c/k_1}(t). \quad (4-47)$$

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, $\left\{ \sigma_{s_j}^{\Delta t}(t) \right\}_{j=1}^{\overline{p}_s(t)}$.

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

$$\hat{\underline{p}}_{c/k_j}^* (t) \triangleq \underline{u}_{c/k_j} (t) \hat{d}_{k_j/c} (t), \quad (4-48)$$

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 $\hat{\underline{p}}_{c/k_j}^* (t)$ is expressed in the j th key frame based on (4-48). Using (4-48), 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

$$\begin{aligned} \frac{d}{dt} \left(\hat{\underline{p}}_{c/k_1} (t) \right) &= R(\hat{q}_{c/k_1}(t)) \underline{v}_c (t) \\ &+ K_p \left(\hat{\underline{p}}_{k_j/k_1} + R(\hat{q}_{k_j/k_1}) \hat{\underline{p}}_{c/k_j}^* (t) - \hat{\underline{p}}_{c/k_1} (t) \right), \end{aligned} \quad (4-49)$$

and

$$\begin{aligned} \frac{d}{dt} \left(\hat{q}_{c/k_1} (t) \right) &= \frac{1}{2} B \left(\hat{q}_{c/k_1} (t) \right) \underline{\omega}_c (t) \\ &+ K_q \left(Q \left(\hat{q}_{k_j/k_1} \right) q_{c/k_j} (t) - \hat{q}_{c/k_1} (t) \right), \end{aligned} \quad (4-50)$$

where $\hat{q}_{k_j/k_1} = \hat{q}_{c/k_1} (\zeta_1^a)$, $\hat{\underline{p}}_{k_j/k_1} = \hat{\underline{p}}_{c/k_1} (\zeta_1^a)$, $K_p \in \mathbb{R}^{3 \times 3}$ 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 $\sigma_L (t) = u \wedge \left(\sigma_{s_j}^{\Delta t} (t) = u \vee \sigma_{s_j}^o (t) = u \right)$, a predictor is designed for the pose as

$$\frac{d}{dt} \left(\hat{\underline{p}}_{c/k_1} (t) \right) = R(\hat{q}_{c/k_1}(t)) \underline{v}_c (t) \quad (4-51)$$

and

$$\frac{d}{dt}(\hat{q}_{c/k_1}(t)) = \frac{1}{2}B(\hat{q}_{c/k_1}(t))\underline{\omega}_c(t). \quad (4-52)$$

Only an analysis of the position estimator design is considered given 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 $\hat{q}_{c/k_1}(\zeta_1^a) = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^T$ implies (4-43) is equivalent to (4-50) and (4-52) 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 (4-49) into the derivative of the position error in (4-44), using $\hat{q}_{c/k_1}(t) = q_{c/k_1}(t)$, and simplifying yields

$$\begin{aligned} \frac{d}{dt}(\tilde{\underline{p}}_{c/k_1}(t)) &= -K_p \tilde{\underline{p}}_{c/k_1}(t) + K_p \tilde{\underline{p}}_{k_j/k_1} \\ &\quad + K_p R(q_{k_j/k_1}) \underline{u}_{c/k_j}(t) \frac{1}{p_{s_j}(t)} \sum_{i=1}^{p_{s_j}(t)} \tilde{d}_{k_j,i/c}(t). \end{aligned} \quad (4-53)$$

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 (4-51) into the derivative of the position error in (4-44), using $\hat{q}_{c/k_1}(t) = q_{c/k_1}(t)$, and simplifying yields

$$\frac{d}{dt}(\tilde{\underline{p}}_{c/k_1}(t)) = 0_{3 \times 1}. \quad (4-54)$$

4.4.1 Stability of Key Frame Position Observer and Predictor Design

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

$$V_{c/k_1}(\tilde{\underline{p}}_{c/k_1}(t)) \triangleq \frac{1}{2} \tilde{\underline{p}}_{c/k_1}^T(t) \tilde{\underline{p}}_{c/k_1}(t), \quad (4-55)$$

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

Theorem 4.6. *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 (4–49) ensure the position error of the camera $\tilde{\underline{p}}_{c/k_1}(t)$ is GUUB while $t \in [\tau_{s_j}^{an}, t_{s_j}^{un}]$ in the sense*

$$\|\tilde{\underline{p}}_{c/k_1}(t)\|^2 \leq \|\tilde{\underline{p}}_{c/k_1}(\tau_{s_j}^{an})\|^2 \exp\left(-\beta_p^a(t - \tau_{s_j}^{an})\right) + 2\frac{\varepsilon_{k_j}}{\beta_p^a} \quad (4-56)$$

where $\beta_p^a \triangleq \lambda_{\min}\{K_p\}$, $\varepsilon_{k_j} \triangleq \frac{(\lambda_{\max}\{K_p\}(\|\tilde{\underline{p}}_{k_j/k_1}\| + \tilde{\eta}_{s_j,i}))^2}{2\lambda_{\min}\{K_p\}}$, and $\lambda_{\max}\{\cdot\}$ is the maximum eigenvalue of $\{\cdot\}$.

Proof. Taking the derivative of (4–55) with respect to time and substituting (4–53) yields

$$\frac{d}{dt}\left(V_{c/k_1}\left(\tilde{\underline{p}}_{c/k_1}(t)\right)\right) \leq -\beta_p^a V_{c/k_1}\left(\tilde{\underline{p}}_{c/k_1}(t)\right) + \varepsilon_{k_j}. \quad (4-57)$$

Invoking the Comparison Lemma [116, Lemma 3.4] on (4–57) then upper bounding yields (4–56). \square

Theorem 4.7. *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 (4–51) ensure the position error of the camera $\tilde{\underline{p}}_{c/k_1}(t)$ is bounded while $t \notin [\tau_{s_j}^{an}, t_{s_j}^{un}]$ in the sense*

$$\|\tilde{\underline{p}}_{c/k_1}(t)\| \leq \|\tilde{\underline{p}}_{c/k_1}(t_{s_j}^{un})\|. \quad (4-58)$$

Proof. Taking the derivative of (4–55) with respect to time and substituting (4–54) yields

$$\frac{d}{dt}\left(V_{c/k_1}\left(\tilde{\underline{p}}_{c/k_1}(t)\right)\right) \leq 0. \quad (4-59)$$

Invoking the Comparison Lemma [116, Lemma 3.4] on the result then upper bounding yields (4–58). \square

4.5 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., [62] and [63]). The experiment assumed no return to previous objects (i.e.,



Figure 4-1. Photo courtesy of author. Image of the Kobuki Turtlebot and iDS uEye camera used for experiments.



Figure 4-2. Photo courtesy of author. 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.

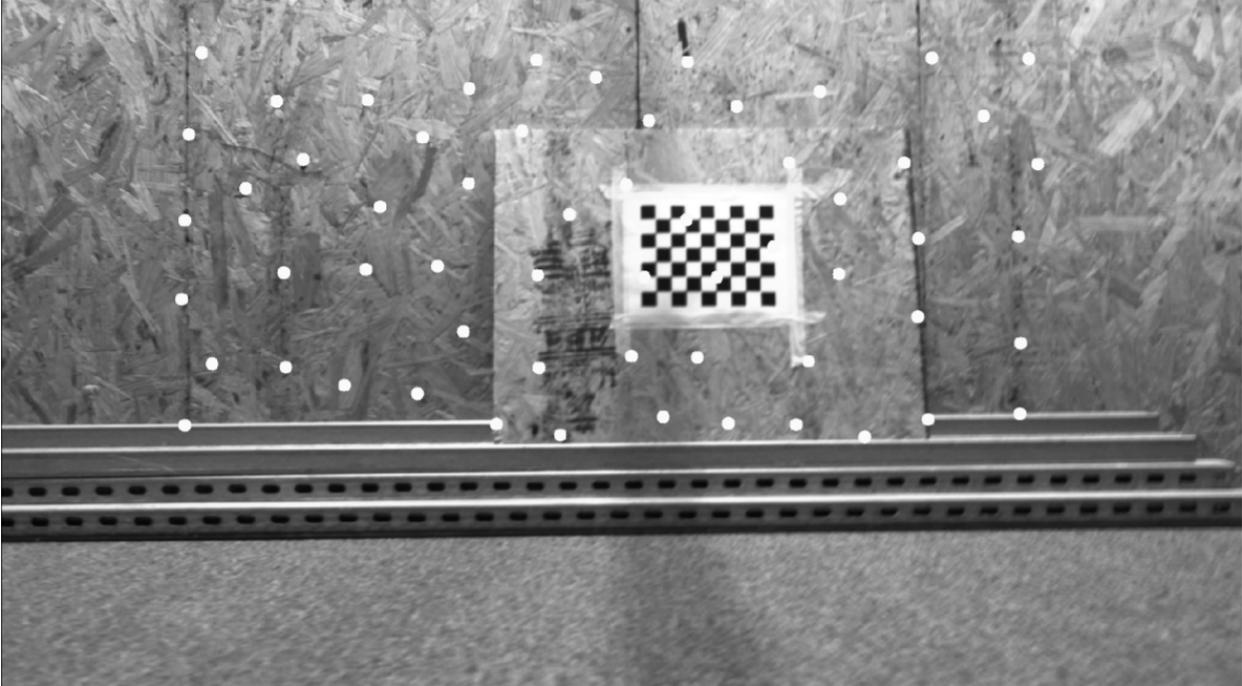


Figure 4-3. Photo courtesy of author. Image of the landmark captured within a key frame (i.e., the checkerboard is leaning on the wooden wall), where the white dots drawn in the image are the extracted corner features.

each new set of features was considered a new object) implying the initial minimum dwell-time condition in (4-37) must always be satisfied before using a new object in the position observer. The performance examined is that of the developed feature observer in (4-16) and the pose estimation strategy using the reset maps in (4-46) and (4-47) when a landmark is in the FOV (i.e., $\sigma_L(t) = a$), the observer update laws in (4-49) and (4-50) 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 (4-51) and (4-52) 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 4-1, provided images and velocity at 30 Hz as it drove through the environment (e.g., Figure 4-2). The estimator can run real-time (i.e., at 30 Hz) and was implemented

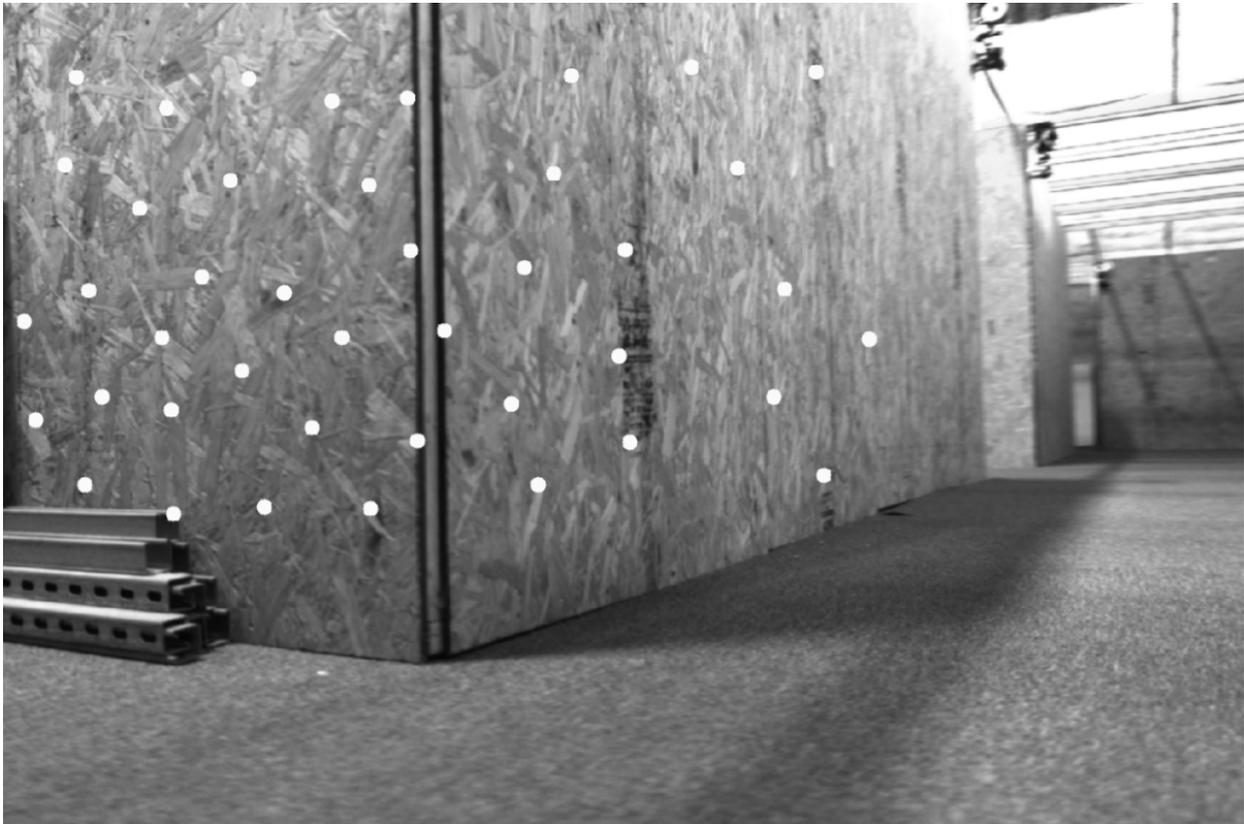


Figure 4-4. Photo courtesy of author. Image of the features extracted from a key frame image of a nonplanar object (i.e., two wooden walls with a 90° angle between them, where the white dots are the extracted corner features).

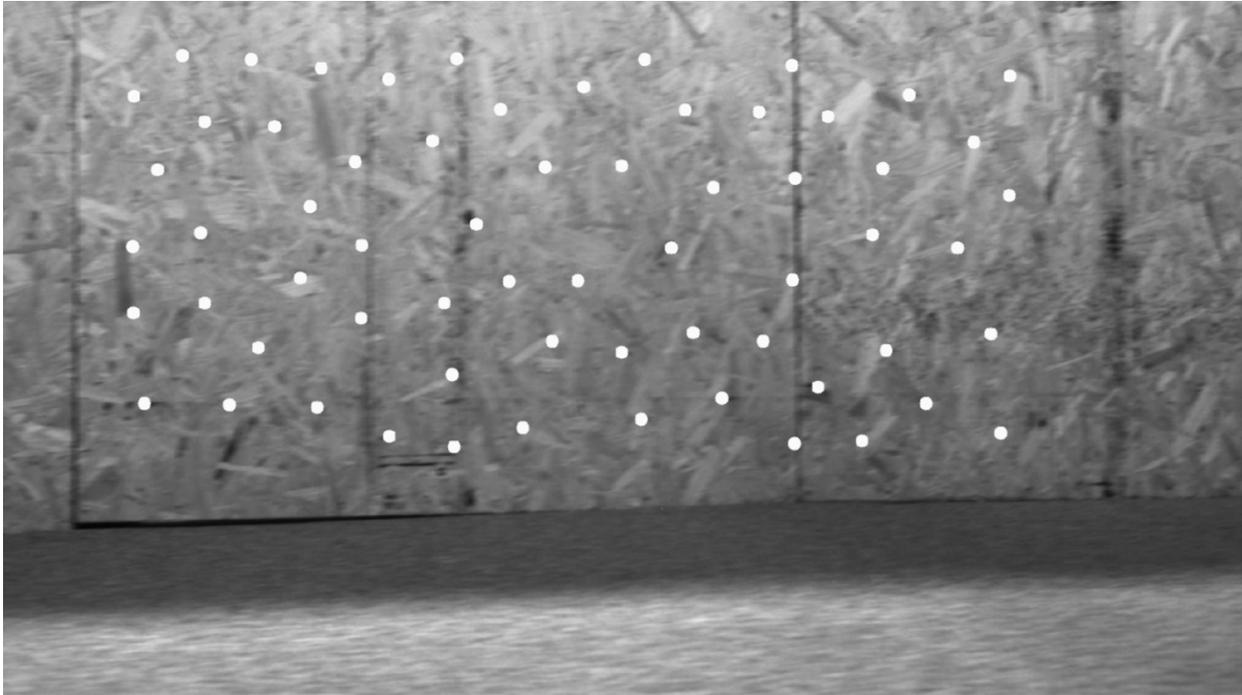


Figure 4-5. Photo courtesy of author. Image of the features extracted from a key frame image of a planar object, where the white dots are the extracted corner features.

using Eigen3, OpenCV, and ROS c++ libraries (cf., [117], [118], and [119], respectively). The landmark was a checkerboard with 8×6 corners where each square is 25.4 millimeters \times 25.4 millimeters. As the example in Figure 4-2 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., [111] or [118]) were extracted from each key frame image with an initial spacing of 100 pixels.⁴ A new object

⁴ The extraction method could only find between 30 and 50 corner features in a key frame image depending on what was in the image.

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 $\sigma_{s_j}^o(t) = u$).⁵ The environment consisted of both nonplanar (e.g., Figure 4-4) and planar (e.g., Figure 4-5) 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 4-3), the ground truth was provided to the vehicle and the reset maps in (4-46) and (4-47) were 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 (4-7) from the previous image at time $t_p \in \mathbb{R}_{>0}$ 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

$$\begin{aligned} \hat{\underline{u}}_{s_{j,i}/c}(t) = & \underline{u}_{s_{j,i}/c}(t_p) - \int_{t_p}^t \underline{\omega}_c^\times(\iota) \underline{u}_{s_{j,i}/c}(\iota) d\iota \\ & + \int_{t_p}^t \frac{1}{\hat{d}_{s_{j,i}/c}(\iota)} \underline{u}_{s_{j,i}/c}(\iota) \underline{u}_{s_{j,i}/c}^T(\iota) \underline{v}_c(\iota) d\iota \\ & - \int_{t_p}^t \frac{1}{\hat{d}_{s_{j,i}/c}(\iota)} \underline{v}_c(\iota) d\iota. \end{aligned} \quad (4-60)$$

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](#)

⁵ While an absolute minimum of 4 features is required, 4 corner features will typically provide a poor estimate of $q_{k_j/c}(t)$ and $\underline{u}_{k_j/c}(t)$ and it was experimentally determined that 20 corner features was the lowest number of features that could consistently provide good estimates.

