PRECISION TRACKING OF EXTENDED OBJECTS VIA NON-TRADITIONAL SENSOR MODELS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Kevin Lee Wyffels

February 2016

© 2016 Kevin Lee Wyffels
ALL RIGHTS RESERVED

PRECISION TRACKING OF EXTENDED OBJECTS VIA
NON-TRADITIONAL SENSOR MODELS

Kevin Lee Wyffels, Ph.D.
Cornell University 2016

Inspired by human perception, novel research into the sub-domain of robotic perception, known as extended object tracking, is presented. This research is motivated by the practical challenges of deploying robot driven cars (aka robo-cars, self-driving cars, autonomous road vehicles, etc.) into the existing transportation infrastructure alongside human-controlled vehicles. In this vein, arbitrary, uncooperative extended objects maneuvering in close proximity to the ego vehicle are considered, without *a priori* knowledge of the environment (e.g. road structure, traffic laws, etc.) or the objects to be tracked (e.g. controls, intent, shape, size, type, etc.). The novel approaches presented here are primarily aimed at improving the precision and accuracy of extended object tracking by mitigating the inherent information loss associated with interpreting sensor data returned from arbitrary extended objects. Specifically, a rigorous probabilistic occlusion model is derived for identifying and fusing *negative* information inherently present in sensor data; information that is heavily leveraged by humans, but largely ignored by robots. Further, a rigorous estimation framework is proposed for estimating the detailed shape of extended objects jointly with their kinematics; detailed object shape estimates enable explicit sensor models, which, in turn, engender improved tracking precision. Additionally, a hierarchical tracking framework is proposed for dynamically allocating computational resources in proportion to an objects continuously changing relevance

to the ego robot task, as defined by novel probabilistic object relevancy metrics. As a result, at any given time, objects considered to be of central importance to the ego robot task are tracked with relatively complex, high-precision methods, while objects deemed merely peripheral to the ego robot task are tracked with relatively low complexity, low precision methods; a direct analog to the human concepts of *attention* and *focus* in allocating finite cognitive resources. The proposed work is evaluated via simulation and experiments involving Cornell University's autonomous Chevy Tahoe, *Skynet*, demonstrating that non-traditional sensor models are necessary for high precision tracking of arbitrary extended objects, and that humans are worth emulating en route to this goal.

BIOGRAPHICAL SKETCH

Kevin Wyffels received his M.S. in Mechanical Engineering from Cornell University in 2013, M.S. in Mechanical Engineering from the Rochester Institute of Technology in 2007, and B.S. in Mechanical Engineering from the University at Buffalo in 2005. His research interests are in Bayesian estimation, probabilistic inference, robotic perception, and autonomous systems.

This work is dedicated to my family and friends who stood by me through over a decade of passionate resistance to joining them in the *real* world.

ACKNOWLEDGEMENTS

I would like to sincerely thank the following people for helping to make this work possible:

- My advisor, Professor Mark Campbell, for his perspective, guidance, expertise, and support.
- My special committee members, Professors Hadas Kress-Gazit and Noah Snavely, for their feedback and expertise.
- Professor Nisar Ahmed and Dr. Frank Havlak for their input and insightful discussion.
- Daniel Lee for his personalized IT support.
- Campbell research group for lengthy discussion and feedback in the weekly group meeting marathons.
- The following research grant organizations:
  - National Science Foundation (NSF):
    - Cyber Physical Systems, CNS-0931686
    - Robust Intelligence, IIS-1320490
  - Army Research Office (ARO):
    - W911NF-09-1-0466

TABLE OF CONTENTS

Biographical Sketch . . . iii
Dedication . . . iv
Acknowledgements . . . v
Table of Contents . . . vi
List of Tables . . . viii
List of Figures . . . ix

1 Introduction 1

2 Negative Information for Occlusion Reasoning in Dynamic Extended Multiobject Tracking 4
2.1 Introduction . . . 4
2.2 Modeling Negative Information . . . 8
2.2.1 Full Occlusion by a Single Object . . . 9
2.2.2 Extension to Partial Occlusions . . . 10
2.2.3 Extension to Multiple Occluding Objects . . . 12
2.3 Fusion Rules . . . 14
2.3.1 Negative Measurements . . . 16
2.3.2 Track Existence Model . . . 22
2.3.3 Data Association . . . 27
2.4 Practical Considerations . . . 31
2.4.1 Ordering of Negative Measurements . . . 31
2.4.2 Computational Complexity . . . 32
2.4.3 Unbiased estimates . . . 34
2.5 Baseline Tracker For Simulation and Experiments . . . 34
2.6 Simulation Results . . . 36
2.7 Experimental results . . . 43
2.7.1 Belief Sensibility Study . . . 44
2.7.2 Track Maintenance Study . . . 47
2.7.3 Model Accuracy Study . . . 50
2.8 Conclusion . . . 52

3 Priority-Based Tracking of Arbitrarily Shaped Extended Objects 53
3.1 Introduction . . . 53
3.2 Object Tracking Problem Formulation . . . 58
3.3 Estimation Framework . . . 60
3.4 Modal Tracking Approaches . . . 62
3.4.1 Vista Space Model . . . 64
3.4.2 Action Space Model . . . 68
3.4.3 Personal Space Model . . . 73
3.5 Mode Transitions . . . 77
3.5.1 Vista $\leftrightarrow$ Action . . . 78

3.5.2 Action ↔ Personal . . . 80
3.5.3 Invariants . . . 83
3.6 Simulation Results . . . 83
3.7 Conclusion . . . 93

4 Precision Tracking via Joint Detailed Shape Estimation of Arbitrary Extended Objects From Lidar 95
4.1 Introduction . . . 95
4.2 Modeling the Shape of Arbitrary Extended Objects . . . 100
4.3 Shape Observation . . . 101
4.4 Object Tracking Problem Formulation . . . 103
4.5 Estimating Shape Jointly with Kinematics . . . 105
4.6 Continuous Problem Partition . . . 106
4.6.1 Dynamics Prediction . . . 108
4.6.2 Measurement Update . . . 111
4.7 Discrete Problem Partition . . . 116
4.7.1 Connectivity of salient scan points . . . 117
4.7.2 Hypothesis Formation . . . 118
4.7.3 Discrete Inference . . . 121
4.8 Baseline Tracker for Simulation and Experiment . . . 122
4.9 Simulation Results . . . 123
4.10 Experimental Results . . . 127
4.11 Conclusions . . . 131

5 Conclusions 133

A Relationship between two Gaussian random variables 136

B Probability of the union of two events 137

C Rate Function Parameters 138

D Log-of-Sums trick 141
D.1 Kalman Filter . . . 142
D.1.1 Prediction . . . 142
D.1.2 Update . . . 143