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., [118] or [15]). 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 (4–8) and subsequently update (4–16), 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, [118]) 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 (4–16) 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 (4–2) 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 thresholds 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 (4-4) where $N = 50$. Using $\lambda_\tau = 0.0001$, when $\lambda_{\min} \left\{ \Sigma_{\mathcal{Y}_{s_{j,i}}} \right\} > \lambda_\tau$, the value of $\mathcal{X}_{s_{j,i}}$ in (4-5) was determined and used to update (4-16), where $K_{\mathcal{X}} = 30I_{3 \times 3}$ was determined to consistently perform well with **low noise sensitivity**. Additionally, after $\lambda_{\min} \left\{ \Sigma_{\mathcal{Y}_{s_{j,i}}} \right\} > \lambda_\tau$, 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 (4-37), 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 $\lambda_{\min} \{K_{\mathcal{X}}\} = 30$, $\lambda_a = 0.3$, and $\bar{d} = 5.0$ meters, $\underline{\Delta t}_{s_j}^{a_1} \geq 0.75$ 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 (4-49) and (4-50) were 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 (4-51) and (4-52) 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 4-1 over a path of approximately 250 meters through

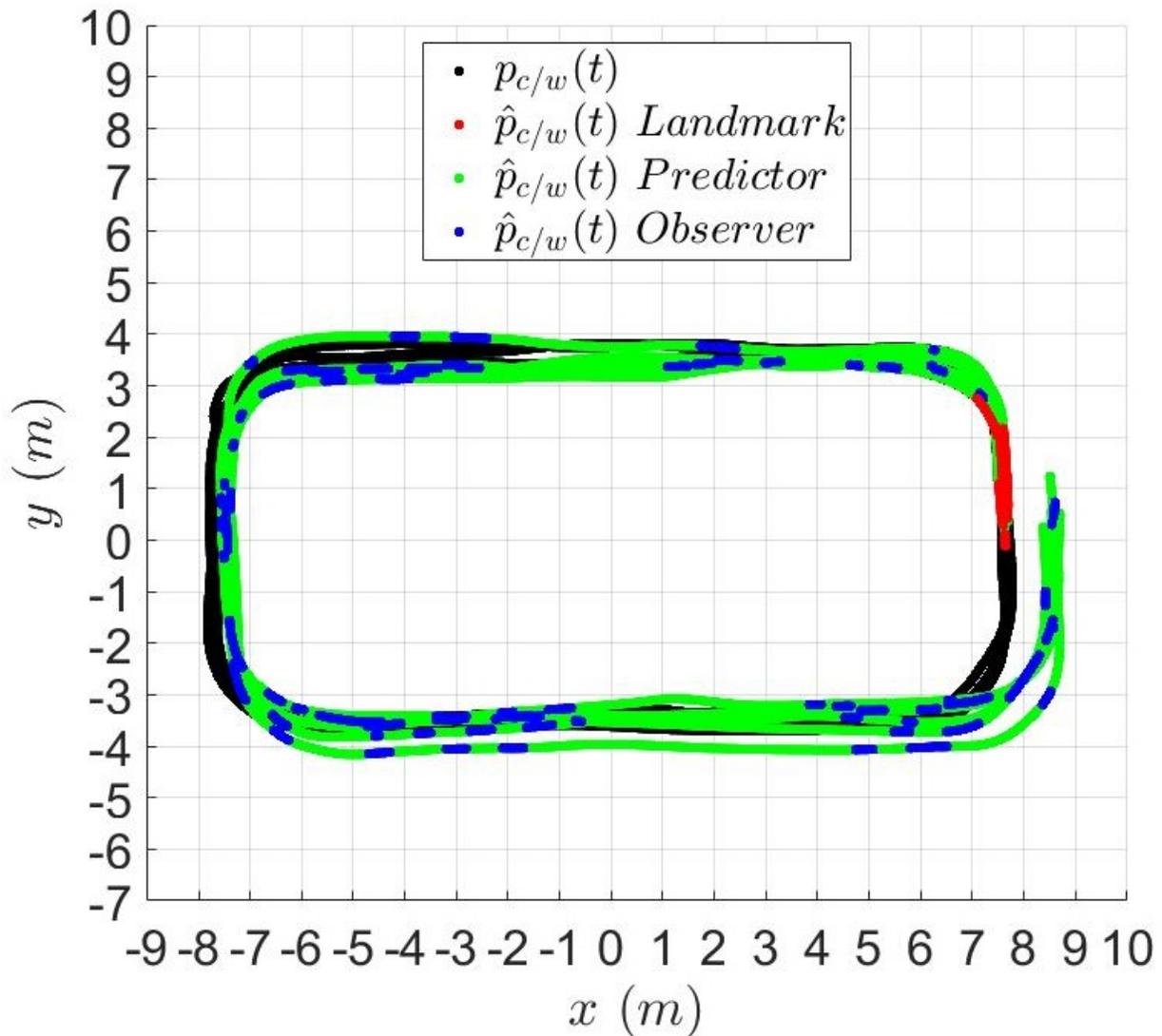


Figure 4-6. Plot of the path of the camera during the experiment and the estimated path of the camera using a standard deviation of 3 pixels for the history stack rejection algorithm. The true path is marked in black. The estimated path shows which estimator is activated over the experiment. While the 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 are active, respectively.

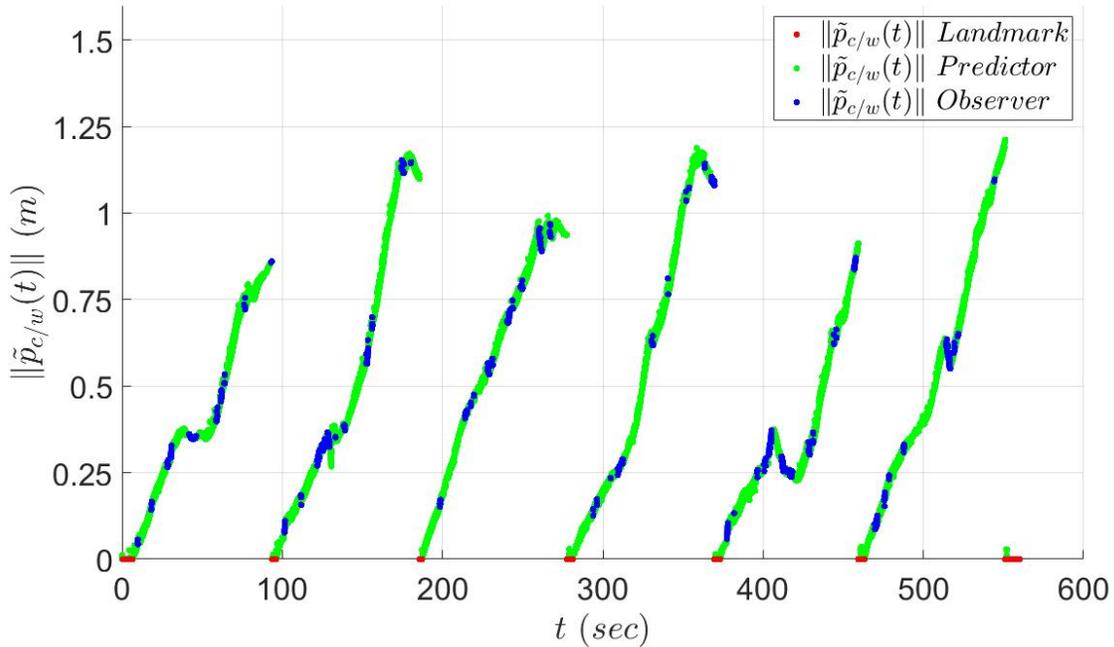


Figure 4-7. Plot of the norm of the camera position error during the experiment using a standard deviation of 3 pixels for accepting data onto the history stack. The camera position error shows which estimator is activated over the experiment. While the landmark is in the FOV, the estimated path is shown using the red marker. Similarly, the estimated path is marked in green or blue when the predictor or observer are 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.

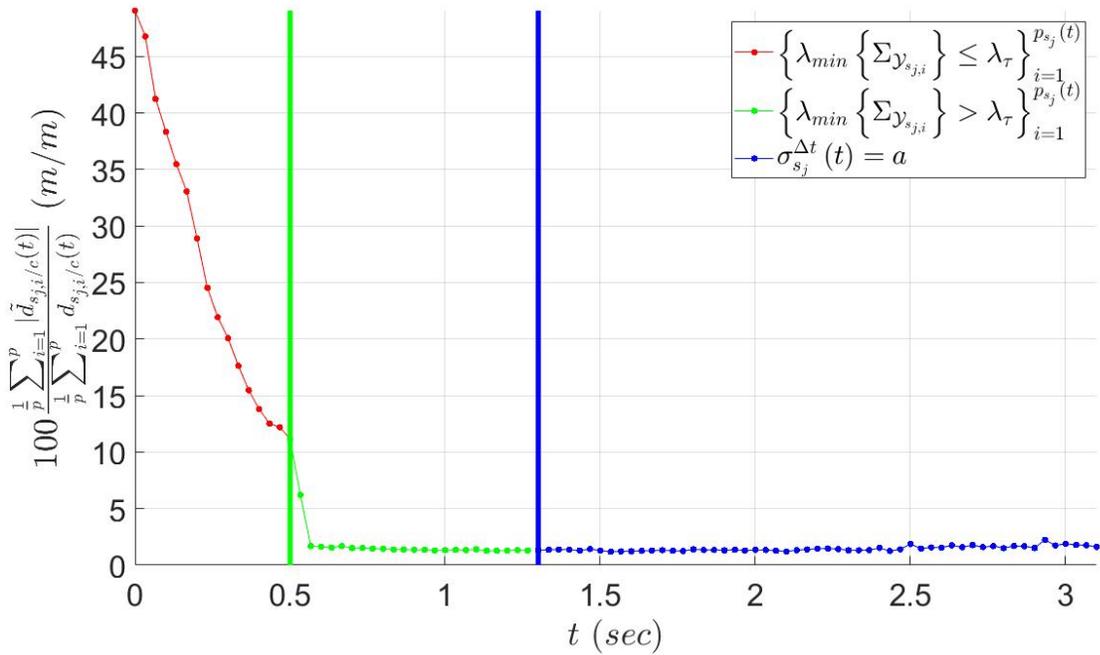


Figure 4-8. Plot of the distance estimator convergence for key frame 33 by showing the average percentage error of the distance to the features relative to the true distance to the features using a standard deviation of 3 pixels for accepting data onto the history stack. The plot shows while the set of estimators have not satisfied the eigenvalue condition, marked in red, the flow type term 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%.

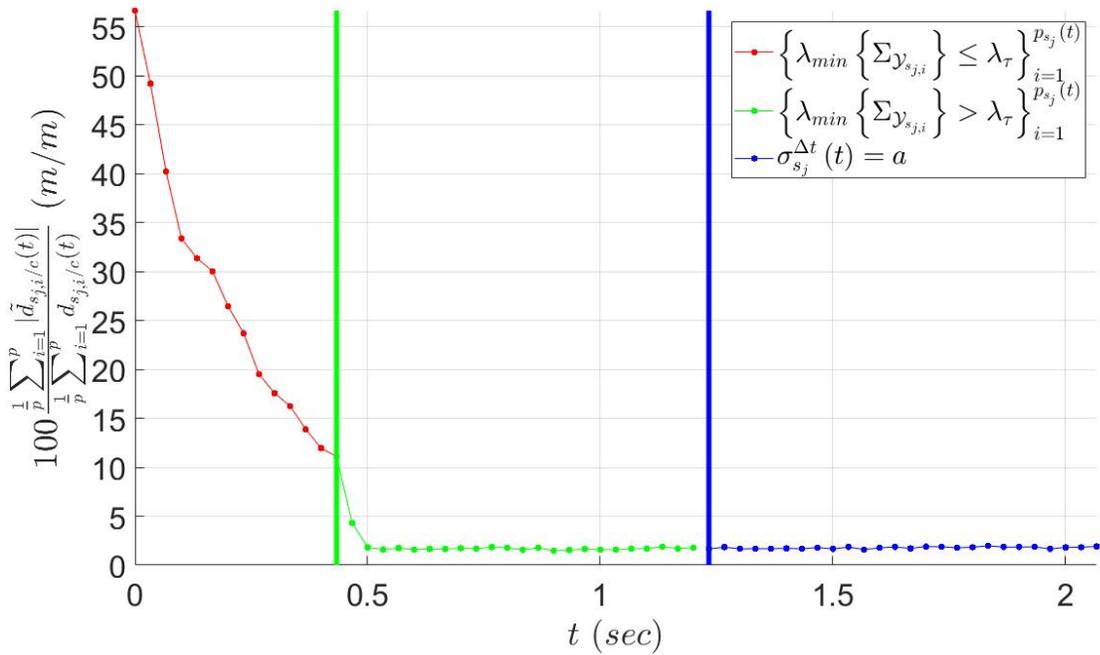


Figure 4-9. Plot of the distance estimator convergence for key frame 173 by showing the average percentage error of the distance to the features relative to the true distance to the features using a standard deviation of 3 pixels for accepting data onto the history stack. The plot shows while the set of estimators have not satisfied the eigenvalue condition, marked in red, the flow type term 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%.

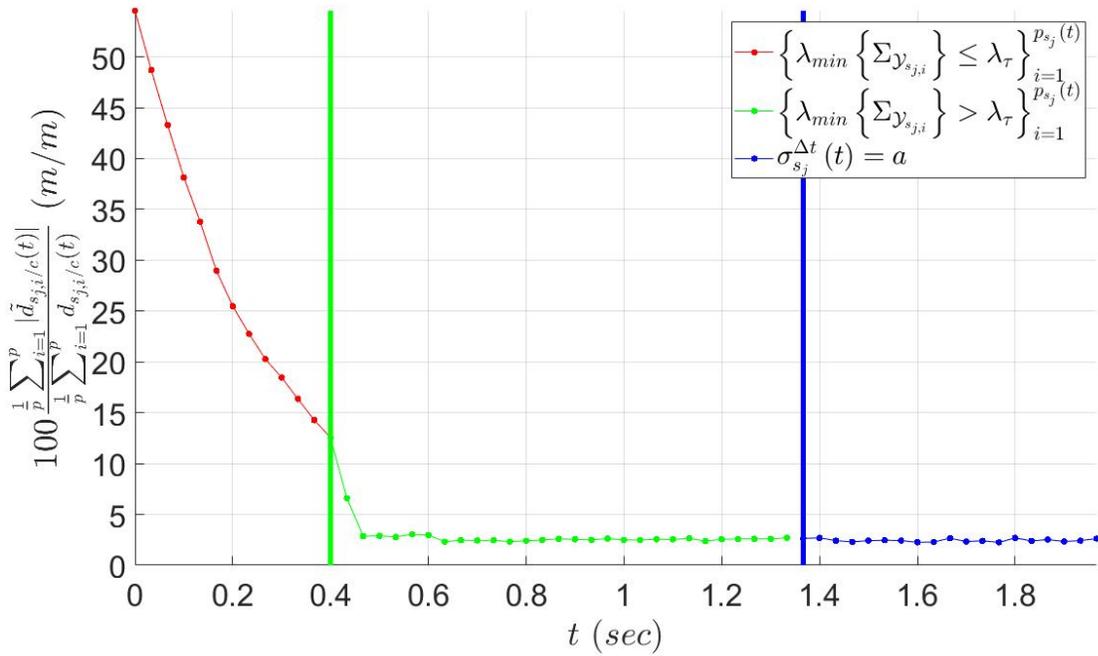


Figure 4-10. Plot of the distance estimator convergence for key frame 215 by showing the average percentage error of the distance to the features relative to the true distance to the features using a standard deviation of 3 pixels for accepting data onto the history stack. The plot shows while the set of estimators have not satisfied the eigenvalue condition, marked in red, the flow type term 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%.

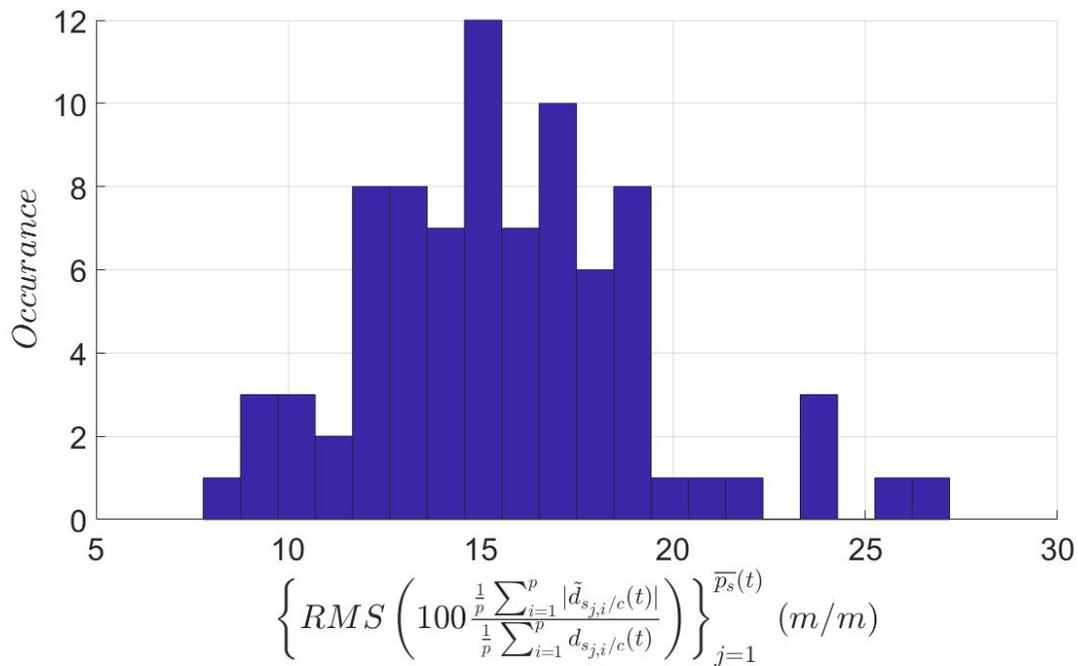


Figure 4-11. Histogram plot of the RMS of the average percentage error of the distance to the features relative to the true distance to the features across the entire experiment using a standard deviation of 3 pixels for accepting data onto the history stack. The histogram shows the the RMS errors over 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.

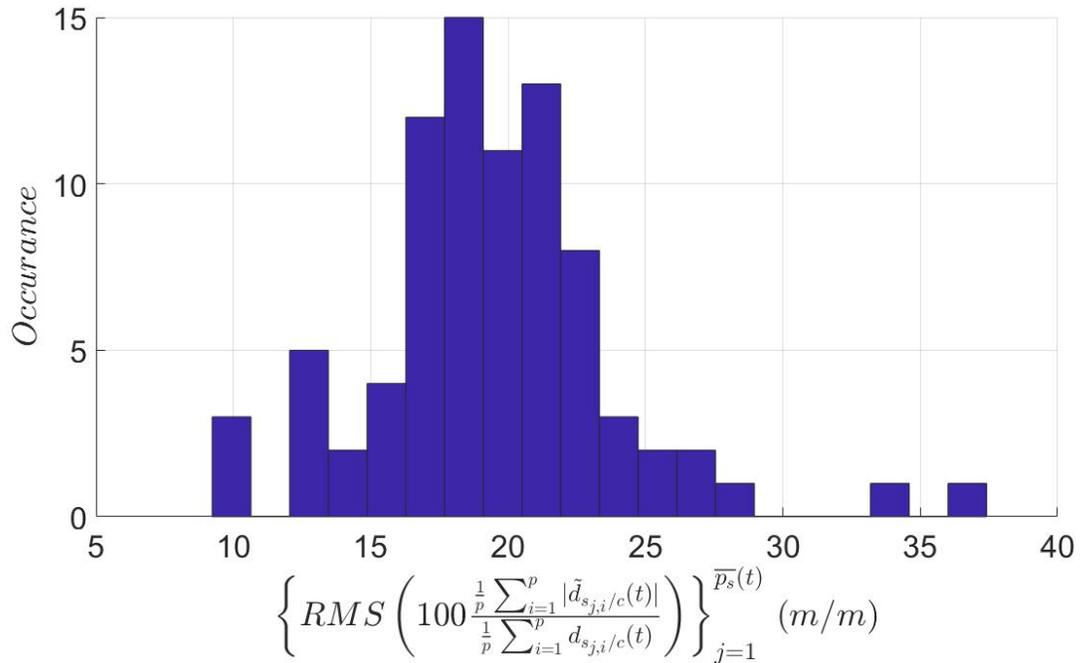


Figure 4-12. Histogram plot of the RMS of the average percentage error of the distance to the features relative to the true distance to the features across before the minimum dwell-time condition is satisfied using a standard deviation of 3 pixels for accepting data onto the history stack. The histogram 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.

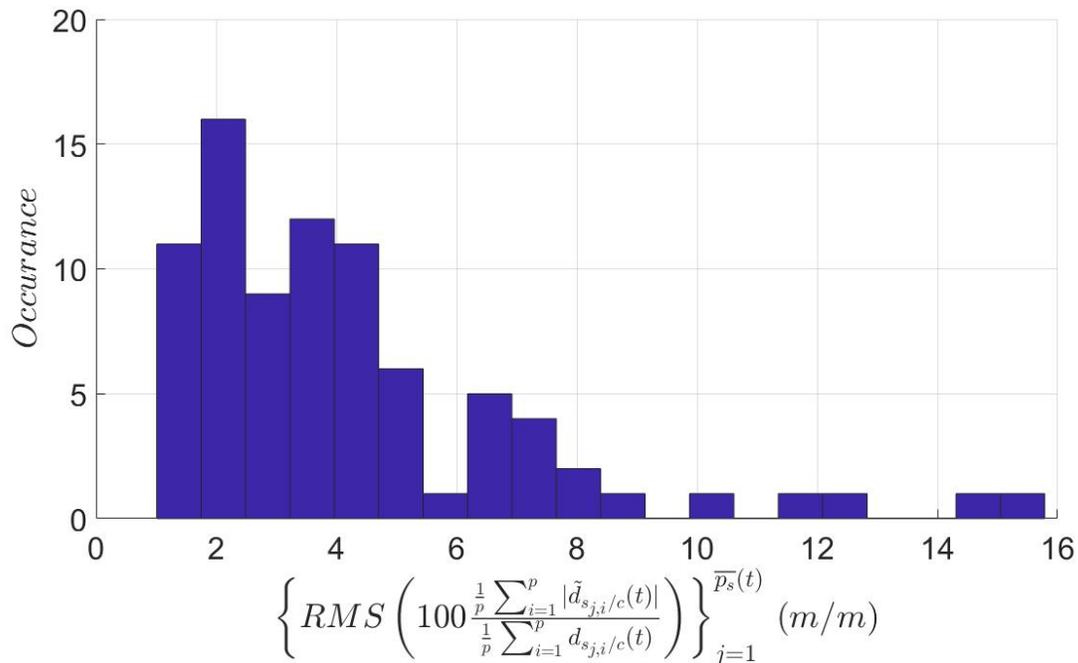


Figure 4-13. Histogram plot of the RMS of the average percentage error of the distance to the features relative to the true distance to the features across after the minimum dwell-time condition is satisfied using a standard deviation of 3 pixels for accepting data onto the history stack. The histogram 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.

the series of wooden hallways shown in Figure 4-2. The true and estimated path over the experiment is shown in Figure 4-6 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 4-7 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 4-7, 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 4-8-4-10, and the histograms of the RMS error (i.e., $\left\{ \text{RMS} \left(100 \frac{\frac{1}{p} \sum_{i=1}^p |\tilde{d}_{s_{j,i}/c}(t)|}{\frac{1}{p} \sum_{i=1}^p d_{s_{j,i}/c}(t)} \right) \right\}_{j=1}^{\bar{p}_s(t)}$) in Figures 4-11-4-13, 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 4-12; 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 [62] and [63], the developed position estimator strategy is to ensure the error in the position does not exceed a desired threshold through the development of dwell-time conditions. However, [62] and [63] 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

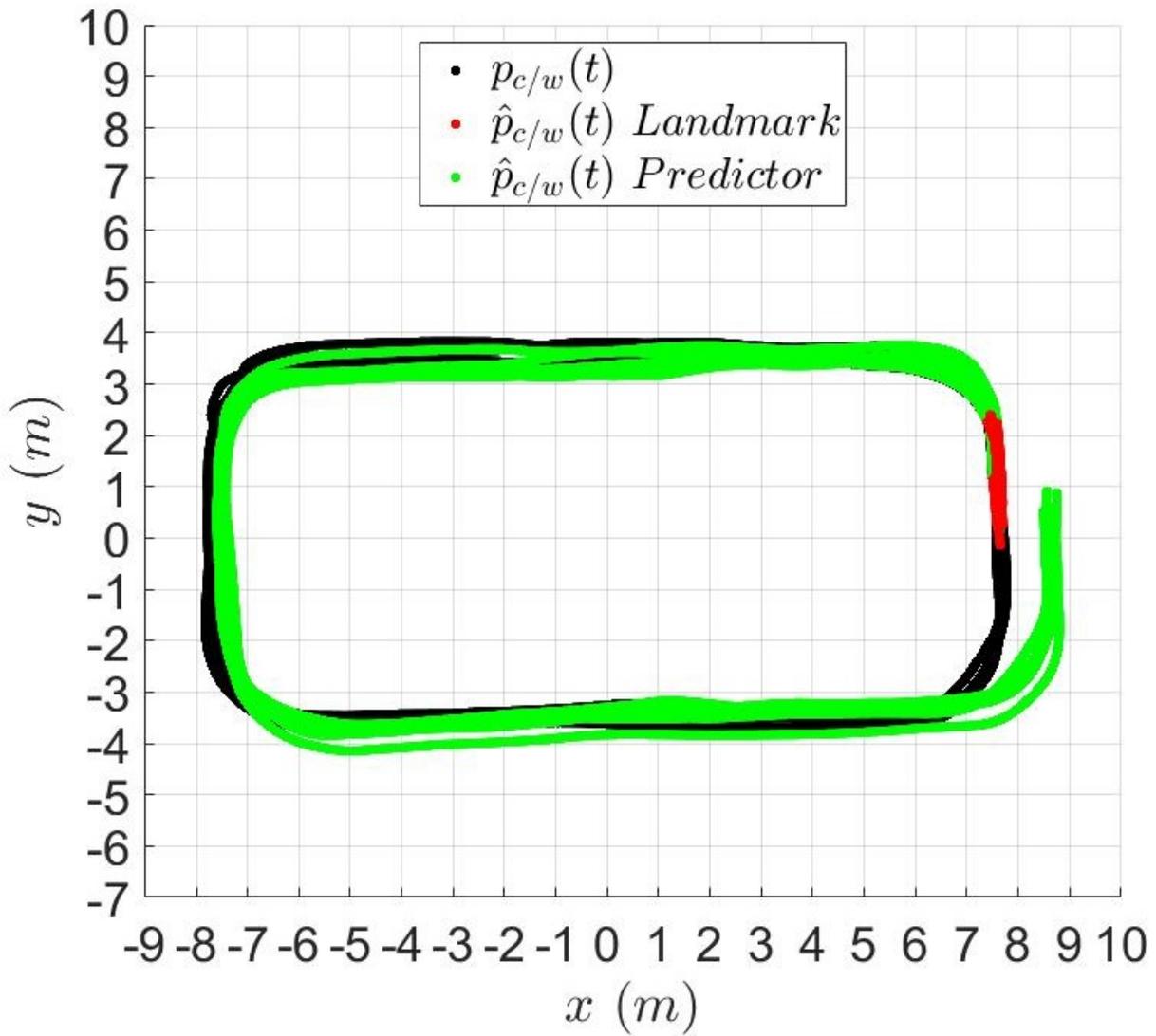


Figure 4-14. Plot of the path of the camera during the experiment and the estimated path of the camera using the predictor-only strategy when no landmark is in the FOV. The true path is marked in black. The estimated path shows if the landmark is in the FOV or the predictor is activated. While the landmark 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.

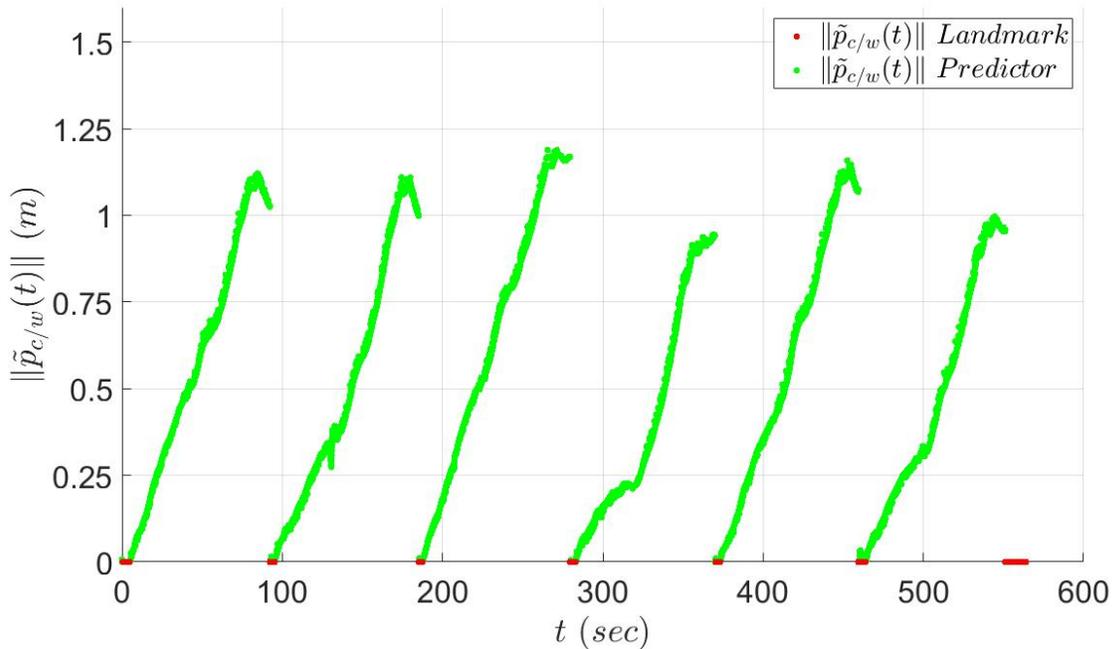


Figure 4-15. Plot of the norm of the camera position error during the experiment using the predictor-only strategy when no landmark is in the FOV. The camera position error shows which estimator is activated over the experiment. While the landmark is in the FOV, the estimated path is shown using the 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.

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 4-14 and 4-15, using a predictor-only strategy, similar to [62] and [63], 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-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 4.1. 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 4-6, 4-7, and 4-13 to Figures 4-16-4-18. Specifically, the maximum position error using a 3 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 to using a 10 pixel standard deviation for accepting data which had a 6.6% average RMS error with a standard

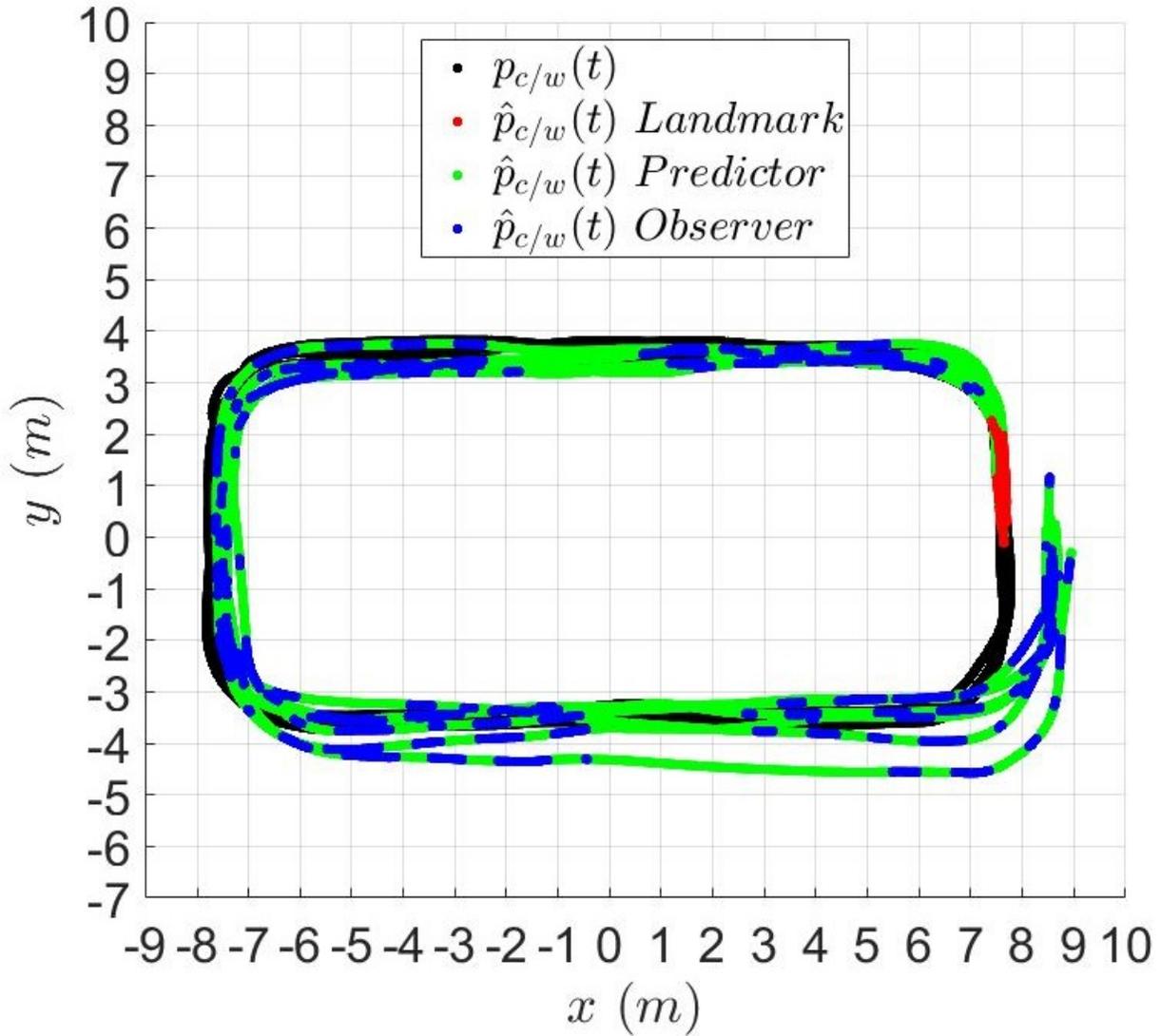


Figure 4-16. Plot of the path of the camera during the experiment and the estimated path of the camera using a standard deviation of 10 pixels for the history stack rejection algorithm. The true path is marked in black. The estimated path shows which estimator is activated over the experiment. While 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 are active, respectively.

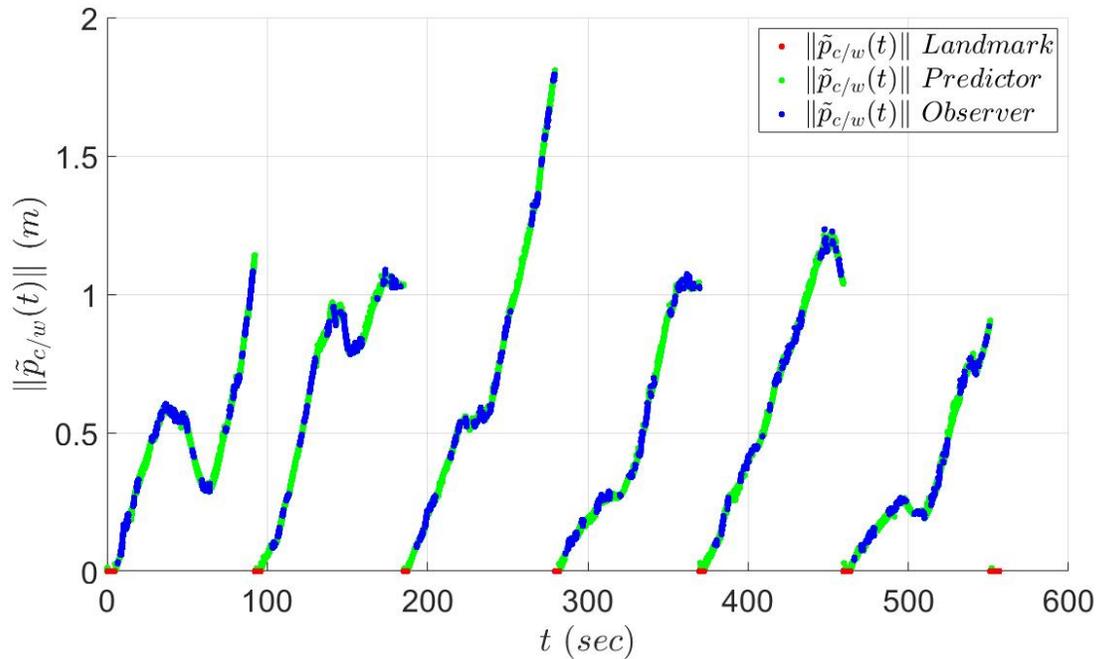


Figure 4-17. Plot of the norm of the camera position error during the experiment using a standard deviation of 10 pixels for accepting data onto the history stack. The camera position error shows which estimator is activated over the experiment. While the landmark is in the FOV, the estimated path is shown using the red marker. Similarly, the estimated path is marked in green or blue when the predictor or observer are 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.