Bibliography 144

LIST OF TABLES

2.1 Objects to be tracked 37
2.2 Occlusion event summary 38
2.3 Precision and Accuracy Results 40

3.1 Suggested modal tracking methods 63
3.2 Example relevance-based metrics 77
3.3 Hybrid tracking parameters 86
3.4 Efficiency vs. Precision 89

C.1 Existence Model Parameters 139

LIST OF FIGURES

2.1 Diagram of the obstacle configuration resulting in full occlusion. The sensor is mounted to the front center of the green object marked "EGO". The sensor-to-occluding object (blue) geometry gives rise to the *occlusion shadow* in red, which is of infinite range extent. 9

2.2 Diagram of the four distinct configurations resulting in occlusion. The sensor is mounted to the front center of the green object marked "EGO". The sensor-to-occluding object (blue) geometry gives rise to the *occlusion shadow* in red. All objects intersecting the occlusion shadow are considered to be occluded. 11

2.3 **Top:** Occlusion likelihood contours for an example 2D scene. **Bottom:** Two 1D cross sections through the occlusion likelihood. 15

2.4 Demonstration of the conservative nature of the occlusion likelihood. **Top:** Point clouds representing the visible surface of the mean of each object belief along with the true and approximated occlusion shadows. **Bottom:** The occlusion likelihood along cross section C computed by (2.12), compared to a Monte Carlo sampling of the true likelihood. Note that the state uncertainty of each object in the scene was reduced from that in Fig. 2.3 to sharpen the transients of the occlusion likelihood, and accentuate the modeling error. 16

2.5 Block diagram of a general multi-object tracker including three uses of negative information. White blocks with black edges are components of common tracking algorithms, for which any appropriate method of choice can be used (EKF, SPF, etc. for prediction/fusion; maximum likelihood, nearest neighbor, etc. for data association). White blocks with green edges represent common tracking components customized to leverage negative information, and solid green blocks represent novel uses of negative information proposed in this paper. 17

2.6 Demonstration of negative measurement update. The process proceeds counterclockwise from the upper left-hand quadrant. The sensor location annotated in the upper left is static throughout the process. 19

2.7 Rate parameter curve, $\mathbf{\Gamma}\left(z_n^{\text{exist}}\right)$. The vertical red lines in the decay region correspond to the left and right boundaries of the 99% of the rate parameter transient for the red α curve. The blue dashed curves show the effect of sweeping α between 0 (a horizontal line) and ∞ (a step function). $\alpha = 0$ is analogous to a traditional tracker, decaying the existence probability of all undetected obstacles at the same rate, independent from the occlusion information. The green line denotes the rate parameter for detected tracks, $D_n = 1$, which is positive for all z^{exist} to yield growth of the existence probability. 25
2.8 Demonstration of the existence probability update for detected tracks, $D_n = 1$, over a range of existence measurement, z^{exist}, and existence prior, $P(E_n)$. 26
2.9 Demonstration of the existence model for undetected tracks. **Top:** Occlusion likelihood contours over a complex 2D scene from the point of view of the indicated sensor. **Second from top:** Existence measurement for undetected tracks, (2.25), corresponding to cross section B in the top pane. **Second from bottom:** Corresponding rate parameter, (2.29), using the suggested parameters for a 15 Hz sensor given in Appendix C. **Bottom:** Resulting existence measurement likelihood, (2.28). 27
2.10 Snapshots of the simulation involving 3 rectangular dynamic objects, 3 circular static objects, and a dynamic ego sensory vehicle denoted by the yellow star. The occlusion shadow cast by each object, as seen from the sensory vehicle, is depicted with a semi-transparent shadow of the same color as the object casting the shadow. 37
2.11 Time snapshots of the baseline filter, F_A, tracking results overlayed with the true scenario as seen from a global reference frame. 41
2.12 Time snapshots of the negative information filter, F_0, tracking results overlayed with the true scenario as seen from a global reference frame. 41
2.13 Photograph of the experiment environment, annotated with the high-level scenario. The trajectories shown are synchronized such that the vehicles continue down the road in the formation implied by the figure. 44
2.14 **Left:** Time snapshots of the experimental tracking results for F_0 and F_A overlayed, as seen from the ego vehicle. **Right:** The global position of the ego vehicle for each of the time snapshots. 45

2.15 Comparison of the F_0 and F_A belief of the occluded object state at $t = 24.4$ sec, just prior to F_0 misassociating a spurious measurement to the occluded object. **Top:** Full-scale time snapshot of the F_A belief. **Middle:** Top pane zoomed to the window denoted by the dashed cyan rectangle. **Bottom:** Time snapshot of the F_0 belief, at the same scale as the middle pane. 46
2.16 Behavior of the existence model for filter F_B. **Top:** Time history of the detection flag, D_n, for each undetected track. **Middle:** Existence measurement, z^{exist}, for each undetected track. **Bottom:** Time history of the existence probability for each undetected track. 48
2.17 Behavior of the existence model for filter F_0. **Top:** Time history of the detection flag, D_n, for each undetected track. **Middle:** Existence measurement, z^{exist}, for each undetected track. **Bottom:** Time history of the existence probability for each undetected track. 49
2.18 **Top:** Time snapshot of the F_0 belief at $t = 24.4$ sec. **Middle:** Time snapshot of the F_C belief at $t = 24.4$ sec. **Bottom:** Marginal distributions over the x_{ego} dimension (major principal axis of the position uncertainty) of the normalized histogram of: the F_C particle set, the Gaussian fit to the F_C particle set, and the F_0 Gaussian. **Right:** Same as bottom plot but for y_{ego} dimension (minor principal axis of the position uncertainty). 51

3.1 Distilled spectrum of extent model complexity and tracking precision, which tend to be directly correlated. 54
3.2 Graphical representation of the Hidden Markov model (HMM) used to represent the single-object tracking problem. Shaded nodes denote observed variables, and unshaded nodes denote the hidden, i.e. latent, variables to be estimated. 59
3.3 Hybrid model of the proposed tracking framework, where $\mathbb{G}$ and $\mathbb{I}$ denote the guards and invariants governing the discrete mode transitions. 61
3.4 Qualitative abstraction of the perceptual space of the ego robot. **Left:** Eight qualitative discrete states comprised of two range sets and four bearing quadrants. **Right:** Graphical model representation of the available transitions between the qualitative states of the discrete abstraction. Dashed arrows denote transitions enabled by the discrete time nature of the filter driven by the finite temporal resolution of the sensor. 65

3.5 Demonstration of the instantaneous collision probability calculation for an object in Action space. **Top:** Belief of the elliptical object extent at time k. **Right:** Anticipated belief of the elliptical object extent at time $k + \ell$. **Bottom:** Probability of collision. Note that the ego-vehicle centroid was swept over the space to generate the probability contours for demonstration purposes, however, in practice, the collision probability only needs to be evaluated at the ego vehicle centroid labeled EGO in the figure. . 82
3.6 Simulated scenario of a star shaped object spiraling in toward the ego-vehicle and then back out over a period of 40 s; 13 time steps are shown. 84
3.7 Simulated scenario of a rectangular object passing the ego vehicle head-on in close proximity at an intersection; three time steps are shown. 84
3.8 Performance in tracking the velocity states for the star scenario: **Left:** MAP estimates overlaid on ground truth, **Right:** tracking error and 1-σ bounds. Note that there is not a concept of velocity in Vista space. 87
3.9 Performance in tracking the velocity states for the intersection scenario: **Left:** MAP estimates overlaid on ground truth, **Right:** tracking error and 1-σ bounds. Note that there is not a concept of velocity in Vista space. 88
3.10 Computational effort, ϵ, and precision, γ, as a function of time (left) and range to the closest point on the object (right) for the star scenario. 89
3.11 Computational effort, ϵ, and precision, γ, as a function of time (left) and range to the closest point on the object (right) for the intersection scenario. 90
3.12 Probabilistic object relevancy metric trajectories for the star scenario. Annotations referring to events in the ground truth scenario are provided for perspective. **Top Row:** Probability that the object is 'Far' away, governing the transitions between Vista and Action modes. **Bottom Row:** Anticipated collision probability for a $h = 1$ s time horizon, governing the transitions between Action and Personal modes. **Left Column:** Variables plotted against time. **Right Column:** Variables plotted against distance to the closest point on the object. 91

3.13 Probabilistic object relevancy metric trajectories for the intersection scenario. Annotations referring to events in the ground truth scenario are provided for perspective. **Top Row:** Probability that the object is 'Far' away, governing the transitions between Vista and Action modes. **Bottom Row:** Anticipated collision probability for a $h = 1$ s time horizon, governing the transitions between Action and Personal modes. **Left Column:** Variables plotted against time. **Right Column:** Variables plotted against distance to the closest point on the object. 92

4.1 Example 2D hull extraction of a 3D Velodyne HDL-64 scan of a 1995 Kubota L2850 tractor using a bearing window of $\Delta\beta = 0.5^\circ$. Scan points are pseudo-colored by height above the local ground model to demonstrate that scan points at varying heights contribute to the 2D hull observation. Tractor features (bucket and tires) were placed by hand to orient the viewer. 102
4.2 Block diagram of the multiple hypothesis framework for estimating the joint posterior in (4.10). 105
4.3 Graphical representation of the Hidden Markov model (HMM) used to represent the continuous partition of the PH-SET inference approach within each of the $N_{\mathrm{H},k}$ parallel filters. 107
4.4 Demonstration of the shape model growth to include previously occluded object surfaces revealed with changing viewpoint. Point correspondences are identified between subplots via matching colors and markers. Notice that only the salient scan point that does not correspond to any existing shape points (blue diamond) is added to the shape model, and the posterior connectivity is unique and immediately apparent given the connectivity of the prior and salient scan points, as well as the point correspondences between the two sets. 120
4.5 Simulated scenario in both global (top) and ego (bottom) coordinates. The blue rectangle denotes the ego (sensory) vehicle, while the green star denotes the tracked object. 124
4.6 Evolution of the MAP star shape estimate projected into a common object-fixed coordinate frame (seven time steps are shown). Each snapshot shows the shape estimate at a unique time step superimposed upon the occlusion shadow cast by the visible surface of the object (light blue region). Beside the snapshots, the full sensor trajectory as seen from the object-fixed coordinate frame is shown, and the shape estimates from each of the seven snapshots are superimposed upon the ground truth shape (green star); the sensor positions corresponding to each of the seven snapshots are denoted with a triangle marker. Sensor positions and shape estimates are pseudo-colored by their time step. 125

4.7 Performance in tracking the kinematic velocity states for both the baseline (left) and PH-SET (right) filters. The MAP estimate errors from ground truth is plotted along with the 1-σ covariance bounds estimated by each filter. 126
4.8 Five time snapshots of sensor data returned from the tractor over the 20 s scenario. Sensor frames are strung together with a dashed line denoting the approximate trajectory of the tractor. Lidar scan points are pseudo-colored by their height above the ground model. 128
4.9 Evolution of the MAP tractor shape estimate projected into a common object-fixed coordinate frame (six time steps are shown). Each snapshot shows the shape estimate at a unique time step superimposed upon the occlusion shadow cast by the visible surface of the object (light blue region). Beside the snapshots, the full sensor trajectory as seen from the object-fixed coordinate frame is shown, and the shape estimates from each of the six snapshots are superimposed upon each other; the sensor positions corresponding to each of the six snapshots are denoted with a triangle marker. Sensor positions and shape estimates are pseudo-colored by their time step. Appropriately sized tractor features are placed by hand to orient the viewer, and are labeled in the top right frame. 129
4.10 Mean and 1-σ bound of the velocity state estimate histories for both the baseline (left) and PH-SET (right) filters. 130
4.11 MAP estimate of object orientation, ϕ, (bottom) and its time derivative, $\dot{\phi}$, (top) for both the baseline and PH-SET filters, superimposed. 131

CHAPTER 1

INTRODUCTION

This dissertation presents novel research into the sub-domain of robotic perception known as Extended Object Tracking (EOT); *object tracking* refers to estimating the state (e.g. kinematics) of objects in the local environment from noisy sensor data, and *extended objects* are defined as objects of non-negligible size relative to the sensor resolution, such that they cannot be accurately modeled as a mathematical point. Extended objects further complicate the traditional tracking problem in two fundamental ways: first, they violate the foundational assumption that each object can produce, at most, a single measurement per sensor query; and second, they cast shadows in the sensor data, known as *occlusions*, which result in missing, or incomplete, measurements of tracked objects for substantial periods of time.

Bayesian inference provides a rigorous means of accounting for the many sources of uncertainty in the tracking problem, deeming it a valuable tool for EOT. Specifically, rather than estimating the latent object states directly, the object states are modeled as random variables, and their joint probability density function (pdf) is estimated, conditioned on the sensor data. High-level robot decisions can be made directly from the state pdf, which inherently reflects the uncertainty in the object states; alternatively, explicit estimates of the object states can be computed from the state pdf using existing methods, such as expected value, maximum *a posteriori* (MAP), or minimum mean squared error (MMSE) estimates.