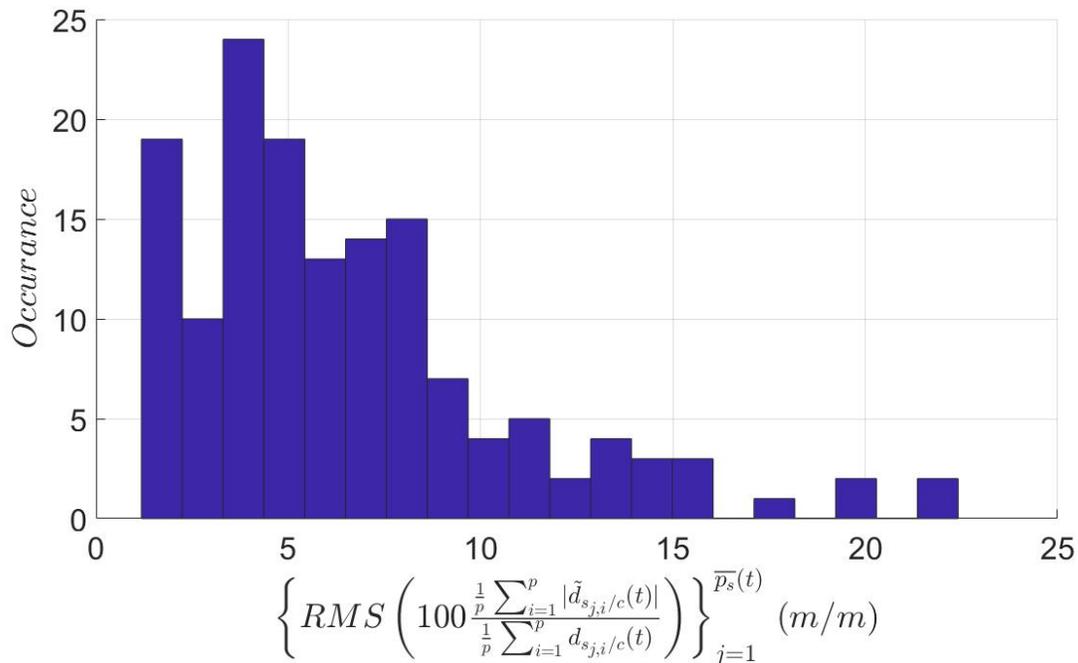


Figure 4-18. Histogram plot of the RMS of the average percentage error of the distance to the features relative to the true distance to the features across after the minimum dwell-time condition is satisfied using a standard deviation of 10 pixels for accepting data onto the history stack. The histogram 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.

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.

4.6 Summary

In this chapter, an extension to the learning approaches in [32], [45], and Chapter 3, is developed that applies a new learning strategy that maintains a continuous estimate of the position of the camera and estimates the structure of features as they become visible. 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. Additionally, this approach shows how the extended observer removes the positive depth constraint required by all previous structure from motion approaches. Using this approach, a camera may travel over large distances without keeping specific features in the FOV for all time and allow objects to permanently leave the FOV if necessary. A Lyapunov based stability analysis proves that the observers for estimating the path of the camera as well as the structure of each set of objects are globally exponentially stable while features are in the camera's FOV. A switched systems analysis is used to develop dwell-time conditions to indicate how long a feature must be tracked to ensure the distance estimation error is below a threshold.

After distance estimates have converged below the threshold, the feature may be used to update the camera position. If a feature does not satisfy the dwell-time condition, it is never used to update the position of the agent. Furthermore, the approach does not require a new set of features to be in the camera's FOV when older features leave the camera's FOV. Finally, if a recognized landmark enters the camera's FOV, the feedback is used to compensate for drift error.

CHAPTER 5 STRUCTURE AND MOTION OF A MOVING TARGET USING A MOVING MONOCULAR CAMERA SUBJECT TO INTERMITTENT FEEDBACK

In this chapter, an approach similar to [45] and Chapter 3 is developed to estimate the initial structure of a moving object. Unlike Chapter 3, intermittent feedback of the object is considered. After estimating the initial structure of the object, an observer and predictor for the object's pose, velocity, and acceleration model (i.e., SaMfM [81]) is guaranteed to be GUUB provided dwell-time conditions on the availability of feedback are satisfied. Specifically, the dwell-time conditions, developed using a Lyapunov-based stability analysis, give upper bounds on the time to learn the object's initial structure and upper bounds on the time feedback of the object's features can be unavailable. The approach to learn the motion model is motivated by [38]; however, an acceleration model is learned in this chapter instead of a velocity model as done in [38]. Estimating an acceleration model relaxes constraints on the motion and enables the use of more general system models. Furthermore, the approach to develop the dwell-times is motivated by [63] where the dwell-times in this chapter are based on the size of the camera's FOV, ensuring estimation error cannot exceed upper thresholds. The dwell-times developed in this chapter ensure an object is recaptured in the FOV after leaving which is not guaranteed in [38]. Specifically, [38] gives an object a number of cycles of leaving and returning to the camera's FOV to guarantee stability; however, in many applications it is not possible to ensure an object returns to the camera's FOV if estimation error grows too large motivating the dwell-times developed in this chapter. Furthermore, this development in this chapter relaxes the positive depth constraint required in [38].

5.1 Learning the First Feature Structure

Similar to the approach for stationary features, the dynamics for the describing the moving object's distances can be integrated over a window of time to determine

a relationship for the initial structure; however, unlike the stationary features, (2–27) and (2–28) are functions of the target’s linear and angular velocity. Because the target’s linear and angular velocity are measurable, relationships must be developed to enable learning the initial structure. Let $q_{m/c}(t)$ be the quaternion representation of the orientation of \mathcal{F}_m with respect to \mathcal{F}_c where its derivative with respect to time is

$$\frac{d}{dt} (q_{m/c}(t)) = \frac{1}{2} B (q_{m/c}(t)) R_{m/c}^T(t) (\underline{\omega}_m(t) - \underline{\omega}_c(t)). \quad (5-1)$$

If the rotation of the object is measurable when in the FOV, $\underline{\omega}_m(t)$ can be estimated by approximating $\frac{d}{dt} (q_{m/c}(t))$ while $\sigma_{\mathcal{O}}(t) = a$ as

$$\underline{\omega}_m(t) = 2R_{m/c}(t) B^T (q_{m/c}(t)) \frac{d}{dt} (q_{m/c}(t)) + \underline{\omega}_c(t), \quad (5-2)$$

where $q_{m/c}(t)$ can be determined while $\sigma_{\mathcal{O}}(t) = a$ implying estimates of $\frac{d}{dt} (q_{m/c}(t))$ can also be determined. The target’s linear velocity does not have any relationship that allows for a direct approach. Yet by examining the rate of change of the direction for the first feature (i.e., the origin of the object) and the rate of change of the relative motion direction, $\underline{v}_m(t)$ can be written as a function of measurable quantities and the initial distance to the first feature. Specifically, using (5–2) in (2–30) for the first feature and (2–31) yields

$$\begin{bmatrix} \xi_{m_1}(t) & 0_{3 \times 1} \\ 0_{3 \times 1} & \xi_{m/m^*}(t) \end{bmatrix} \begin{bmatrix} d_{m_1/c}(t) \\ d_{m/m^*}(t) \end{bmatrix} + \begin{bmatrix} 0_{3 \times 1} \\ \xi_{m/m_1^*}(t) \end{bmatrix} d_{m_1^*/k_1} + \begin{bmatrix} \rho_{m_1}(t) \\ \rho_{m/m^*}(t) \end{bmatrix} = \begin{bmatrix} \Psi_{m_1}(t) \\ \Psi_{m/m^*}(t) \end{bmatrix} \underline{v}_m(t), \quad (5-3)$$

where $\xi_{m_i}(t) \triangleq \left(\frac{d}{dt} (\underline{u}_{m_i/c}(t)) + \underline{\omega}_c^\times(t) \underline{u}_{m_i/c}(t) \right)$, $\rho_{m_i}(t) \triangleq \Psi_{m_i}(t) \underline{v}_c(t)$, $\xi_{m/m^*}(t) \triangleq \left(\frac{d}{dt} (\underline{u}_{m/m^*}(t)) + \underline{\omega}_c^\times(t) \underline{u}_{m/m^*}(t) \right)$, $\rho_{m/m^*}(t) \triangleq \Psi_{m/m^*}(t) \underline{v}_c(t)$, and $\xi_{m/m_1^*}(t) \triangleq \Psi_{m/m^*}(t) \underline{\omega}_m^\times(t) R_{m/c}(t) \underline{u}_{m_1^*/k_1}$ are all measurable while the target is in the camera’s FOV. Substituting the relationship in (2–24) for the first feature into (5–3)

for $\begin{bmatrix} d_{m_1/c}(t) \\ d_{m/m^*}(t) \end{bmatrix}$ and simplifying yields

$$\Xi_{\underline{v}_m}(t) d_{m_1^*/k_1} + P_{\underline{v}_m}(t) = \Psi_{\underline{v}_m}(t) \underline{v}_m(t), \quad (5-4)$$

where $\Psi_{\underline{v}_m}(t) \triangleq \begin{bmatrix} \Psi_{m_1}(t) \\ \Psi_{m/m^*}(t) \end{bmatrix}$, $\Xi_{\underline{v}_m}(t) \triangleq \left(\begin{bmatrix} \xi_{m_1}(t) & 0_{3 \times 1} \\ 0_{3 \times 1} & \xi_{m/m^*}(t) \end{bmatrix} \psi_{m_1}(t) + \begin{bmatrix} 0_{3 \times 1} \\ \xi_{m/m_1^*}(t) \end{bmatrix} \right)$,

and $P_{\underline{v}_m}(t) \triangleq \begin{bmatrix} \rho_{m_1}(t) \\ \rho_{m/m^*}(t) \end{bmatrix}$ are all measurable while the target is in the camera's FOV.

Let $\sigma_{v_m}(t) \in \{u, a\}$ be an indicator signal indicating if $\lambda_{\min} \left\{ \Psi_{\underline{v}_m}^T(t) \Psi_{\underline{v}_m}(t) \right\} \leq \lambda_{v_m}$ or $\lambda_{\min} \left\{ \Psi_{\underline{v}_m}^T(t) \Psi_{\underline{v}_m}(t) \right\} > \lambda_{v_m}$, respectively. If $\sigma_O(t) = a \wedge \sigma_{v_m}(t) = a \wedge \sigma_{m_1}(t) = a$, the body velocity can be written as a function of the constant initial distance to the origin of the object (i.e., the first feature) as

$$\underline{v}_m(t) = \Psi_{\underline{v}_m}^+(t) \Xi_{\underline{v}_m}(t) d_{m_1^*/k_1} + \Psi_{\underline{v}_m}^+(t) P_{\underline{v}_m}(t), \quad (5-5)$$

where $\Psi_{\underline{v}_m}^+(t) = \left(\Psi_{\underline{v}_m}^T(t) \Psi_{\underline{v}_m}(t) \right)^{-1} \Psi_{\underline{v}_m}^T(t)$

Remark 5.1. The set of features, $\{\sigma_{m_i}(t)\}_{m_i \in \mathcal{O}_m}$, are set as $\sigma_{m_i}(t) = u$ if the velocity does not satisfy the eigenvalue condition (i.e., $\sigma_{v_m}(t) = u$). This is done because the features are all dependent on the velocity in the subsequent development.

While $\sigma_O(t) = a \wedge \sigma_{v_m}(t) = a \wedge \sigma_{m_1}(t) = a$, the relationships in (5-5) are used to learn the initial distance to the first feature by substituting (5-5) into (2-27) for the first feature and (2-28), then integrating over a time window $\varsigma \in \mathbb{R}_{>0}$ yielding

$$\begin{aligned} \begin{bmatrix} d_{m_1/c}(t) \\ d_{m/m^*}(t) \end{bmatrix} - \begin{bmatrix} d_{m_1/c}(t-\varsigma) \\ d_{m/m^*}(t-\varsigma) \end{bmatrix} &= d_{m_1^*/k_1} \int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{m_1/c}^T(t) \Psi_{\underline{v}_m}^+(t) \Xi_{\underline{v}_m}(t) \\ \underline{u}_{m/m^*}^T(t) \Psi_{\underline{v}_m}^+(t) \Xi_{\underline{v}_m}(t) \end{bmatrix} dt \\ &\quad - d_{m_1^*/k_1} \int_{t-\varsigma}^t \begin{bmatrix} 0 \\ \underline{u}_{m/m^*}^T(t) \underline{\omega}_m^\times(t) R_{m/c}(t) \underline{u}_{m_1^*/k_1} \end{bmatrix} dt \end{aligned}$$

$$+ \int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{m_1/c}^T(\iota) \Psi_{v_m}^+(\iota) P_{v_m}(\iota) - \underline{u}_{m_1/c}^T(\iota) \underline{v}_c(\iota) \\ \underline{u}_{m/m^*}^T(\iota) \Psi_{v_m}^+(\iota) P_{v_m}(\iota) - \underline{u}_{m/m^*}^T(\iota) \underline{v}_c(\iota) \end{bmatrix} d\iota, \quad (5-6)$$

where ς may be constant in size or change over time, and $\sigma_{\mathcal{O}}(\iota) = a \wedge \sigma_{v_m}(\iota) =$

$a \wedge \sigma_{m_1}(\iota) = a, \forall \iota \in [t - \varsigma, t]$. Again substituting the relationship in (2-24) for the first

feature into (5-6) for $\begin{bmatrix} d_{m_1/c}(t) \\ d_{m/m^*}(t) \end{bmatrix}$ and $\begin{bmatrix} d_{m_1/c}(t - \varsigma) \\ d_{m/m^*}(t - \varsigma) \end{bmatrix}$ yields

$$\mathcal{Y}_{m_1}(t) d_{m_1^*/k_1} = \mathcal{U}_{m_1}(t), \quad (5-7)$$

where

$$\mathcal{Y}_{m_1}(t) \triangleq \begin{cases} \left\{ \begin{array}{l} \psi_{m_1}(t) - \psi_{m_1}(t - \pi_{j,m_1}^{a_i}) \\ - \int_{t-\pi_{j,m_1}^{a_i}}^t \begin{bmatrix} \underline{u}_{m_1/c}^T(\iota) \Psi_{v_m}^+(\iota) \Xi_{v_m}(\iota) \\ \underline{u}_{m/m^*}^T(\iota) \Psi_{v_m}^+(\iota) \Xi_{v_m}(\iota) \end{bmatrix} d\iota \\ + \int_{t-\pi_{j,m_1}^{a_i}}^t \begin{bmatrix} 0 \\ \underline{u}_{m/m^*}^T(\iota) \underline{\omega}_m^\times(\iota) R_{m/c}(\iota) \underline{u}_{m_1^*/k_1} \end{bmatrix} d\iota, & t - \pi_{j,m_1}^{a_i} < \varsigma, \end{array} \right. \\ \left\{ \begin{array}{l} \psi_{m_1}(t) - \psi_{m_1}(t - \varsigma) \\ - \int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{m_1/c}^T(\iota) \Psi_{v_m}^+(\iota) \Xi_{v_m}(\iota) \\ \underline{u}_{m/m^*}^T(\iota) \Psi_{v_m}^+(\iota) \Xi_{v_m}(\iota) \end{bmatrix} d\iota \\ + \int_{t-\varsigma}^t \begin{bmatrix} 0 \\ \underline{u}_{m/m^*}^T(\iota) \underline{\omega}_m^\times(\iota) R_{m/c}(\iota) \underline{u}_{m_1^*/k_1} \end{bmatrix} d\iota, & t - \pi_{j,m_1}^{a_i} \geq \varsigma, \end{array} \right. \end{cases}$$

and

$$\mathcal{U}_{m_1}(t) \triangleq \begin{cases} \int_{t-\pi_j^a}^t \begin{bmatrix} \underline{u}_{m_1/c}^T(\ell) \Psi_{\underline{v}_m}^+(\ell) P_{\underline{v}_m}(\ell) - \underline{u}_{m_1/c}^T(\ell) \underline{v}_c(\ell) \\ \underline{u}_{m/m^*}^T(\ell) \Psi_{\underline{v}_m}^+(\ell) P_{\underline{v}_m}(\ell) - \underline{u}_{m/m^*}^T(\ell) \underline{v}_c(\ell) \end{bmatrix} d\ell, & t - \pi_{j,m_1}^{a_l} < \varsigma, \\ \int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{m_1/c}^T(\ell) \Psi_{\underline{v}_m}^+(\ell) P_{\underline{v}_m}(\ell) - \underline{u}_{m_1/c}^T(\ell) \underline{v}_c(\ell) \\ \underline{u}_{m/m^*}^T(\ell) \Psi_{\underline{v}_m}^+(\ell) P_{\underline{v}_m}(\ell) - \underline{u}_{m/m^*}^T(\ell) \underline{v}_c(\ell) \end{bmatrix} d\ell, & t - \pi_{j,m_1}^{a_l} \geq \varsigma, \end{cases}$$

and $\pi_{j,m_1}^{a_l}$ is the l th instance the first feature satisfies the eigenvalue condition during the j th instance the object enters the camera's FOV (i.e., $t \in [\pi_j^a, \pi_j^u]$). The time $\pi_{j,m_1}^{a_l}$ must be considered given there is no guarantee the object is learned before time $\pi_{j,m_1}^{u_l}$ or time π_j^u .

Multiplying both sides of (5–7) by $\mathcal{Y}_{m_1}^T(t)$ yields

$$\mathcal{Y}_{m_1}^T(t) \mathcal{Y}_{m_1}(t) d_{m_1^*/k_1} = \mathcal{Y}_{m_1}^T(t) \mathcal{U}_{m_1}(t). \quad (5-8)$$

In general, $\mathcal{Y}_{m_1}(t)$ will not have full column rank while $\sigma_{\mathcal{O}}(\ell) = a \wedge \sigma_{v_m}(\ell) = a \wedge \sigma_{m_1}(\ell) = a, \forall \ell \in [t - \varsigma, t]$ (e.g. when the camera and object are stationary implying $\mathcal{Y}_{m_1}^T(t) \mathcal{Y}_{m_1}(t) \geq 0$). However, the equality in (5–8) may be evaluated at any instance in time and summed together (i.e., history stacks) yielding

$$\Sigma_{\mathcal{Y}_{m_1}} d_{m_1^*/k_1} = \Sigma_{\mathcal{U}_{m_1}}, \quad (5-9)$$

where $\Sigma_{\mathcal{Y}_{m_1}} \triangleq \sum_{h=1}^N \mathcal{Y}_{m_1}^T(t_h) \mathcal{Y}_{m_1}(t_h)$, $\Sigma_{\mathcal{U}_{m_1}} \triangleq \sum_{h=1}^N \mathcal{Y}_{m_1}^T(t_h) \mathcal{U}_{m_1}(t_h)$, $t_h \in (\pi_1^a, t]$, and $N \in \mathbb{Z}_{>1}$.

Assumption 5.1. There is sufficient relative motion between the camera and target so there exists a time $\tau_{m_1} \in \mathbb{R}_{>\pi_1^a}$, such that for all time $t > \tau_{m_1}$, $\lambda_{\min} \{ \Sigma_{\mathcal{Y}_{m_1}} \} > \lambda_{\tau}$.

Remark 5.2. Learning the initial distance to the first feature enables learning the remaining features on the target. Specific to the target tracking objective, learning the initial distance to the first feature provides sufficient information to determine the relative position of the target with respect to the camera, while the target is in the FOV and the relative motion is not parallel. However, since it is often desirable to obtain a continuous

estimate of the targets velocity, learning the structure of all the features will enable a more robust estimate of the velocity. Specifically, $\sigma_{v_m}(t) = u$ may occur, implying the first feature cannot estimate $\underline{v}_m(t)$. The time τ_{m_1} is unknown; however, it can be determined online by checking the minimum eigenvalue of $\Sigma_{\mathcal{Y}_{m_1}}$.

Once sufficient relative motion occurs as discussed in Assumption 5.1, the constant unknown distance, $d_{m_1^*/k_1}$, can be determined from (5–9) yielding

$$d_{m_1^*/k_1} = \mathcal{X}_{m_1}, \quad (5-10)$$

where $\mathcal{X}_{m_1} \triangleq \Sigma_{\mathcal{Y}_{m_1}}^{-1} \Sigma \underline{u}_{m_1}$. Given $\underline{p}_{m_1/c}(t) = \underline{u}_{m_1/c}(t) d_{m_1/c}(t)$, while $\sigma_{\mathcal{O}}(t) = a \wedge \sigma_{m_1}(t) = a \wedge t > \tau_{m_1}$ using (5–10) in (2–24) yields

$$\underline{p}_{m_1/c}(t) = \underline{u}_{m_1/c}(t) \psi_{m_1,1}(t) \mathcal{X}_{m_1}, \quad (5-11)$$

where $\psi_{m_1,1}(t)$ is the first element of $\psi_{m_1}(t)$.

5.2 Learning the Structure of the Remaining Features

After the initial structure for the origin has been learned, the linear velocity of the object can be determined while $\sigma_{\mathcal{O}}(t) = a \wedge \sigma_{v_m}(t) = a \wedge \sigma_{m_1}(t) = a \wedge t > \tau_{m_1}$ using (5–10) in (5–5) yielding

$$\underline{v}_m(t) = \Psi_{\underline{v}_m}^+(t) \Xi_{\underline{v}_m}(t) \mathcal{X}_{m_1} + \Psi_{\underline{v}_m}^+(t) P_{\underline{v}_m}(t). \quad (5-12)$$

Using (5–10) and (5–12) while $\sigma_{m_i}(t) = a \wedge \sigma_{\mathcal{O}}(t) = a \wedge \sigma_{v_m}(t) = a \wedge \sigma_{m_1}(t) = a \wedge t > \tau_{m_1}$ the i th feature on the object, $m_i \in \mathcal{O}_m$, is learned while it satisfies the eigenvalue condition (i.e., $\sigma_{m_i}(t) = a$) by integrating (2–27) for m_i and (2–28) similar to (5–6) yielding

$$\begin{bmatrix} d_{m_i/c}(t) \\ d_{m/m^*}(t) \end{bmatrix} - \begin{bmatrix} d_{m_i/c}(t - \varsigma) \\ d_{m/m^*}(t - \varsigma) \end{bmatrix} = d_{m_i^*/k_1} \int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{m_i/c}^T(t) \underline{\omega}_m^\times(t) R_{m/c}(t) \underline{u}_{m_i^*/k_1} \\ 0 \end{bmatrix} dt$$

$$\begin{aligned}
& + \mathcal{X}_{m_1} \int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{m_i/c}^T(\iota) \Psi_{\underline{v}_m}^+(\iota) \Xi_{\underline{v}_m}(\iota) \\ \underline{u}_{m/m^*}^T(\iota) \Psi_{\underline{v}_m}^+(\iota) \Xi_{\underline{v}_m}(\iota) \end{bmatrix} d\iota \\
& - \mathcal{X}_{m_1} \int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{m_i/c}^T(\iota) \underline{\omega}_m^\times(\iota) R_{m/c}(\iota) \underline{u}_{m_i^*/k_1} \\ \underline{u}_{m/m^*}^T(\iota) \underline{\omega}_m^\times(\iota) R_{m/c}(\iota) \underline{u}_{m_i^*/k_1} \end{bmatrix} d\iota \\
& + \int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{m_i/c}^T(\iota) \Psi_{\underline{v}_m}^+(\iota) P_{\underline{v}_m}(\iota) - \underline{u}_{m_i/c}^T(\iota) \underline{v}_c(\iota) \\ \underline{u}_{m/m^*}^T(\iota) \Psi_{\underline{v}_m}^+(\iota) P_{\underline{v}_m}(\iota) - \underline{u}_{m/m^*}^T(\iota) \underline{v}_c(\iota) \end{bmatrix} d\iota, \quad (5-13)
\end{aligned}$$

where $\sigma_{\mathcal{O}}(\iota) = a \wedge \sigma_{v_m}(\iota) = a \wedge \sigma_{m_1}(\iota) = a \wedge \sigma_{m_i}(\iota) = a, \forall \iota \in [t - \varsigma, t] \wedge t > \tau_{m_1}$.

Again substituting the relationship in (2-24) for m_i into (5-13) for $\begin{bmatrix} d_{m_i/c}(t) \\ d_{m/m^*}(t) \end{bmatrix}$ and

$\begin{bmatrix} d_{m_i/c}(t - \varsigma) \\ d_{m/m^*}(t - \varsigma) \end{bmatrix}$ yields

$$\mathcal{Y}_{m_i}(t) d_{m_i^*/k_1} = \mathcal{U}_{m_i}(t), \quad (5-14)$$

where

$$\mathcal{Y}_{m_1}(t) \triangleq \begin{cases} \begin{bmatrix} \psi_{m_i}(t) - \psi_{m_i}(t - \pi_{j,m_i}^{a_i}) \\ - \int_{t-\pi_{j,m_i}^{a_i}}^t \begin{bmatrix} \underline{u}_{m_i/c}^T(\iota) \underline{\omega}_m^\times(\iota) R_{m/c}(\iota) \underline{u}_{m_i^*/k_1} \\ 0 \end{bmatrix} d\iota, & t - \pi_{j,m_i}^{a_i} < \varsigma, \\ \psi_{m_i}(t) - \psi_{m_i}(t - \varsigma) \\ - \int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{m_i/c}^T(\iota) \underline{\omega}_m^\times(\iota) R_{m/c}(\iota) \underline{u}_{m_i^*/k_1} \\ 0 \end{bmatrix} d\iota, & t - \pi_{j,m_i}^{a_i} \geq \varsigma, \end{cases}$$

and

$$\mathcal{U}_{m_i}(t) \triangleq \begin{cases} \mathcal{X}_{m_1} \int_{t-\pi_{j,m_i}^{a_l}}^t \begin{bmatrix} \underline{u}_{m_i/c}^T(\iota) \Psi_{\underline{v}_m}^+(\iota) \Xi_{\underline{v}_m}(\iota) \\ \underline{u}_{m/m^*}^T(\iota) \Psi_{\underline{v}_m}^+(\iota) \Xi_{\underline{v}_m}(\iota) \end{bmatrix} d\iota \\ - \mathcal{X}_{m_1} \int_{t-\pi_{j,m_i}^{a_l}}^t \begin{bmatrix} \underline{u}_{m_i/c}^T(\iota) \underline{\omega}_m^\times(\iota) R_{m/c}(\iota) \underline{u}_{m_1^*/k_1} \\ \underline{u}_{m/m^*}^T(\iota) \underline{\omega}_m^\times(\iota) R_{m/c}(\iota) \underline{u}_{m_1^*/k_1} \end{bmatrix} d\iota \\ + \int_{t-\pi_{j,m_i}^{a_l}}^t \begin{bmatrix} \underline{u}_{m_i/c}^T(\iota) \Psi_{\underline{v}_m}^+(\iota) P_{\underline{v}_m}(\iota) - \underline{u}_{m_i/c}^T(\iota) \underline{v}_c(\iota) \\ \underline{u}_{m/m^*}^T(\iota) \Psi_{\underline{v}_m}^+(\iota) P_{\underline{v}_m}(\iota) - \underline{u}_{m/m^*}^T(\iota) \underline{v}_c(\iota) \end{bmatrix} d\iota, \quad t - \pi_{j,m_i}^{a_l} < \varsigma, \\ \mathcal{X}_{m_1} \int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{m_i/c}^T(\iota) \Psi_{\underline{v}_m}^+(\iota) \Xi_{\underline{v}_m}(\iota) \\ \underline{u}_{m/m^*}^T(\iota) \Psi_{\underline{v}_m}^+(\iota) \Xi_{\underline{v}_m}(\iota) \end{bmatrix} d\iota \\ - \mathcal{X}_{m_1} \int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{m_i/c}^T(\iota) \underline{\omega}_m^\times(\iota) R_{m/c}(\iota) \underline{u}_{m_1^*/k_1} \\ \underline{u}_{m/m^*}^T(\iota) \underline{\omega}_m^\times(\iota) R_{m/c}(\iota) \underline{u}_{m_1^*/k_1} \end{bmatrix} d\iota \\ + \int_{t-\varsigma}^t \begin{bmatrix} \underline{u}_{m_i/c}^T(\iota) \Psi_{\underline{v}_m}^+(\iota) P_{\underline{v}_m}(\iota) - \underline{u}_{m_i/c}^T(\iota) \underline{v}_c(\iota) \\ \underline{u}_{m/m^*}^T(\iota) \Psi_{\underline{v}_m}^+(\iota) P_{\underline{v}_m}(\iota) - \underline{u}_{m/m^*}^T(\iota) \underline{v}_c(\iota) \end{bmatrix} d\iota, \quad t - \pi_{j,m_i}^{a_l} \geq \varsigma, \end{cases}$$

and $\pi_{j,m_i}^{a_l}$ is the l th time the i th feature satisfies the eigenvalue condition during the j th time the object enters the camera's FOV and **sufficient data has been collected for the first feature** (i.e., $\pi_{j,m_i}^{a_l} \in [\pi_j^a, \pi_j^u] \cap t > \tau_{m_1}$). The time $\pi_{j,m_i}^{a_l} \in [\pi_j^a, \pi_j^u] \cap t > \tau_{m_1}$ must be considered given there is no guarantee the object is learned before time $\pi_{j,m_i}^{u_l}$ or π_j^u .

Multiplying both sides of (5-14) by $\mathcal{Y}_{m_i}^T(t)$ yields

$$\mathcal{Y}_{m_i}^T(t) \mathcal{Y}_{m_i}(t) d_{m_i^*/k_1} = \mathcal{Y}_{m_i}^T(t) \mathcal{U}_{m_i}(t). \quad (5-15)$$

In general, $\mathcal{Y}_{m_i}(t)$ will not have full column rank while $\sigma_{\mathcal{O}}(\iota) = a \wedge \sigma_{v_m}(\iota) = a \wedge \sigma_{m_1}(\iota) = a \wedge \sigma_{m_i}(\iota) = a, \forall \iota \in [t - \varsigma, t] \wedge t > \tau_{m_1}$ (e.g. when the camera and object are stationary implying $\mathcal{Y}_{m_i}^T(t) \mathcal{Y}_{m_i}(t) \geq 0$). However, the equality in (5-15) may be evaluated at any instance in time and summed together (i.e., history stacks) yielding

$$\Sigma_{\mathcal{Y}_{m_i}} d_{m_i^*/k_1} = \Sigma_{\mathcal{U}_{m_i}}, \quad (5-16)$$

where $\Sigma_{\mathcal{Y}_{m_i}} \triangleq \sum_{h=1}^N \mathcal{Y}_{m_i}^T(t_h) \mathcal{Y}_{m_i}(t_h)$, $\Sigma_{\mathcal{U}_{m_i}} \triangleq \sum_{h=1}^N \mathcal{Y}_{m_i}^T(t_h) \mathcal{U}_{m_i}(t_h)$, $t_h \in (\tau_{m_i}, t]$, and $N \in \mathbb{Z}_{>1}$.

Assumption 5.2. There is sufficient relative between the camera and target so there exists a time $\tau_{m_i} \in \mathbb{R}_{>\pi_1^a}$, such that for all time $t > \tau_{m_i}$, $\lambda_{\min} \{ \Sigma_{\mathcal{Y}_{m_i}} \} > \lambda_\tau$.

Once sufficient relative motion occurs as discussed in Assumption 5.2, the constant unknown distance, $d_{m_i^*/k_1}$, can be determined from (5-16) yielding

$$d_{m_i^*/k_1} = \mathcal{X}_{m_i}, \quad (5-17)$$

where $\mathcal{X}_{m_i} \triangleq \Sigma_{\mathcal{Y}_{m_i}}^{-1} \Sigma_{\mathcal{U}_{m_i}}$.

5.3 Learning the Object Motion Model

After learning the structure of the object, estimates of object pose and velocity will be available while the object is in the camera's FOV and the eigenvalue conditions are satisfied; however, the object may periodically leave the camera's FOV over time and the eigenvalue conditions will not always be satisfied. The work in [38] used motion model learning to design a predictor for the pose of a target while the target is outside the camera's FOV which is naturally extended to include time periods where the eigenvalue conditions are not satisfied. In the subsequent development, an estimator for the motion model of the vehicle is presented; however, the primary difference in the subsequent design is the model learned is for the acceleration of the vehicle and not the velocity. Specifically, [38] developed a method for modeling the velocity of the target but estimating a model of the acceleration enables the use of kinetic models of targets to be exploited rather than only using kinematic models. As described in [38], there are numerous applications where a target's velocity is directly a function of its pose in the world or the relative pose between the target and the camera; however, these kinematic models or approximations of the desired trajectory of a target don't accurately model

vehicle trajectories in numerous applications. In the subsequent development, a more general model is presented which enables the estimation of a larger class of systems.

Let $\eta_c(t) \triangleq \begin{bmatrix} p_{c/k_1}(t) \\ q_{c/k_1}(t) \end{bmatrix}$ and $\eta_m(t) \triangleq \begin{bmatrix} p_{m_1/c}(t) \\ q_{m/c}(t) \end{bmatrix}$ represent the pose of the camera expressed in \mathcal{F}_{k_1} and the pose of the object with respect to the camera expressed in \mathcal{F}_c .

Additionally, let $\phi_m(t) \triangleq \begin{bmatrix} v_m(t) \\ \omega_m(t) \end{bmatrix}$ represent the velocity of the object expressed in \mathcal{F}_c .

Assumption 5.3. The pose of the camera, $\eta_c(t)$, is known.

Assumption 5.4. The pose and velocity of the moving object and camera are bounded.

Specifically, $\eta_m(t) \in N_{\eta_m}$, $\eta_c(t) \in N_{\eta_c}$, and $\phi_m(t) \in N_{\phi_m}$ where $N_{\eta_m}, N_{\eta_c} \subset \mathbb{R}^7$ and $N_{\phi_m} \subset \mathbb{R}^6$ are convex, compact sets and $\|\phi_m(t)\| \leq \overline{\phi_m}$.

Remark 5.3. Assumption 5.4 is a general requirement for any estimator to converge (i.e., the state to be estimated must remain bounded for an estimator to remain bounded). This is equivalent to the requirement of desired trajectories remaining bounded in control problems.

Assumption 5.5. The acceleration of the moving object is bounded and limited to the class of systems that are bounded and are locally Lipschitz functions of the pose and velocity of the moving object and pose of the camera. Specifically, the derivative of the velocity with respect to time is

$$\frac{d}{dt}(\phi_m(t)) = f_m(\eta_m(t), \eta_c(t), \phi_m(t)) \quad (5-18)$$

where $f_m : \mathbb{R}^7 \times \mathbb{R}^7 \times \mathbb{R}^6 \rightarrow \mathbb{R}^6$ is a locally Lipschitz and bounded function.

Remark 5.4. Assumption 5.5 ensures the kinetic model of the target can be approximated using universal function approximators to an arbitrary level of accuracy via the Stone-Weierstrass theorem [120]. Specifically, a neural network (NN) is subsequently used to approximate $f_m(\eta_m(t), \eta_c(t), \phi_m(t))$. This assumption holds in various target tracking objectives (e.g., projectile and orbital motion, pursuit-evasion games, and applications where a target is traveling through an environment with stationary obstacles).

Remark 5.5. The acceleration model is not limited to targets that are a function of $\eta_m(t)$, $\eta_c(t)$, and $\phi_m(t)$, specifically, the target's motion model may be a function of any combination of $\eta_m(t)$, $\eta_c(t)$, and $\phi_m(t)$.