The research presented in this dissertation is primarily focused on tracking arbitrary, uncooperative extended objects maneuvering in close proximity to the

ego robot in unknown environments; *arbitrary* refers to the most general case in which the object type, size, and shape is unknown *a priori*, and *uncooperative* refers to the general scenario in which information is not exchanged between the ego robot and the objects to be tracked (i.e. they operate independently). This problem is motivated by the recent effort to integrate fully autonomous road vehicles (aka self-driving cars, robot cars, etc.) alongside human driven vehicles in the existing transportation infrastructure; an effort spawned primarily by the Defense Advanced Research Projects Agency (DARPA) in issuing their Urban Challenge in 2007; the third in a series of robocar challenges.

The novelties presented in this dissertation are inspired by human perception, and primarily focus on the efficient extraction of sensor information (i.e. mitigate information loss) and detailed interpretation of its relationship to the object states via non-traditional sensor models. This efficient use of sensor information improves the accuracy and precision of inferred object state pdfs, thereby facilitating safety by providing lower entropy probabilistic beliefs of the local environment to dependent processes, such as behavioral planners. Specifically, Chapter 2 provides a framework for identifying and fusing negative information inherently present in sensor data to reason about occlusion. *Negative information* is defined as information inferred from expected, but missing sensor measurements; information that is heavily relied upon by humans, yet largely ignored by robots. This work is published in the IEEE Transactions on Robotics (T-Ro) [51], with preliminary work presented at the IEEE International Conference on Robotics and Automation (ICRA) in May of 2013 [48] and the American Control Conference (ACC) in June of 2014 [49]. Chapter 3 presents a hierarchical tracking framework for dynamically allocating computational resources in proportion to the time-dependent relevance of the object to the ego robot task,

where *relevance* is defined via novel probabilistic relevancy metrics. As a result, at any given time, objects considered to be of central importance to the ego robot task are tracked with relatively complex, high-precision methods, while objects deemed merely peripheral to the ego robot task are tracked with relatively low complexity, low precision methods; a direct analog to the human concepts of *attention* and *focus* in allocating finite cognitive resources. This work is currently under review for publication in the Journal of Advances in Information Fusion's (JAIF) special issue on EOT. Lastly, Chapter 4 presents a rigorous framework for estimating the detailed shape of arbitrary extended objects jointly with their kinematics. In addition to providing a richer set of information for higher level reasoning about extended objects (e.g. about object type, or occupied space), detailed shape estimates enable efficient use of sensor information by way of explicit, detailed, sensor models. This work is currently under review for publication in T-Ro, with preliminary work presented at ICRA in May of 2015 [50]. Further introductory information, including existing work in the EOT field as it pertains to each contribution, is provided in the introductory sections of each of the respective chapters.

CHAPTER 2

NEGATIVE INFORMATION FOR OCCLUSION REASONING IN DYNAMIC EXTENDED MULTIOBJECT TRACKING

2.1 Introduction

Fundamentally, accurate perception is critical for fully autonomous robots operating in uncontrolled environments. Practical deployment relies heavily on robots making safe, robust, and productive decisions in the face of uncertainty. This ability is contingent upon perception algorithms providing a practical belief of the surroundings, in which the uncertainty is properly characterized.

Full-scale deployment of autonomous robots in uncontrolled environments has been limited despite certain hardware advantages favoring robots over humans. For instance, sensors providing metric information over a wide field-of-view (FOV) are readily available for robots, while human sensors provide ordinal information, at best, over a limited FOV. Human foveal vision, in particular, has a FOV about the size of a thumbnail held at arms length, and infers distances based on relative data, such as occlusion, relative size, binocular disparities, and motion parallax [16]. Despite these hardware disadvantages, humans continue to outperform robots in perceptual tasks, suggesting human cognition is superior to analogous robotic algorithms in these areas, and potentially worth emulating.

The goal of multi-object tracking is to infer the posterior probability density of objects maneuvering in the local scene, given a history of noisy measurements. General multi-object tracking is more challenging than textbook fil-

tering, in that the object controls, number of objects, and measurement-object correspondences are unknown and changing with time. *Extended objects* refer to objects of non-negligible size relative to the sensor resolution, such that they cannot be accurately modeled as a mathematical point. Extended objects further complicate the multi-object tracking problem, in that the usual assumption that each object can produce, at most, a single measurement per sensor scan is no longer valid. Additionally, extended objects cast shadows in the sensor data, known as *occlusions*, which pose a serious challenge as they result in missing, or incomplete, measurements of tracked objects for substantial periods of time.

Traditionally, occlusions are either ignored [40], or identified but not addressed beyond the goal of track preservation [3, 20, 21, 37]. In [40], tracks undetected in sensor data are quickly deleted with confidence that new, different tracks will be birthed if the corresponding objects re-appear. Alternatively, [37] allows undetected tracks satisfying some occlusion reasoning to persist for up to 1 min in the tracker, during which time their state dynamics are propagated in the usual way. Occlusion was analytically modeled in [3] as the overlap of Gaussian position beliefs in 2-dimensional camera coordinates. This model was leveraged for tracking humans in batch processing of vision data by defining the cost penalty for an undetected track in a given camera frame as a function of the modeled track visibility. In this way, the tracking algorithm allows for occlusion by relaxing some tracking objectives with an informed approach. To a similar end, [20] presents an occlusion model as part of a variable sensor detection probability within a Probability Hypothesis Density (PHD) tracking framework, and [21] extends this model to consider estimated object extent.

Methods such as these foster incomplete, inaccurate, and imprecise beliefs

of the local scene, engendering safety concerns and hindering high-level behaviors. The work in this paper incorporates a key aspect of human perception by inferring information from expected, but missing, sensor measurements, to address the challenge of occlusion. Consistent with past work in robotics and sensor fusion, such inferences are referred to as *negative information* [29].

The concept of negative information has been successfully leveraged within the estimation framework [1, 13, 14, 26, 29, 30, 40, 41, 43, 46, 47]. Human observation has been fused as sensor data for static target search applications in [1], [13]. The complexity of human observation often requires off-line learning of intricate sensor models to capture the negative information (e.g. "observer cannot see the target" implies that the target is not within line-of-sight of the observer). The concept of negative information in [1] and [13] is similar to the work presented here, but considers only static targets in known static maps with strong correspondence assumptions. The work in this paper addresses dynamic targets in fully unknown, probabilistic environments, by leveraging negative information naturally derived from tracked random variables.

The most common use of negative information in robotics is in occupancy grid mapping, where grid cells traversed by the sensor scan short of the measured range are assumed to be unoccupied [46]. In [47], a centralized occupancy grid is generated from multiple cooperative agents, and used to correct the centralized tracking results by discarding object tracks residing in unoccupied regions of space. In [43], a centralized binary *detection map* is calculated for multiple sensors by combining line-of-sight calculations over occupancy grids computed for each sensor. This gridded binary detection map is leveraged in a lidar sensor model for a Sequential Monte Carlo pedestrian tracking filter using

Random Finite Sets.

Stanford University's submission to the DARPA Urban Challenge (DUC) used negative information to detect dark colored objects in 3-dimensional laser data [40]. Specifically, a missing laser return is interpreted as evidence of a dark colored object having absorbed the beam, and a distance measurement is inferred from the nearest ground return at the same bearing.

Negative information has also been successfully leveraged in a particle filter for localization within a known map [26] and for tracking a known number of objects within a known map [41], by discounting particles that do not predict the correct set of visible landmarks/objects detected by the sensor.

Negative information has also been used to address known sensor limitations such as finite resolution, Doppler blindness, and jamming [14, 29, 30]. Refined sensor models are developed for a variety of applications by inferring information from unique properties of each sensor.

Existing work highlights the benefits of negative information, however, outside of particle-based methods, it is a widely untapped resource in the field of dynamic extended multi-object tracking. While attractive, particle-based methods quickly become intractable for practical applications involving an unknown, potentially large map and number of objects.

Section 2.2 presents a novel occlusion likelihood model derived naturally from tracked random variables, providing the probability that each tracked object is occluded given the fully probabilistic belief of the local scene. In Section 2.3, three unique methods for considering negative information to leverage the occlusion likelihood model in extended multi-object tracking filters is pre-

sented. Specifically, it is assumed that expected, but missing, measurements imply occlusion of the undetected objects, *or* that the undetected objects do not truly exist in the local environment. This negative information is used in a filter to update the parametric probability density of undetected objects, infer the true existence of tracked objects in the local environment, and inform data association by imposing probabilistic visibility and object existence constraints on assignment candidates. Section 2.4 discusses practical considerations regarding the implementation of the methods proposed in Section 2.3. Section 2.5 presents details of the particular baseline tracking filter chosen for simulation and experimental evaluation. Lastly, simulation and experimental results are presented in Sections 2.6 and 2.7, demonstrating that fusing this occlusion information into a tracking filter improves the precision, accuracy, and sensibility of state estimates for occluded objects, as well as the overall belief of the local scene. This work builds on previous work, [48], by further developing each of the key ideas presented, providing additional implementation details, and providing a more rigorous experimental evaluation.

2.2 Modeling Negative Information

The sensor shadow cast by a tracked object, henceforth referred to as the *occlusion shadow*, is modeled by a minimal and necessary set of sufficient conditions regarding the following object metadata:

$$\begin{aligned} \theta_i &\sim \mathcal{N}\left(\bar{\theta}_i, \sigma^2_{\theta_i}\right) \quad \text{–clockwise most bearing} \\ \phi_i &\sim \mathcal{N}\left(\bar{\phi}_i, \sigma^2_{\phi_i}\right) \quad \text{–counterclockwise most bearing} \\ \rho_i &\sim \mathcal{N}\left(\bar{\rho}_i, \sigma^2_{\rho_i}\right) \quad \text{–minimum range} \end{aligned} \tag{2.1}$$

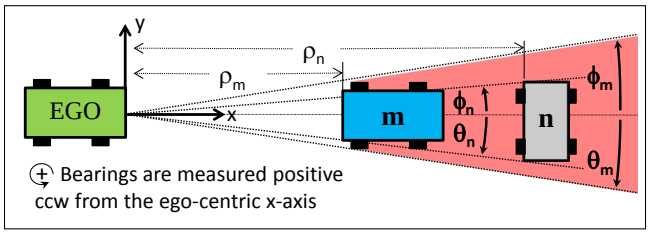

Figure 2.1: Diagram of the obstacle configuration resulting in full occlusion. The sensor is mounted to the front center of the green object marked "EGO". The sensor-to-occluding object (blue) geometry gives rise to the *occlusion shadow* in red, which is of infinite range extent.

where clockwise (cw) and counterclockwise (ccw) designations are with respect to the mean bearing to the object. The metadata in (2.1) are chosen to parameterize the occlusion model because they naturally describe the line-of-sight visibility of a point observer, with reasonable accuracy and minimal number of parameters. It is assumed that the tracked object belief maintains a notion of object extent, such that the Gaussian distribution over the metadata in (2.1) can be computed.

2.2.1 Full Occlusion by a Single Object

As depicted in Fig. 2.1, object n is *fully* occluded by object m, $O_{n,m}$, if object m exists, E_m, and the following bearing and range conditions, $B^{\{1\}}_{n,m}$ and $R_{n,m}$, are satisfied:

$$\begin{aligned} B^{\{1\}}_{n,m} &= (\theta_n \geq \theta_m) \cap (\phi_m \geq \phi_n) \\ R_{n,m} &= (\rho_n \geq \rho_m) \end{aligned} \tag{2.2}$$

Therefore, the occlusion likelihood is found as the joint probability that the necessary conditions for occlusion are satisfied. Specifically, given that the range and bearing conditions are independent, the likelihood that object n is fully oc-

cluded by object m is:

$$\begin{aligned} p\left(O_{n,m}|D_n=0\right) &= p\left(B_{n,m}^{\{1\}}, R_{n,m}, E_m\right) \\ &= p\left(B_{n,m}^{\{1\}}\middle|E_m\right)\cdot p\left(R_{n,m}\middle|E_m\right)\cdot p\left(E_m\right) \end{aligned} \tag{2.3}$$

where $D_n = 0$ is the binary observation that object n is undetected in sensor data, $p\left(E_m\right)$ is the probability that the occluding object exists, for which an inference strategy is proposed in Section 2.3.2, and the remaining two terms are the probability that the bearing and range conditions in (2.2) are satisfied, given that the occluding object exists. Note that objects detected in the sensor data, $D_n = 1$, are assumed to be fully visible to the sensor:

$$p\left(O_{n,m}|D_n=1\right) = 0 \;\forall m \tag{2.4}$$

The bearing and range probabilities in (2.3) are computed via the Gaussian relational probability provided in Appendix A, where $\mathbf{\Phi}\left(\cdot\right)$ denotes the standard normal cumulative distribution function:

$$\begin{aligned} p\left(B_{n,m}^{\{1\}}\middle|E_m\right) &= \mathbf{\Phi}\left(\frac{\theta_n-\theta_m}{\sqrt{\sigma_{\theta_n}^2+\sigma_{\theta_m}^2}}\right)\mathbf{\Phi}\left(\frac{\phi_m-\phi_n}{\sqrt{\sigma_{\phi_m}^2+\sigma_{\phi_n}^2}}\right) \\ p\left(R_{n,m}|E_m\right) &= \mathbf{\Phi}\left(\frac{\rho_n-\rho_m}{\sqrt{\sigma_{\rho_n}^2+\sigma_{\rho_m}^2}}\right) \end{aligned} \tag{2.5}$$

2.2.2 Extension to Partial Occlusions

Partially occluded objects emit incomplete measurements, which can be ambiguous to associate to the source object; therefore, it may be prudent to discard these measurements to avoid costly errors in data association. This gives rise to three additional configurations of undetected objects to consider in the occlusion likelihood function, depicted by cases 2, 3, and 4 in Fig. 2.2 (case 1 is the fully occluded case addressed in Section 2.2.1).