In this development, the target's acceleration model is approximated using a neural network of the form

$$f_m(\eta_m(t), \eta_c(t), \phi_m(t)) = W_m^T \sigma_m(\eta_m(t), \eta_c(t), \phi_m(t)) + \varepsilon_m(\eta_m(t), \eta_c(t), \phi_m(t)), \quad (5-19)$$

where $L \in \mathbb{Z}_{>0}$ is the number of basis functions, $W_m \in \mathbb{R}^{L \times 6}$ is a matrix of the constant unknown ideal weights, $\sigma_m : \mathbb{R}^7 \times \mathbb{R}^7 \times \mathbb{R}^6 \rightarrow \mathbb{R}^6$ is a designed matrix of basis functions that are bounded and locally Lipschitz, and $\varepsilon_m : \mathbb{R}^7 \times \mathbb{R}^7 \times \mathbb{R}^6 \rightarrow \mathbb{R}^6$ is the function approximation residual, which is locally Lipschitz and can be bounded with a bound that can be made arbitrarily small based on the Stone-Weierstrass theorem $\overline{\varepsilon}_m \triangleq \sup_{\eta_m \in N_m, \eta_c \in N_c, \phi_m \in \Phi_m, t \in [0, \infty)} \|\varepsilon_m(\eta_m(t), \eta_c(t), \phi_m(t))\|$. Furthermore, $\|W_m\| \leq \overline{W}_m \in \mathbb{R}_{>0}$, $\overline{\sigma}_m \triangleq \sup_{\eta_m \in N_m, \eta_c \in N_c, \phi_m \in \Phi_m, t \in [0, \infty)} \|\sigma_m(\eta_m(t), \eta_c(t), \phi_m(t))\|$, $\overline{\sigma}_{m, \eta_m} \triangleq \sup_{\eta_m \in N_m, \eta_c \in N_c, \phi_m \in \Phi_m, t \in [0, \infty)} \left\| \frac{\partial \sigma_m}{\partial \eta_m(t)}(\eta_m(t), \eta_c(t), \phi_m(t)) \right\|$, and $\overline{\sigma}_{m, \phi_m} \triangleq \sup_{\eta_m \in N_m, \eta_c \in N_c, \phi_m \in \Phi_m, t \in [0, \infty)} \left\| \frac{\partial \sigma_m}{\partial \phi_m(t)}(\eta_m(t), \eta_c(t), \phi_m(t)) \right\|$.

Remark 5.6. After the initial structure of the first feature $m_1 \in \mathcal{O}_m$ is known, while the eigenvalue condition is satisfied and the object remains in the camera's FOV (i.e., $\sigma_{\mathcal{O}}(t) = a \wedge \sigma_{v_m}(t) = a \wedge \sigma_{m_1}(t) = a \wedge t > \tau_{m_1}$), $\eta_m(t)$ can be approximated by using (5-11). Using (5-12) and (5-2), $\phi_m(t)$ can be determined. Furthermore, after the initial structure for the other features has been learned, any or all of the features can be used to estimate the $\eta_m(t)$ and $\phi_m(t)$; however, in the subsequent development only the origin of the object is used (i.e., the first feature $m_1 \in \mathcal{O}_m$).

Similar to the approach taken for the moving features, using (5-19), the derivative of the velocity with respect to time is integrated over a time window ς while $\sigma_{\mathcal{O}}(t) =$

$$a \wedge \sigma_{v_m}(t) = a \wedge \sigma_{m_1}(t) = a \wedge t > \tau_{m_1}$$

$$\begin{aligned} \phi_m(t) - \phi_m(t - \varsigma) &= W_m^T \int_{t-\varsigma}^t \sigma_m(\eta_m(\iota), \eta_c(\iota), \phi_m(\iota)) d\iota \\ &+ \int_{t-\varsigma}^t \varepsilon_m(\eta_m(\iota), \eta_c(\iota), \phi_m(\iota)) d\iota, \end{aligned} \quad (5-20)$$

where $\sigma_{\mathcal{O}}(\iota) = a \wedge \sigma_{v_m}(\iota) = a \wedge \sigma_{m_1}(\iota) = a, \forall \iota \in [t - \varsigma, t] \wedge t > \tau_{m_1}$. Using (5-12) and (5-2),

$$\phi_m(t) = \begin{bmatrix} \Psi_{\underline{v}_m}^+(t) \Xi_{\underline{v}_m}(t) \mathcal{X}_{m_1} + \Psi_{\underline{v}_m}^+(t) P_{\underline{v}_m}(t) \\ 2R_{m/c}(t) B^T(q_{m/c}(t)) \frac{d}{dt}(q_{m/c}(t)) + \underline{\omega}_c(t) \end{bmatrix}, \quad (5-21)$$

implying (5-20) can be written as

$$\mathcal{Y}_m(t) W_m = \mathcal{U}_m(t) + \mathcal{E}_m(t), \quad (5-22)$$

where

$$\mathcal{Y}_m(t) \triangleq \begin{cases} \int_{t-\pi_{j,m_1}^{a_l}}^t \sigma_m^T(\eta_m(\iota), \eta_c(\iota), \phi_m(\iota)) d\iota, & t - \pi_{j,m_1}^{a_l} < \varsigma, \\ \int_{t-\varsigma}^t \sigma_m^T(\eta_m(\iota), \eta_c(\iota), \phi_m(\iota)) d\iota, & t - \pi_{j,m_1}^{a_l} \geq \varsigma, \end{cases}$$

$$\mathcal{U}_m(t) \triangleq \left\{ \begin{array}{l} \left[\begin{array}{l} \Psi_{\underline{v}_m}^+(t) \Xi_{\underline{v}_m}(t) \mathcal{X}_{m_1} \\ 2R_{m/c}(t) B^T(q_{m/c}(t)) \frac{d}{dt}(q_{m/c}(t)) \end{array} \right]^T + \left[\begin{array}{l} \Psi_{\underline{v}_m}^+(t) P_{\underline{v}_m}(t) \\ \underline{\omega}_c(t) \end{array} \right]^T \\ - \left[\begin{array}{l} \Psi_{\underline{v}_m}^+(t - \pi_{j,m_1}^{a_l}) \Xi_{\underline{v}_m}(t - \pi_{j,m_1}^{a_l}) \mathcal{X}_{m_1} \\ 2R_{m/c}(t - \pi_{j,m_1}^{a_l}) B^T(q_{m/c}(t - \pi_{j,m_1}^{a_l})) \frac{d}{dt}(q_{m/c}(t - \pi_{j,m_1}^{a_l})) \end{array} \right]^T \\ - \left[\begin{array}{l} \Psi_{\underline{v}_m}^+(t - \pi_{j,m_1}^{a_l}) P_{\underline{v}_m}(t - \pi_{j,m_1}^{a_l}) \\ \underline{\omega}_c(t - \pi_{j,m_1}^{a_l}) \end{array} \right]^T, \quad t - \pi_{j,m_1}^{a_l} < \varsigma, \\ \left[\begin{array}{l} \Psi_{\underline{v}_m}^+(t) \Xi_{\underline{v}_m}(t) \mathcal{X}_{m_1} \\ 2R_{m/c}(t) B^T(q_{m/c}(t)) \frac{d}{dt}(q_{m/c}(t)) \end{array} \right]^T + \left[\begin{array}{l} \Psi_{\underline{v}_m}^+(t) P_{\underline{v}_m}(t) \\ \underline{\omega}_c(t) \end{array} \right]^T \\ - \left[\begin{array}{l} \Psi_{\underline{v}_m}^+(t - \varsigma) \Xi_{\underline{v}_m}(t - \varsigma) \mathcal{X}_{m_1} \\ 2R_{m/c}(t - \varsigma) B^T(q_{m/c}(t - \varsigma)) \frac{d}{dt}(q_{m/c}(t - \varsigma)) \end{array} \right]^T \\ - \left[\begin{array}{l} \Psi_{\underline{v}_m}^+(t - \varsigma) P_{\underline{v}_m}(t - \varsigma) \\ \underline{\omega}_c(t - \varsigma) \end{array} \right]^T, \quad t - \pi_{j,m_1}^{a_l} \geq \varsigma, \end{array} \right.$$

and

$$\mathcal{E}_m(t) \triangleq \begin{cases} - \int_{t - \pi_{j,m_1}^{a_l}}^t \varepsilon_m^T(\eta_m(\iota), \eta_c(\iota), \phi_m(\iota)) d\iota, & t - \pi_{j,m_1}^{a_l} < \varsigma, \\ - \int_{t - \varsigma}^t \varepsilon_m^T(\eta_m(\iota), \eta_c(\iota), \phi_m(\iota)) d\iota, & t - \pi_{j,m_1}^{a_l} \geq \varsigma. \end{cases}$$

Multiplying both sides of (5-22) by $\mathcal{Y}_m^T(t)$ yields

$$\mathcal{Y}_m^T(t) \mathcal{Y}_m(t) W_m = \mathcal{Y}_m^T(t) \mathcal{U}_m(t) + \mathcal{Y}_m^T(t) \mathcal{E}_m(t). \quad (5-23)$$

The matrix $\mathcal{Y}_m(t)$ will never have full column rank; however, the equality in (5-23) may be evaluated at any instance in time and summed together (i.e., history stacks) yielding

$$\Sigma \mathcal{Y}_m W_m = \Sigma \mathcal{U}_m + \Sigma \mathcal{E}_m, \quad (5-24)$$

where $\Sigma_{\mathcal{Y}_m} \triangleq \sum_{h=1}^{N_W} \mathcal{Y}_m^T(t_h) \mathcal{Y}_m(t_h)$, $\Sigma_{\mathcal{U}_m} \triangleq \sum_{h=1}^{N_W} \mathcal{Y}_m^T(t_h) \mathcal{U}_m(t_h)$, $\Sigma_{\mathcal{E}_m} \triangleq \sum_{h=1}^{N_W} \mathcal{Y}_m^T(t_h) \mathcal{E}_m(t_h)$, $t_h \in (\tau_{m_1}, t]$, and $N_W \in \mathbb{Z}_{>L}$.

Assumption 5.6. There is sufficient relative motion between the camera and target so there exists a time $\tau_m \in \mathbb{R}_{>\tau_{m_1}}$, such that for all time $t > \tau_m$, $\lambda_{\min} \{\Sigma_{\mathcal{Y}_m}\} > \lambda_\tau$.

Remark 5.7. The time τ_m is unknown; however, it can be determined online by checking the minimum eigenvalue of $\Sigma_{\mathcal{Y}_m}$.

5.4 Target Estimators

To quantify the pose and velocity estimation objective, let

$$\tilde{\eta}_m(t) \triangleq \eta_m(t) - \hat{\eta}_m(t), \quad (5-25)$$

$$\tilde{\phi}_m(t) \triangleq \phi_m(t) - \hat{\phi}_m(t), \quad (5-26)$$

and

$$\tilde{W}_m(t) \triangleq W_m - \hat{W}_m(t), \quad (5-27)$$

where $\hat{\eta}_m(t) \in \mathbb{R}^7$, $\hat{\phi}_m(t) \in \mathbb{R}^6$, and $\hat{W}_m(t) \in \mathbb{R}^{L \times 6}$ are the estimates of $\eta_m(t)$, $\phi_m(t)$, and W_m , respectively. Taking the time derivative of (5-25)-(5-27) and using (2-25) for feature m_1 , (5-1), (5-18), (5-19), and given $\frac{d}{dt}(W_m) = 0_{L \times 6}$ yields

$$\frac{d}{dt}(\tilde{\eta}_m(t)) = \Phi_m(q_{m/c}(t)) \phi_m(t) + \begin{bmatrix} -\underline{v}_c(t) - \underline{\omega}_c^\times(t) \underline{p}_{m_1/c}(t) \\ -\frac{1}{2}B(q_{m/c}(t)) R_{m/c}^T(t) \underline{\omega}_c(t) \end{bmatrix} - \frac{d}{dt}(\hat{\eta}_m(t)), \quad (5-28)$$

$$\frac{d}{dt}(\tilde{\phi}_m(t)) = W_m^T \sigma_m(\eta_m(t), \eta_c(t), \phi_m(t)) + \varepsilon_m(\eta_m(t), \eta_c(t), \phi_m(t)) - \frac{d}{dt}(\hat{\phi}_m(t)), \quad (5-29)$$

and

$$\frac{d}{dt}(\tilde{W}_m(t)) \triangleq -\frac{d}{dt}(\hat{W}_m(t)), \quad (5-30)$$

where $\Phi_m(q_{m/c}(t)) \triangleq \begin{bmatrix} I_{3 \times 3} & 0_{3 \times 3} \\ 0_{4 \times 3} & \frac{1}{2}B(q_{m/c}(t)) R_{m/c}^T(t) \end{bmatrix}$, $\|\Phi_m(q_{m/c}(t))\| \leq 1$, and

$$\overline{\Phi}_{m, q_{m/c}} \triangleq \sup_{\eta_m \in N_m, t \in [0, \infty)} \left\| \frac{\partial \Phi_m}{\partial q_{m/c}(t)}(q_{m/c}(t)) \right\|$$

While the target is in the camera's FOV and the eigenvalue conditions are satisfied but the first feature learning condition is not satisfied, $\sigma_{\mathcal{O}}(t) = a \wedge \sigma_{v_m}(t) = a \wedge \sigma_{m_1}(t) = a \wedge t < \tau_{m_1}$, observer update laws are designed for $\frac{d}{dt}(\hat{\eta}_m(t))$, $\frac{d}{dt}(\hat{\phi}_m(t))$, and $\frac{d}{dt}(\hat{W}_m(t))$ as

$$\frac{d}{dt}(\hat{\eta}_m(t)) \triangleq \text{proj} \left(\Phi_m(q_{m/c}(t)) \hat{\phi}_m(t) + \begin{bmatrix} -\underline{v}_c(t) - \underline{\omega}_c^\times(t) \underline{p}_{m_1/c}(t) \\ -\frac{1}{2}B(q_{m/c}(t)) R_{m/c}^T(t) \underline{\omega}_c(t) \end{bmatrix} \right), \quad (5-31)$$

$$\frac{d}{dt}(\hat{\phi}_m(t)) \triangleq \text{proj} \left(\hat{W}_m^T \sigma_m(\hat{\eta}_m(t), \eta_c(t), \hat{\phi}_m(t)) \right), \quad (5-32)$$

and

$$\frac{d}{dt}(\hat{W}_m(t)) \triangleq 0_{L \times 6}. \quad (5-33)$$

After the learning condition is satisfied (i.e., $t \geq \tau_{m_1}$), and given $\underline{u}_{m_1/c}(t)$ and $q_{m/c}(t)$ can be determined while $\sigma_{\mathcal{O}}(t) = a \wedge \sigma_{v_m}(t) = a \wedge \sigma_{m_1}(t) = a$, (5-11) and (5-21) can be used to determine $\eta_m(t)$ and $\phi_m(t)$ as

$$\eta_m(t) = \begin{bmatrix} \underline{u}_{m_1/c}(t) \psi_{m_1,1}(t) \mathcal{X}_{m_1} \\ q_{m/c}(t) \end{bmatrix}$$

and

$$\phi_m(t) = \begin{bmatrix} \Psi_{\underline{v}_m}^+(t) \Xi_{\underline{v}_m}(t) \mathcal{X}_{m_1} + \Psi_{\underline{v}_m}^+(t) P_{\underline{v}_m}(t) \\ 2R_{m/c}(t) B^T(q_{m/c}(t)) \frac{d}{dt}(q_{m/c}(t)) + \underline{\omega}_c(t) \end{bmatrix}.$$

After (5-24) begins saving data, the observer update laws use $\eta_m(t)$, $\phi_m(t)$, $\Sigma_{\mathcal{U}_m}$, and $\Sigma_{\mathcal{Y}_m}$ as

$$\frac{d}{dt}(\hat{\eta}_m(t)) \triangleq \text{proj} \left(\Phi_m(q_{m/c}(t)) \phi_m(t) + \begin{bmatrix} -\underline{v}_c(t) - \underline{\omega}_c^\times(t) \underline{p}_{m_1/c}(t) \\ -\frac{1}{2}B(q_{m/c}(t)) R_{m/c}^T(t) \underline{\omega}_c(t) \end{bmatrix} + K_{\eta_m} \tilde{\eta}_m(t) \right), \quad (5-34)$$

$$\frac{d}{dt}(\hat{\phi}_m(t)) \triangleq \text{proj} \left(\hat{W}_m^T \sigma_m(\eta_m(t), \eta_c(t), \phi_m(t)) + K_{\phi_m} \tilde{\phi}_m(t) \right), \quad (5-35)$$

$$\frac{d}{dt} \left(\hat{W}_m(t) \right) \triangleq \text{proj} \left(\Gamma_m \sigma_m \left(\eta_m(t), \eta_c(t), \phi_m(t) \right) \tilde{\phi}_m^T(t) + \Gamma_m K_{W_m} \left(\Sigma_{\mathcal{U}_m} - \Sigma_{\mathcal{Y}_m} \hat{W}_m(t) \right) \right), \quad (5-36)$$

where $K_{\eta_m} \in \mathbb{R}^{7 \times 7}$, $K_{\phi_m} \in \mathbb{R}^{6 \times 6}$, $\Gamma_m \in \mathbb{R}^{L \times L}$, and $K_{W_m} \in \mathbb{R}^{L \times L}$ are constant, positive definite gain matrices.