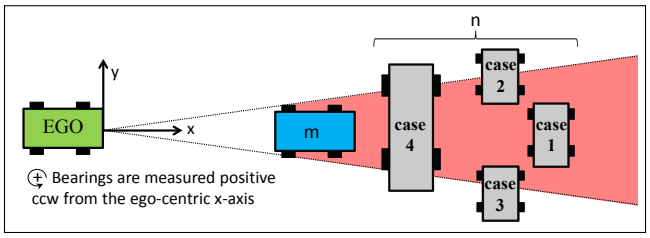

Figure 2.2: Diagram of the four distinct configurations resulting in occlusion. The sensor is mounted to the front center of the green object marked "EGO". The sensor-to-occluding object (blue) geometry gives rise to the *occlusion shadow* in red. All objects intersecting the occlusion shadow are considered to be occluded.

These configurations share the range condition from (2.2), but require the following two additional bearing conditions:

$$\begin{aligned} B_{n,m}^{\{2,4\}} &= (\phi_m \geq \theta_n) \cap (\phi_n \geq \phi_m) \\ B_{n,m}^{\{3,4\}} &= (\theta_m \geq \theta_n) \cap (\phi_n \geq \theta_m) \end{aligned} \tag{2.6}$$

where the superscripts refer to the corresponding sets of occlusion cases in Fig. 2.2 described by the bearing condition. The likelihood that object n is *at least* partially occluded by object m is found by replacing the full occlusion bearing condition in (2.3) with the union of the full occlusion bearing condition from (2.2), and two partial occlusion bearing conditions from (2.6):

$$p\left(O_{n,m}|D_n=0\right) = p\left(B_{n,m}|E_m\right) \cdot p\left(R_{n,m}|E_m\right) \cdot p\left(E_m\right) \tag{2.7}$$

where:

$$B_{n,m} = B_{n,m}^{\{1\}} \cup B_{n,m}^{\{2,4\}} \cup B_{n,m}^{\{3,4\}} \tag{2.8}$$

Given that $B_{n,m}^{\{2,4\}}$ and $B_{n,m}^{\{3,4\}}$ both include case 4, they are not mutually exclusive (two conditions to describe three configurations). However, $B_{n,m}^{\{1\}}$ is mutually exclusive from both $B_{n,m}^{\{2,4\}}$ and $B_{n,m}^{\{3,4\}}$. With this in mind, the bearing likelihood in (2.7), $p\left(B_{n,m}|E_m\right)$, is computed using the union probability rule provided in

Appendix B:

$$\begin{aligned} p\left(B_{n,m}|E_m\right) &= p\left(B_{n,m}^{\{1\}} \cup B_{n,m}^{\{2,4\}} \cup B_{n,m}^{\{3,4\}} \middle| E_m\right) \\ &= p\left(B_{n,m}^{\{1\}} \middle| E_m\right) + p\left(B_{n,m}^{\{2,4\}} \middle| E_m\right) + p\left(B_{n,m}^{\{3,4\}} \middle| E_m\right) - p\left(B_{n,m}^{\{2,4\}}, B_{n,m}^{\{3,4\}} \middle| E_m\right) \end{aligned} \tag{2.9}$$

where $p\left(B_{n,m}^{\{1\}} \middle| E_m\right)$ is defined in (2.5), and the remaining terms are computed via the Gaussian relational probability provided in Appendix A:

$$\begin{aligned} p\left(B_{n,m}^{\{2,4\}} \middle| E_m\right) &= \mathbf{\Phi}\left(\frac{\phi_m - \theta_n}{\sqrt{\sigma_{\phi_m}^2 + \sigma_{\theta_n}^2}}\right)\mathbf{\Phi}\left(\frac{\phi_n - \phi_m}{\sqrt{\sigma_{\phi_n}^2 + \sigma_{\phi_m}^2}}\right) \\ p\left(B_{n,m}^{\{3,4\}} \middle| E_m\right) &= \mathbf{\Phi}\left(\frac{\theta_m - \theta_n}{\sqrt{\sigma_{\theta_m}^2 + \sigma_{\theta_n}^2}}\right)\mathbf{\Phi}\left(\frac{\phi_n - \theta_m}{\sqrt{\sigma_{\phi_n}^2 + \sigma_{\theta_m}^2}}\right) \\ p\left(B_{n,m}^{\{2,4\}}, B_{n,m}^{\{3,4\}} \middle| E_m\right) &= \mathbf{\Phi}\left(\frac{\theta_m - \theta_n}{\sqrt{\sigma_{\theta_m}^2 + \sigma_{\theta_n}^2}}\right)\mathbf{\Phi}\left(\frac{\phi_n - \phi_m}{\sqrt{\sigma_{\phi_n}^2 + \sigma_{\phi_m}^2}}\right) \end{aligned} \tag{2.10}$$

Note that jointly satisfying (i.e. the intersection of) bearing conditions $B_{n,m}^{\{2,4\}}$ and $B_{n,m}^{\{3,4\}}$ corresponds to case 4 from Fig. 2.2.

2.2.3 Extension to Multiple Occluding Objects

The full and partial occlusion likelihoods for a single object occluder are given in (2.3) and (2.7), respectively. This concept must be extended to the general case of an arbitrary number of tracked objects, N. Specifically, the total likelihood that object n is occluded is computed as the probability that object n is occluded by *any* of the other $N-1$ tracked objects:

$$p\left(O_n|D_n\right) = p\left(\bigcup_{\substack{m \neq n \\ m = 1,}}^{N} O_{n,m} \middle| D_n\right) \tag{2.11}$$

In the event that object n is detected in the sensor data, $D_n = 1$, the probability of occlusion by each tracked object is defined in (2.4) to be 0; therefore, evaluating

the occlusion model in (2.11) for detected tracks is trivial. Conversely, in the event that object n is not detected in sensor data, $D_n = 0$, the union probability rule provided in Appendix B is applied to (2.11); note that the $N-1$ occlusion events in (2.11) are independent and not mutually exclusive, as a given object can be occluded by multiple objects simultaneously. Therefore, applying the union probability rule to (2.11) gives:

$$p\left(O_n | D_n = 0\right) = \sum_{i=1}^{N\text{-}1} \left\{ (\text{-}1)^{(i+1)} \sum_{j=1}^{{}^{(N-1)}\mathrm{C}_i} \left[\prod_{k=1}^{i} p\left(O_{n,\mathbb{C}_j^i(k)}\right) \right] \right\} \tag{2.12}$$

where $\mathbb{C}_j^i(k)$ denotes the kth object in the jth i-combination (unique set of i objects), and ${}^{(N-1)}\mathrm{C}_i$ denotes the number of i-combinations in the set of $N-1$ potential occluders:

$${}^{(N-1)}\mathrm{C}_i = \frac{(N-1)!}{i!\,(N-1-i)!} \tag{2.13}$$

Note that (2.12) is identically equal to (2.11) for independent occlusion events that are not mutually exclusive.