Before the learning condition is satisfied (i.e., $t < \tau_{m_1}$), when the object is not in the camera's FOV or the origin or velocity eigenvalue conditions are not satisfied (i.e., $\sigma_{\mathcal{O}}(t) = u \vee \sigma_{v_m}(t) = u \vee \sigma_{m_1}(t) = u$), the predictor update laws are designed to update the estimates as

$$\frac{d}{dt} \left(\hat{\eta}_m(t) \right) \triangleq \text{proj} \left(\Phi_m \left(\hat{q}_{m/c}(t) \right) \hat{\phi}_m(t) + \begin{bmatrix} -\underline{v}_c(t) - \underline{\omega}_c^\times(t) \hat{p}_{m_1/c}(t) \\ -\frac{1}{2} B \left(\hat{q}_{m/c}(t) \right) \hat{R}_{m/c}^T(t) \underline{\omega}_c(t) \end{bmatrix} \right), \quad (5-37)$$

$$\frac{d}{dt} \left(\hat{\phi}_m(t) \right) \triangleq \text{proj} \left(\hat{W}_m^T(t) \sigma_m \left(\hat{\eta}_m(t), \eta_c(t), \hat{\phi}_m(t) \right) \right), \quad (5-38)$$

and

$$\frac{d}{dt} \left(\hat{W}_m(t) \right) \triangleq 0_{L \times 6}, \quad (5-39)$$

where after the the learning condition is satisfied, $t \geq \tau_{m_1}$, the acceleration model weights are updated using the history stacks as

$$\frac{d}{dt} \left(\hat{W}_m(t) \right) \triangleq \Gamma_m K_{W_m} \left(\Sigma_{\mathcal{U}_m} - \Sigma_{\mathcal{Y}_m} \hat{W}_m(t) \right). \quad (5-40)$$

5.5 Object Observer and Predictor Analysis

To simplify the subsequent analysis, if $\sigma_{m_1}(t) = u$ then $\sigma_{\mathcal{O}}(t) = u$ implying $\pi_{j,m_1}^{a_i} = \pi_j^a$ and $\pi_{j,m_1}^u = \pi_j^u$. Let a Lyapunov candidate function, $V_m(Z_m(t)) : \mathbb{R}^{7+6+6L} \rightarrow \mathbb{R}$, be defined as

$$V_m(Z_m(t)) \triangleq \frac{1}{2} \tilde{\eta}_m^T(t) \tilde{\eta}_m(t) + \frac{1}{2} \tilde{\phi}_m^T(t) \tilde{\phi}_m(t) + \frac{1}{2} \text{tr} \left(\tilde{W}_m^T(t) \Gamma_m^{-1} \tilde{W}_m(t) \right) \quad (5-41)$$

where

$$\frac{1}{2} \min \{1, \lambda_{\min} \{ \Gamma_m^{-1} \} \} \|Z_m(t)\|^2 \leq V_m(Z_m(t)) \leq \frac{1}{2} \max \{1, \lambda_{\max} \{ \Gamma_m^{-1} \} \} \|Z_m(t)\|^2$$

and $Z_m(t) \in \mathbb{R}^{7+6+6L}$ is a stacked error vector defined as

$$Z_m(t) \triangleq \begin{bmatrix} \tilde{\eta}_m^T(t) & \tilde{\phi}_m^T(t) & \text{vec} \left(\tilde{W}_m(t) \right)^T \end{bmatrix}^T.$$

Lemma 5.1. *The observer designs in (5–31)-(5–33) ensure the stacked error in $Z_m(t)$ is exponentially bounded while feedback from the object is available (i.e., object is in the camera’s FOV), the motion eigenvalue conditions are satisfied, and the history stack eigenvalue conditions are unsatisfied, $\sigma_{\mathcal{O}}(t) = a \wedge \sigma_{v_m}(t) = a \wedge \sigma_{m_1}(t) = a \wedge t < \tau_{m_1} \wedge t < \tau_m$, (i.e., $t \in [\pi_j^a, \pi_j^u] \cap t < \tau_m$).*

Proof. Taking the derivative of (5–41) with respect to time and substituting the error derivative in (5–28)-(5–30), the update laws in (5–31)-(5–33), and upper bounding using the bounds in (5–41) yields

$$\frac{d}{dt} (V_m(Z_m(t))) \leq c_1 V_m(Z_m(t)) + c_2,$$

which invoking the Comparison Lemma [116, Lemma 3.4] implies

$$V_m(Z_m(t)) \leq \left(V_m(Z_m(\pi_j^a)) + \frac{c_2}{c_1} \right) \exp(c_1(t - \pi_j^a)) - \frac{c_2}{c_1}, \quad (5–42)$$

where

$$c_1 \triangleq \frac{\max \left\{ \left(\frac{1}{2} + \overline{\omega}_c + \frac{1}{2} \overline{W}_m \overline{\sigma}_{m, \eta_m} \right), \left(1 + \frac{1}{2} \overline{\sigma}_m + \frac{1}{2} \overline{W}_m \overline{\sigma}_{m, \eta_m} + \overline{W}_m \overline{\sigma}_{m, \phi_m} \right), \frac{1}{2} \overline{\sigma}_m \right\}}{\frac{1}{2} \min \{1, \lambda_{\min} \{ \Gamma_m^{-1} \} \}}$$

and

$$c_2 \triangleq \frac{1}{2} \overline{\varepsilon}_m^2.$$

□

Lemma 5.2. *The observer designs in (5–34)-(5–36) ensure the stacked error $Z_m(t)$ exponentially decays while feedback from the object is available (i.e., object is in the camera’s FOV), the motion eigenvalue conditions are satisfied, and the history stack*

eigenvalue conditions are satisfied, $\sigma_{\mathcal{O}}(t) = a \wedge \sigma_{v_m}(t) = a \wedge \sigma_{m_1}(t) = a \wedge t \geq \tau_{m_1} \wedge t \geq \tau_m$, (i.e., $t \in [\pi_j^a, \pi_j^u] \cap t \geq \tau_m$).

Proof. Taking the derivative of (5–41) with respect to time and substituting the error derivative in (5–28)-(5–30), the update laws in (5–34)-(5–36), using the eigenvalue conditions in Assumptions 5.1 and 5.6, and upper bounding using the bounds in (5–41) yields

$$\frac{d}{dt}(V_m(Z_m(t))) \leq -c_3 V_m(Z_m(t)) + c_4,$$

which invoking the Comparison Lemma [116, Lemma 3.4] implies

$$V_m(Z_m(t)) \leq \left(V_m(Z_m(\pi_j^a)) - \frac{c_4}{c_3} \right) \exp(-c_3(t - \pi_j^a)) + \frac{c_4}{c_3}, \quad (5-43)$$

where

$$c_3 \triangleq \frac{\min \{ \lambda_{\min} \{ K_{\eta_m} \}, \frac{1}{2} \lambda_{\min} \{ K_{\phi_m} \}, \frac{1}{2} \lambda_{\min} \{ K_{W_m} \} \lambda_{\tau} \}}{\frac{1}{2} \max \{ 1, \lambda_{\max} \{ \Gamma_m^{-1} \} \}}$$

and

$$c_4 \triangleq \frac{\overline{\varepsilon}_m^2}{2 \lambda_{\min} \{ K_{\phi_m} \}} + \frac{\left(\lambda_{\max} \{ K_{W_m} \} \sqrt{\lambda_{\max} \{ \Sigma_{y_m} \}} N_W \overline{\varepsilon}_m \varsigma \right)^2}{2 \lambda_{\min} \{ K_{W_m} \} \lambda_{\tau}}.$$

As described in Remark 5.4, $\overline{\varepsilon}_m$ decreases as L increases implying $\frac{c_4}{c_3}$ decreases as the number of different basis functions increases. □

Lemma 5.3. *The predictor designs in (5–37)-(5–40) ensure the stacked error $Z_m(t)$ exponentially grows while feedback from the object is unavailable (i.e., the object is outside the camera’s FOV) or the motion eigenvalue conditions are unsatisfied, ($\sigma_{\mathcal{O}}(t) = u \vee \sigma_{v_m}(t) = u \vee \sigma_{m_1}(t) = u$), (i.e., $t \in [\pi_j^u, \pi_{j+1}^a]$).*

Proof. Taking the derivative of (5–41) with respect to time and substituting the error derivative in (5–28)-(5–30), the update laws in (5–37)-(5–39), and upper bounding using the bounds in (5–41) yields

$$\frac{d}{dt}(V_m(Z_m(t))) \leq c_5 V_m(Z_m(t)) + c_2,$$

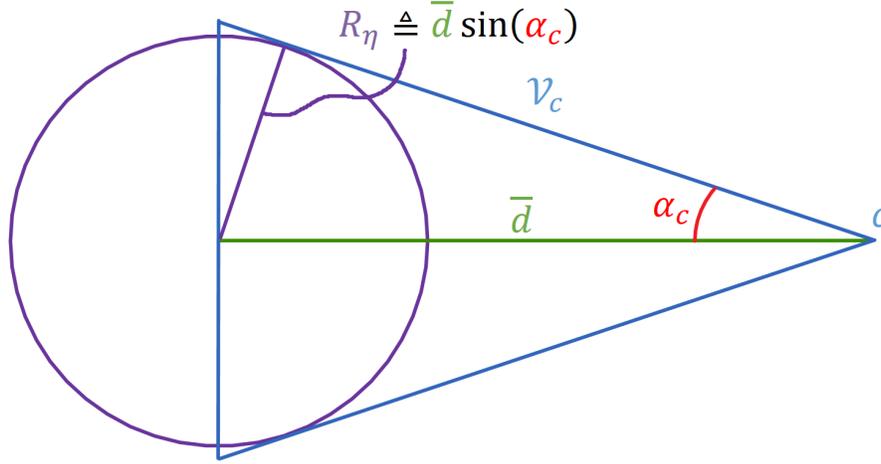


Figure 5-1. Example geometry for a simplified camera with origin at c , angle α_c , and FOV ν_c . The maximum radius of an inscribed sphere in the camera's FOV at a distance \bar{d} is $R_\eta \triangleq \bar{d} \sin(\alpha_c)$.

which invoking the Comparison Lemma [116, Lemma 3.4] implies

$$V_m(Z_m(t)) \leq \left(V_m(Z_m(\pi_j^u)) + \frac{c_2}{c_5} \right) \exp(c_5(t - \pi_j^u)) - \frac{c_2}{c_5}, \quad (5-44)$$

where $c_5 \triangleq \frac{\max\{c_{5,\eta_m}, c_{5,\phi_m}, \frac{1}{2}\sigma_m\}}{\frac{1}{2} \min\{1, \lambda_{\min}\{\Gamma_m^{-1}\}\}}$, $c_{5,\eta_m} \triangleq \left(\overline{\phi_m \Phi_{m,q_m/c}} + \frac{1}{2} + \underline{\omega}_c + \underline{\omega}_c \overline{\Phi_{m,q_m/c}} + \frac{1}{2} \overline{W_m \sigma_{m,\eta_m}} \right)$, and $c_{5,\phi_m} \triangleq \left(1 + \frac{1}{2}\sigma_m + \frac{1}{2} \overline{W_m \sigma_{m,\eta_m}} + \overline{W_m \sigma_{m,\phi_m}} \right)$. \square

5.6 Object Dwell-Time Analysis

As shown in (5-42) and (5-44), the observer designs before Assumptions 5.1 and 5.6 and the predictor designs are always exponentially growing. Given the objective is to track the moving object, dwell-times must be developed to ensure the $Z_m(t)$ remains bounded during periods of time where $Z_m(t)$ grows. The analysis in [38] assumes the structure of the object is known and measurable and shows that an estimator design (i.e., switching between an observer and predictor) is stable provided dwell-times can upper bound the total time spent in the unstable periods over a constant number of cycles; however, growth of the estimation error beyond some threshold is not always possible. Specifically, as described in [63], feedback regions are finite in size and constrained by sensor modality (e.g., the size of a camera's FOV or regions where a

positioning system is accurate). Furthermore, the object's structure is unknown while $t < \tau_m$ implying the assumptions required in [38] are unsatisfied.

For this image-based target tracking objective, the feedback region is defined by the camera's FOV, \mathcal{V}_c . As shown in the simplified camera model in Figure 5-1, the largest inscribed sphere that can fit within the camera's FOV is defined by $R_\eta \triangleq \bar{d} \sin(\alpha_c)$, where \bar{d} is the maximum distance a camera can reasonably estimate and $\alpha_c \in \mathbb{R}_{>0}$ is the minimum angle of the camera's FOV \mathcal{V}_c . Specifically, the set of features on the object must be captured within the camera's FOV for feedback to be available, $\sigma_{\mathcal{O}}(t) = a$. This requirement is covered by Assumption 2.6 where the set of features on the target, \mathcal{O}_m , are assumed to fit within the camera's FOV, $\mathcal{M} \subset \mathcal{V}_c$. Let $D_{\mathcal{M}} \triangleq \max_{p_{m_i/m_1} \in \mathcal{M}} \{m_i \in \mathcal{O}_m\}_{i=2}^n$ represent the maximum distance between the origin of the object, m_1 , and another feature, m_i , and $\overline{D_{\mathcal{M}}} \in \mathbb{R}_{>R_{\mathcal{M}}}$ represent a known bound on $D_{\mathcal{M}}$. To ensure the estimation error remains bounded, the maximum value for (5-41) must be bounded as

$$V_m(Z_m(t)) < \frac{1}{2} \tilde{\eta}_m^2, \quad (5-45)$$

where $\tilde{\eta}_m \triangleq R_\eta - \overline{D_{\mathcal{M}}}$ is the maximum error for the camera's FOV, $R_\eta > \overline{D_{\mathcal{M}}}$, and $\|\tilde{\eta}_m(t)\| \leq \tilde{\eta}_m$.

Ensuring the maximum time where feedback is unavailable (i.e., maximum dwell-time) also guarantees the camera will have feedback and the estimator will remain stable when Assumptions 5.1 and 5.6 are satisfied. This implies that an initial minimum dwell-time must reflect the learning objective, specifically, Assumptions 5.1 and 5.6 are satisfied after $t \geq \tau_{m_1} \cap t \geq \tau_m$ (i.e., the times the history stacks for the initial structure of the origin and acceleration model have sufficient data). By design, $\tau_m > \tau_{m_1}$ implying the initial minimum dwell-time must initially exceed τ_m otherwise it is not possible to guarantee the object is captured within the camera's FOV. Specifically, the maximum amount of time it can take to learn, $\underline{\Delta t}_m \in \mathbb{R}_{>0}$ must be greater than the finite excitation

condition in Assumptions 5.1 and 5.6 implying

$$\underline{\Delta t}_m > \tau_m. \quad (5-46)$$

If this is not true, then it is not possible to ensure stability using the proposed observer and predictor design.

For the subsequent development, let $\Delta\pi_j^a \triangleq \pi_j^u - \pi_j^a$ and $\Delta\pi_j^u \triangleq \pi_{j+1}^a - \pi_j^u$ represent the time spent with feedback available and unavailable over the j th cycle.

Theorem 5.1. *The switched system defined by the switching signals $\sigma_{\mathcal{O}}(t)$, $\sigma_{v_m}(t)$, and $\sigma_{m_1}(t)$, and the bounds in (5-42) and (5-44) ensure the estimation error Z_m remains bounded while $t < \tau_m$ provided*

$$\tau_m < \min \left\{ \frac{1}{c_1} \ln \left(\frac{\frac{1}{2}\tilde{\eta}_m^{-2} + \frac{c_2}{c_1}}{V_m(Z_m(\pi_1^a)) + \frac{c_2}{c_1}} \right), \frac{1}{c_5} \ln \left(\frac{\frac{1}{2}\tilde{\eta}_m^{-2} + \frac{c_2}{c_5}}{V_m(Z_m(\pi_1^a)) + \frac{c_2}{c_5}} \right) \right\}. \quad (5-47)$$

Proof. Using (5-45), to ensure stability of the system, $V_m(Z_m(\tau_m)) < \frac{1}{2}\tilde{\eta}_m^{-2}$. Considering the bounds in (5-42) and (5-44), the initial minimum dwell-time must satisfy the minimum of

$$\left(V_m(Z_m(\pi_1^a)) + \frac{c_2}{c_1} \right) \exp(c_1 \underline{\Delta t}_m) - \frac{c_2}{c_1} \leq \frac{1}{2}\tilde{\eta}_m^{-2} \quad (5-48)$$

and

$$\left(V_m(Z_m(\pi_1^a)) + \frac{c_2}{c_5} \right) \exp(c_5 \underline{\Delta t}_m) - \frac{c_2}{c_5} \leq \frac{1}{2}\tilde{\eta}_m^{-2}. \quad (5-49)$$

Solving (5-48) and (5-49) for $\underline{\Delta t}_m$ and substituting (5-46) yields (5-47). \square

Remark 5.8. The learning condition in (5-47) requires reasonable initial values to be known for the estimates; however, this is a general requirement to guarantee the motion model is learned when a target intermittently leaves the camera's FOV. Furthermore, reasonable initial values for the estimates are often available.

Theorem 5.2. *The switched system defined by the switching signals $\sigma_{\mathcal{O}}(t)$, $\sigma_{v_m}(t)$, and $\sigma_{m_1}(t)$, and the bounds in (5-43) and (5-44) ensure the estimation error Z_m remains*

GUUB while $t \geq \tau_m$ provided the j th cycle always satisfies the loss of feedback dwell-time condition

$$\Delta\pi_j^u \leq \frac{1}{c_5} \ln \left(\frac{\frac{1}{2}\tilde{\eta}_m^{-2} + \frac{c_2}{c_5}}{\left(\frac{1}{2}\tilde{\eta}_m^{-2} - \frac{c_4}{c_3}\right) \exp(-c_3\Delta\pi_j^a) + \frac{c_4}{c_3} + \frac{c_2}{c_5}} \right). \quad (5-50)$$

Proof. Using (5-45), to ensure stability of the system, $V_m(Z_m(\pi_j^a)) < \frac{1}{2}\tilde{\eta}_m^{-2}$. Considering the observer bound in (5-43), the worst case for each cycle j is the estimation error growing to the maximum during the previous cycle implying when $t = \pi_j^u$,

$$V_m(Z_m(\pi_j^u)) \leq \left(\frac{1}{2}\tilde{\eta}_m^{-2} - \frac{c_4}{c_3}\right) \exp(-c_3\Delta\pi_j^a) + \frac{c_4}{c_3}. \quad (5-51)$$

Using the worst case, $V_m(Z_m(\pi_{j+1}^a)) < \frac{1}{2}\tilde{\eta}_m^{-2}$ for the predictor bound in (5-44) and solving for the $\Delta\pi_j^u$ yields

$$\Delta\pi_j^u \leq \frac{1}{c_5} \ln \left(\frac{\frac{1}{2}\tilde{\eta}_m^{-2} + \frac{c_2}{c_5}}{V_m(Z_m(\pi_j^u)) + \frac{c_2}{c_5}} \right). \quad (5-52)$$

Substituting (5-51) into (5-50) yields the loss of feedback dwell-time condition in (5-50). □

Remark 5.9. A minimum feedback dwell-time condition is not developed here given the objective is to always track the target.

5.7 Summary

In this chapter, a novel approach to estimating the pose, velocity, and acceleration of a target is developed while considering intermittent feedback. This approach utilizes a new approach to image geometry that relaxes the requirement to have continuous observation of the target, to know structure, velocity, or acceleration of the target, and does not require the persistence of excitation assumption or positive depth constraint.