Fig. 2.3 demonstrates the occlusion likelihood function for undetected objects from the point of view of a sensor scanning a complex scene including rectangular objects of certain existence, $p\left(E_m\right) = 1 \;\forall\, m \in \{1, \ldots, N\}$; the top pane shows the contours that result from computing the likelihood over a 2-dimensional (2D) grid, while the bottom pane displays 1-dimensional (1D) slices through the likelihood function. These slices can be thought of as potential paths through the scene. As intuition would suggest, the occlusion likelihood approaches 1 in the occlusion shadows, and 0 in areas visible to the sensor. The likelihood does not exceed 1 when simultaneously occluded by multiple objects, as a consequence of correctly accounting for the shadow overlap in (2.12). Likewise, there are smooth transitions between the occlusion cases depicted in Fig. 2.2, indicating (2.9) is derived correctly. The smoothness of the curves and small peak on

the right-hand side of cross section B are consequences of the uncertainty in the state of each object; this peak diminishes in cross section A, as the confidence that the path passes in front of the object in the far right-hand corner increases.

Note that the occlusion conditions in (2.2) model the visible surface of each object, i, as an arc of radius ρ_i, spanning the extreme bearings $[\theta_i, \phi_i]$. The accuracy of this model depends on the true size, shape, and orientation of the objects. Fig. 2.4 demonstrates the accuracy of the occlusion likelihood model, (2.12), by comparing to *ground truth*, which is computed by sampling the scene one thousand times at each grid point along cross section C, and counting the fraction of samples that result in occlusion. This is the same scene from Fig. 2.3, but with reduced object state uncertainty to accentuate the error present in the occlusion model; cross section C is halfway between cross sections A and B. Notice that model inaccuracies result in a conservative likelihood, in that the true occlusion shadow is guaranteed to be a subset of the approximation. In this way, truly occluded space is not mistaken to be visible, and scene estimates are updated conservatively.

2.3 Fusion Rules

The negative information proposed in this work assumes that expected, but missing, measurements from tracked objects imply that the undetected objects are either occluded, or do not truly exist:

$$(D_n = 0) \rightarrow (O_n = 1) \cup (E_n = 0) \tag{2.14}$$

With this assumption, the occlusion likelihood model developed in Section 2.2 is used to supplement traditional parametric tracking filters in the following

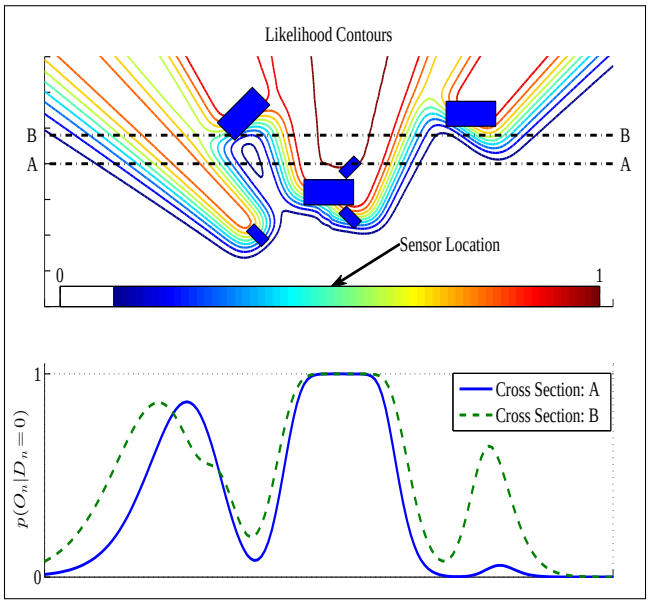

Figure 2.3: **Top:** Occlusion likelihood contours for an example 2D scene. **Bottom:** Two 1D cross sections through the occlusion likelihood.

three ways:

1. Update the pdf over undetected object states via *negative measurements*
2. Infer the true existence of tracked objects in the local scene
3. Impose probabilistic visibility and object existence constraints on the data assignment solution

Fig. 2.5 shows a block diagram of a general multi-object tracker including these three uses of negative information, the details of which are provided in the following three sections.

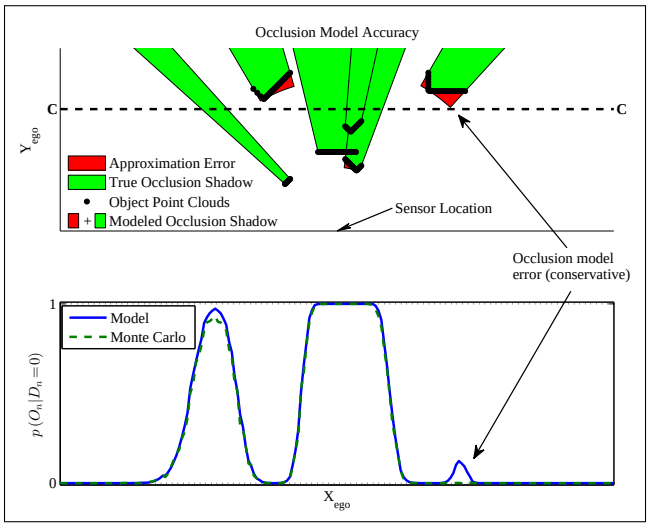

Figure 2.4: Demonstration of the conservative nature of the occlusion likelihood. **Top:** Point clouds representing the visible surface of the mean of each object belief along with the true and approximated occlusion shadows. **Bottom:** The occlusion likelihood along cross section C computed by (2.12), compared to a Monte Carlo sampling of the true likelihood. Note that the state uncertainty of each object in the scene was reduced from that in Fig. 2.3 to sharpen the transients of the occlusion likelihood, and accentuate the modeling error.