CHAPTER 6 CONCLUSIONS

In applications where agents are required to track a moving target through uncertain environments, it is necessary to estimate the structure of local features in the environment (e.g., relative positions of objects in the immediate surrounding environment), the pose of an agent (i.e., position and orientation), and the pose and velocity of the target. Many of these applications require traveling over large distances implying the local environment for an agent is always changing introducing further difficulty. It is often only possible to intermittently sense the target (e.g., environmental obstructions or path constraints of the agent may cause occlusions of the target). A typical assumption is that global sensing is available to measure the state of an agent. However, state feedback generally requires a sensor that can relate all the states to a common coordinate system (e.g., global positioning system (GPS)). However, GPS may be unavailable (e.g., agents could operate in environments where GPS is restricted or denied). Assuming that the entire environment is known and state information from the target is available is a restrictive assumption since targets are not likely to communicate such information and directly sensing the pose and velocity of a target is challenging and not possible in many scenarios. These challenges motivate the development of techniques that rely on local sensing but still allow agents to estimate their own state (i.e., pose) as well as the state of a target (i.e., pose and velocity). Additionally, efforts are motivated by the fact that local sensing often has intermittent availability.

In this dissertation, cameras are proven to be a sensor that can provide local feedback of the environment where coordinates of the target can be related to a common reference frame. Numerous estimators are developed that enable a monocular camera system to estimate the state of an agent and target despite not having the ability to inherently measure scale, have a limited FOV, and being susceptible to intermittent sensing (e.g., due to occlusions). Specifically, novel estimators using a single camera

and SfM theory are developed to estimate the Euclidean distance to features on a stationary objects and the Euclidean trajectory the camera takes while tracking the target. These estimators are extended to develop a novel estimator that used a single camera and SaMfM theory to estimate the pose of the target relative to the agent and the velocity of the target. Unlike previous results that estimate the inverse depth to features, the developed observers do not require the positive depth constraint, allowing for more general trajectories to be taken by an agent. In Chapter 1, the target tracking problem is discussed and a survey of previous work on using camera systems was presented. In Chapter 2, the dynamics for a moving monocular camera tracking stationary features and a moving target's features is developed. The dynamics present a unique approach to the SfM and SaMfM where relationships are developed showing how the Euclidean distance to stationary features relates to the pose of the agent and the Euclidean distance to moving features relates to the pose and velocity of the moving target.

In Chapter 3, a global exponentially stable observer for feature scale is developed under a finite excitation condition through the use of ICL. Since the observer only requires finite excitation to be globally exponentially stable, the observer is more general than previous results. The result indicates that the Euclidean distance to a set of features on a stationary object and the path the camera travels while viewing that object are estimated exponentially fast implying the structure (i.e., Euclidean coordinates of the tracked features) and path are reconstructed exponentially. Furthermore, the developed estimation method does not require the features on the objects to be planar and does not require the positive depth constraint. An experimental study is presented which compares the developed Euclidean distance observer to previous observers demonstrating the effectiveness of this result.

In Chapter 4, an extension to the learning approaches in Chapter 3 is developed that applies a new learning strategy that maintains a continuous estimate of the position

of the camera and estimates the structure of features as they come into the FOV. Furthermore, 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. Additionally, this approach shows how the extended observer removes the positive depth constraint required by all previous SfM approaches. Using this approach, a camera may travel over large distances without keeping specific features in the FOV for all time and allow objects to permanently leave the FOV if necessary. A Lyapunov based stability analysis proves that the observers for estimating the path of the camera as well as the structure of each set of objects are globally exponentially stable while features are in the FOV. A switched systems analysis is used to develop dwell-time conditions to indicate how long a feature must be tracked to ensure the distance estimation error is below a threshold. After the distance estimates have converged below the threshold, the feature may be used to update the position of the camera. If a feature does not satisfy the dwell-time condition, it is never used to update the position of the agent. Furthermore, the approach does not require a new set of features to be in the camera's FOV when older features leave the camera's FOV. Finally, if a recognized landmark enters the camera's FOV, the feedback is used to compensate for drift error.

In Chapter 5, the approach in Chapter 4 is used to provide pose estimates of the camera and an extension of Chapter 3 is developed to exponentially estimate the pose and velocity of the moving target. Specifically, using the pose and velocity of the camera, the estimation error of the Euclidean trajectory of the target as well as the structure of the target, is globally exponentially convergent to an ultimate bound assuming the target velocity and acceleration are bounded and dwell-time conditions are satisfied. The developed estimator relaxes the requirement to have continuous observation of the target, to know structure or velocity of the target, and does not require the persistence of excitation assumption or positive depth constraint.

The monocular camera estimators developed in this dissertation only consider a single agent tracking a single target and assume that a controller exists to satisfy the tracking objective. Future work may include incorporating the estimators in this dissertation into a guidance and control framework that informs the agent about potential trajectory issues (e.g., a building may block the path or occlude the target) and estimate the optimal trajectory to track that enables learning the target structure. Additionally, this work could be the foundation for a cooperative network target tracking system where a multiple agents are tracking multiple targets. Future work can also focus on extending this result to consider disturbances in the dynamics and developing a bundle adjustment strategy that is proven to be stable using a Lyapunov-based analysis enabling improved estimates of the path through the feedback-denied region without sacrificing stability guarantees. The extended result would be the foundation of a novel simultaneous localization and mapping algorithm that improves the observer and predictor strategy while operating in feedback-denied environments and ensures stability.

REFERENCES

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] J. Sola, A. Monin, M. Devy, and T. Vidal-Calleja, "Fusing monocular information in multicamera SLAM," *IEEE Trans. Robot.*, vol. 24, no. 5, pp. 958–968, Oct. 2008.
- [7] S. Y. Chen, "Kalman filter for robot vision: A survey," *IEEE Trans. Ind. Electron.*, vol. 59, no. 11, pp. 4409–4420, Aug. 2011.
- [8] G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis, "A quadratic-complexity observability-constrained unscented Kalman filter for SLAM," *IEEE Trans. Robot.*, vol. 29, no. 5, pp. 1226–1243, Oct. 2013.
- [9] J. Engel, T. Schöps, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," in *Computer Vision – ECCV 2014*, 2014, pp. 834–849.
- [10] 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.
- [11] 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.
- [12] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, Oct 2017.
- [13] 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.

- [14] 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.
- [15] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.
- [16] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry, *An Invitation to 3-D Vision*. Springer, 2004.
- [17] 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.
- [18] 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.
- [19] S. Soatto, R. Frezza, and P. Perona, "Motion estimation via dynamic vision," *IEEE Trans. Autom. Control*, vol. 41, no. 3, pp. 393–413, 1996.
- [20] H. Kano, B. K. Ghosh, and H. Kanai, "Single camera based motion and shape estimation using extended Kalman filtering," *Math. Comput. Modell.*, vol. 34, pp. 511–525, 2001.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.

- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] J. Oliensis, "A critique of structure-from-motion algorithms," *Comput. Vis. Image Understand.*, vol. 80, pp. 172–214, 2000.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] 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.

- [40] 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.
- [41] 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.
- [42] G. Chowdhary, M. Mühlegg, 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.
- [43] G. Chowdhary, T. Yucelen, M. Mühlegg, and E. N. Johnson, "Concurrent learning adaptive control of linear systems with exponentially convergent bounds," *Int. J. Adapt. Control Signal Process.*, vol. 27, no. 4, pp. 280–301, 2013.
- [44] R. Kamalapurkar, P. Walters, and W. E. Dixon, "Model-based reinforcement learning for approximate optimal regulation," *Automatica*, vol. 64, pp. 94–104, 2016.
- [45] 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.
- [46] 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.
- [47] 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.
- [48] 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.
- [49] Z. Bell, J. Nezvadovitz, 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.
- [50] S. Hutchinson, G. Hager, and P. Corke, "A tutorial on visual servo control," *IEEE Trans. Robot. Autom.*, vol. 12, no. 5, pp. 651–670, Oct. 1996.
- [51] N. R. Gans, P. I. Corke, and S. A. Hutchinson, "Performance tests partitioned approaches to visual servo control," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2002.
- [52] J. Chen, D. M. Dawson, W. E. Dixon, and A. Behal, "Adaptive homography-based visual servo tracking for fixed and camera-in-hand configurations," *IEEE Trans. Control Syst. Technol.*, vol. 13, pp. 814–825, 2005.

- [53] J. Chen, D. M. Dawson, W. E. Dixon, and V. Chitrakaran, "Navigation function based visual servo control," *Automatica*, vol. 43, pp. 1165–1177, 2007.
- [54] N. R. Gans and S. A. Hutchinson, "A stable vision-based control scheme for nonholonomic vehicles to keep a landmark in the field of view," in *Proc. IEEE Int. Conf. Robot. Autom.*, Roma, Italy, Apr. 2007, pp. 2196–2201.
- [55] N. R. Gans, G. Hu, and W. E. Dixon, "Keeping multiple objects in the field of view of a single ptz camera," in *Proc. Am. Control Conf.*, St. Louis, Missouri, Jun. 2009, pp. 5259–5264.
- [56] G. Hu, W. Mackunis, N. Gans, W. E. Dixon, J. Chen, A. Behal, and D. Dawson, "Homography-based visual servo control with imperfect camera calibration," *IEEE Trans. Autom. Control*, vol. 54, no. 6, pp. 1318–1324, 2009.
- [57] G. Hu, N. Gans, N. Fitz-Coy, and W. E. Dixon, "Adaptive homography-based visual servo tracking control via a quaternion formulation," *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 1, pp. 128–135, 2010.
- [58] G. Hu, N. Gans, and W. E. Dixon, "Quaternion-based visual servo control in the presence of camera calibration error," *Int. J. Robust Nonlinear Control*, vol. 20, no. 5, pp. 489–503, 2010.
- [59] G. Lopez-Nicolas, N. R. Gans, S. Bhattacharya, C. Sagues, J. J. Guerrero, and S. Hutchinson, "Homography-based control scheme for mobile robots with nonholonomic and field-of-view constraints," *IEEE Trans. Syst. Man Cybern.*, vol. 40, no. 4, pp. 1115–1127, Aug. 2010.
- [60] N. Gans, G. Hu, J. Shen, Y. Zhang, and W. E. Dixon, "Adaptive visual servo control to simultaneously stabilize image and pose error," *Mechatron.*, vol. 22, no. 4, pp. 410–422, 2012.
- [61] G. Palmieri, M. Palpacelli, M. Battistelli, and M. Callegari, "A comparison between position-based and image-based dynamic visual servoings in the control of a translating parallel manipulator," *J. Robot.*, vol. 2012, 2012.
- [62] 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.
- [63] 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.
- [64] K. Granstrom and U. Orguner, "A PHD filter for tracking multiple extended targets using random matrices," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5657–5671, Nov. 2012.

- [65] K. Wyffels and M. Campbell, “Negative observations for multiple hypothesis tracking of dynamic extended objects,” in *Proc. Am. Control Conf.*, 2014.
- [66] —, “Negative information for occlusion reasoning in dynamic extended multiobject tracking,” *IEEE Trans. Robot.*, vol. 31, no. 2, pp. 425–442, Apr. 2015.
- [67] A. Andriyenko, S. Roth, and K. Schindler, “An analytical formulation of global occlusion reasoning for multi-target tracking,” in *Proc. IEEE Int. Conf. Comput. Vis. Workshop*, 2011, pp. 1839–1846.
- [68] M. K. C. Tay and C. Laugier, *Field and Service Robotics*, ser. Springer Tracts in Advanced Robotics. Springer Berlin Heidelberg, 2008, vol. 42, ch. Modelling Smooth Paths Using Gaussian Processes, pp. 381–390.
- [69] J. Ko and D. Fox, “Gp-bayesfilters: Bayesian filtering using gaussian process prediction and observation models,” *Auton. Robot.*, vol. 27, no. 1, pp. 75–90, 2009.
- [70] D. Ellis, E. Sommerlade, and I. Reid, “Modelling pedestrian trajectory patterns with gaussian processes,” in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2009, pp. 1229–1234.
- [71] W. Lu, G. Zhang, S. Ferrari, and I. Palunko, “An information potential approach for tracking and surveilling multiple moving targets using mobile sensor agents,” in *Proc. SPIE 8045 Unmanned Syst. Technol. XIII*, 2011.
- [72] J. Joseph, F. Doshi-Velez, A. S. Huang, and N. Roy, “A bayesian nonparametric approach to modeling motion patterns,” *Auton. Robot.*, vol. 31, no. 4, pp. 383–400, 2011.
- [73] H. Wei, W. Lu, P. Zhu, S. Ferrari, R. H. Klein, S. Omidshafiei, and J. P. How, “Camera control for learning nonlinear target dynamics via bayesian nonparametric dirichlet-process gaussian-process (dp-gp) models,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, 2014, pp. 95–102.
- [74] 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.
- [75] 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.
- [76] 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.
- [77] 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.

- [78] 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.
- [79] 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.
- [80] 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.
- [81] A. Dani, Z. Kan, N. Fischer, and W. E. Dixon, "Structure and motion estimation of a moving object using a moving camera," in *Proc. Am. Control Conf.*, Baltimore, MD, 2010, pp. 6962–6967.
- [82] R. Horaud, B. Conio, O. Le Boulleux, and B. Lacolle, "An analytic solution for the perspective 4-point problem," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 1989, pp. 500–507.
- [83] D. DeMenthon and L. S. Davis, "Exact and approximate solutions of the perspective-three-point problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 11, pp. 1100–1105, 1992.
- [84] 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.
- [85] D. DeMenthon and L. Davis, "Model-based object pose in 25 lines of code," *Int. J. Comput. Vision*, vol. 15, pp. 123–141, 1995.
- [86] T. Q. Phong, R. Horaud, A. Yassine, and P. D. Tao, "Object pose from 2-d to 3-d point and line correspondences," *Int. J. Comput. Vision*, vol. 15, no. 3, pp. 225–243, 1995.
- [87] 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.
- [88] 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.
- [89] N. R. Gans, A. Dani, and W. E. Dixon, "Visual servoing to an arbitrary pose with respect to an object given a single known length," in *Proc. Am. Control Conf.*, Seattle, WA, USA, Jun. 2008, pp. 1261–1267.
- [90] S. Avidan and A. Shashua, "Trajectory triangulation: 3D reconstruction of moving points from a monocular image sequence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 4, pp. 348–357, Apr. 2000.

- [91] J. Kaminski and M. Teicher, "A general framework for trajectory triangulation," *J. Math. Imag. Vis.*, vol. 21, no. 1, pp. 27–41, 2004.
- [92] A. Bartoli and P. Sturm, "Structure-from-motion using lines: Representation, triangulation, and bundle adjustment," *Comput. Vis. Image. Understand.*, vol. 100, no. 3, pp. 416–441, 2005.
- [93] A. Bartoli, "The geometry of dynamic scenes – on coplanar and convergent linear motions embedded in 3D static scenes," *Comput. Vis. Image. Understand.*, vol. 98, no. 2, pp. 223–238, 2005.
- [94] V. Chitrakaran, D. M. Dawson, W. E. Dixon, and J. Chen, "Identification of a moving object's velocity with a fixed camera," *Automatica*, vol. 41, pp. 553–562, 2005.
- [95] M. Fujita, H. Kawai, and M. W. Spong, "Passivity-based dynamic visual feedback control for three-dimensional target tracking: Stability and l_2 -gain performance analysis," *IEEE Transactions on Control Systems Technology*, vol. 15, no. 1, pp. 40–52, 2007.
- [96] T. Ibuki, T. Hatanaka, and M. Fujita, "Passivity-based visual pose regulation for a moving target object in three dimensions: Structure design and convergence analysis," in *Proc. IEEE Conf. Decis. Control*. IEEE, 2012, pp. 5655–5660.
- [97] A. Dani, Z. Kan, N. Fischer, and W. E. Dixon, "Structure estimation of a moving object using a moving camera: An unknown input observer approach," in *Proc. IEEE Conf. Decis. Control*, Orlando, FL, 2011, pp. 5005–5012.
- [98] S. Jang, A. Dani, C. Crane, and W. E. Dixon, "Experimental results for moving object structure estimation using an unknown input observer approach," in *Proc. ASME Dyn. Syst. Control Conf.*, Fort Lauderdale, Florida, Oct. 2012.
- [99] D. Chwa, A. Dani, H. Kim, and W. E. Dixon, "Camera motion estimation for 3-d structure reconstruction of moving objects," in *Proc. of the IEEE Int. Conf. on Systems, Man, and Cybernetics*, 2012, pp. 1788–1793.
- [100] D. Chwa, A. Dani, and W. E. Dixon, "Range and motion estimation of a monocular camera using static and moving objects," *IEEE Trans. Control Syst. Tech.*, vol. 24, no. 4, pp. 1174–1183, July 2016.
- [101] Z. I. Bell, C. Harris, R. Sun, and W. E. Dixon, "Structure and velocity estimation of a moving object via synthetic persistence by a network of stationary cameras," in *Proc. IEEE Conf. Decis. Control*, Nice, Fr, Dec. 2019.
- [102] J. Lewis, "Fast template matching," *Vis. Interface*, pp. 120–123, 1995.
- [103] D. Lowe, "Distinctive image features from scale-invariant keypoints," in *IJCV*, 2004.
- [104] 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.

- [105] 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.
- [106] 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.
- [107] 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.
- [108] 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.
- [109] 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.
- [110] C. Harris and M. Stephens, “A combined corner and edge detector,” in *Alvey Vis. Conf.*, vol. 15, 1988, pp. 147–151.
- [111] J. Shi and C. Tomasi, “Good features to track,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1994, pp. 593–600.
- [112] 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.
- [113] 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.
- [114] G. Chowdhary and E. Johnson, “A singular value maximizing data recording algorithm for concurrent learning,” in *Proc. Am. Control Conf.*, 2011, pp. 3547–3552.
- [115] 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.
- [116] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2002.
- [117] G. Guennebaud, B. Jacob *et al.*, “Eigen v3,” <http://eigen.tuxfamily.org>, 2010.
- [118] G. Bradski, “The OpenCV Library,” *Dr. Dobb’s Journal of Software Tools*, 2000.

- [119] 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.
- [120] M. H. Stone, “The generalized weierstrass approximation theorem,” *Math. Mag.*, vol. 21, no. 4, pp. 167–184, 1948.

BIOGRAPHICAL SKETCH

Zachary Ian Bell received his bachelor's degree in mechanical engineering in 2015 from the University of Florida. That same year, Zachary joined the Nonlinear Controls and Robotics group under the advisement of Dr. Warren E. Dixon to pursue his doctoral studies. He received his master's degree in mechanical engineering from the University of Florida in 2017. In 2018, he received the Science, Mathematics, and Research for Transformation (SMART) Scholarship. He received his Ph.D. in mechanical engineering from the University of Florida in 2019. His research interests include image-based sensing, switched systems, network control, and Lyapunov-based nonlinear and adaptive control.