2.3.1 Negative Measurements

Negative measurements are fused for undetected object tracks to rigorously leverage the negative information that undetected objects are not likely to reside in areas of space visible to the sensor (i.e. they are likely to be occluded):

$$(D_n = 0) \rightarrow (O_n = 1) \tag{2.15}$$

Conceptually, the missed detection of an object is considered an observation of that object being occluded, and negative measurements are fused to ensure the object state belief reflects this observation. In this way, as depicted in the block diagram of Fig. 2.5, a sensor scan provides information about *all* tracked objects; detected objects, $D_n = 1$, are updated via traditional *positive* measurements, while

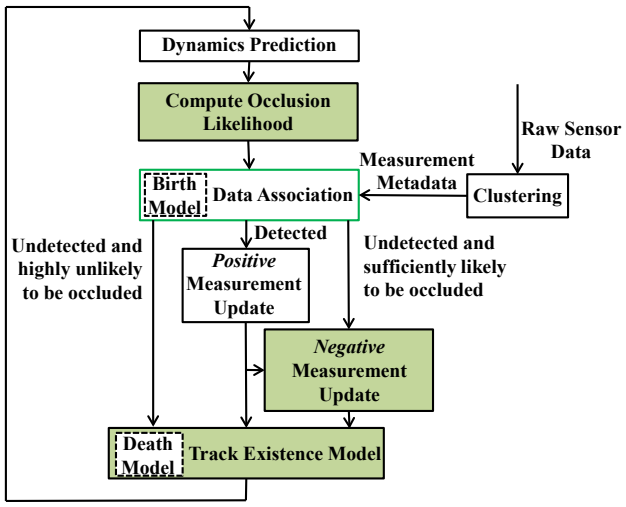

Figure 2.5: Block diagram of a general multi-object tracker including three uses of negative information. White blocks with black edges are components of common tracking algorithms, for which any appropriate method of choice can be used (EKF, SPF, etc. for prediction/fusion; maximum likelihood, nearest neighbor, etc. for data association). White blocks with green edges represent common tracking components customized to leverage negative information, and solid green blocks represent novel uses of negative information proposed in this paper.

undetected objects, $D_n = 0$, are updated via the proposed *negative* measurements.

Note that the exact negative measurement fusion process depends on the particular representation of the object state belief chosen in the tracking filter. In this section, the fusion process is derived for the case that the object state belief is represented as a Gaussian (or mixture of Gaussians), as this is a common choice, however, it is possible to derive the fusion process for other representations.

Given the negative information in (2.15), the occlusion likelihood provided in (2.12) naturally serves as the negative measurement likelihood for undetected objects. However, given the complex form of the occlusion likelihood, as demonstrated in Fig. 2.3, there lacks a closed-form fusion rule for a Gaus-

sian state belief. Therefore, unnormalized Gaussian Mixture Models (GMM) are used to approximate the negative measurement likelihood and facilitate the fusion process, as has recently been done in similar parametric Bayesian filters [1]:

$$\hat{L}(x) = \sum_{j=0}^{N_{\text{mix}}-1} w_j \cdot \exp\left(-\frac{\left(x-\bar{x}_j\right)^2}{2\sigma_j^2}\right) \tag{2.16}$$

where $(\hat{\cdot})$ denotes the Gaussian mixture representation of the underlying variable.

The bearing and range portions of the likelihood function are uncoupled, as demonstrated in (2.3) and (2.7). Therefore, the fusion process is simplified by considering the bearing and range dimensions independently, and fusing the resulting likelihoods sequentially as measurements using existing Gaussian mixture filtering methods [2, 23, 42, 45]. Note that fusing a GMM likelihood, such as (2.16), results in a mixture model posterior, where the number of mixands grows exponentially with each negative measurement update. Therefore, the negative measurement posterior must be compressed to a manageable number of mixands using existing methods, such as those presented in [2,23,42,45]. Similarly, for filters representing the object state as a single Gaussian, such as a traditional EKF or SPF, the negative measurement posterior can be approximated by a single Gaussian by matching first and second moments of the distribution.

In order to gain a conceptual understanding, Fig. 2.6 steps through the process of fusing negative measurements to update the probability density of the undetected object states. Specifically, the 'Prior' pane shows the scene from Fig. 2.3, along with the *a priori* mean and 1-sigma ellipse of the Gaussian belief of an undetected object of interest; note that the prior has significant probability mass in areas likely to be visible to the sensor. The 'Likelihood Contours' pane shows the same likelihood contours from Fig. 2.3, along with range and bear-

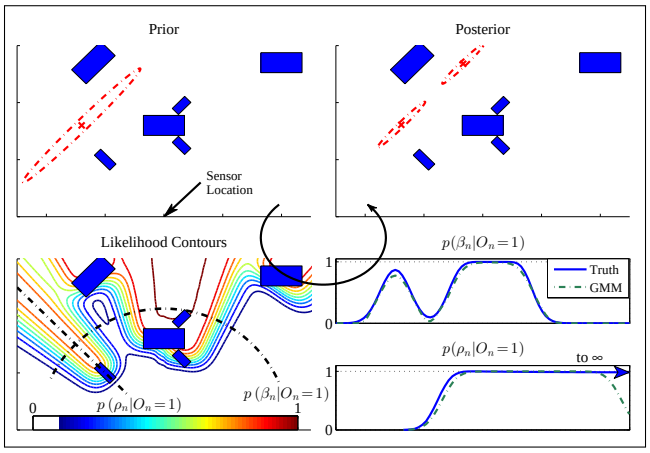

Figure 2.6: Demonstration of negative measurement update. The process proceeds counterclockwise from the upper left-hand quadrant. The sensor location annotated in the upper left is static throughout the process.

ing cross section markers; this step is included for visualization purposes, and is not computed in practice. The '$p(\beta_n|O_n=1)$' and '$p(\rho_n|O_n=1)$' panes show the corresponding 1D range and bearing likelihoods, overlayed with their unnormalized Gaussian mixture representations. Lastly, the 'Posterior' pane shows the bimodal posterior that results from fusing the negative measurements. Note that the posterior state belief is more sensible than the prior, because it is consistent with the negative information in (2.15), that undetected objects are unlikely to exist in areas of space visible to the sensor.

Further details of the negative bearing and range measurement fusion are provided the following two subsections:

Bearing Measurement

Given that the perceived size of an object does not change with bearing, a reduction in the dimension of the negative bearing measurement is possible by

fixing the bearing spanned by the object, δ_n, and considering a mean bearing measurement, β_n, as opposed to the extreme bearings, $\{\theta_n, \phi_n\}$. Specifically, the negative bearing likelihood function in (2.9) can be rewritten in terms of β_n by making the following substitutions in (2.5) and (2.10):

$$\begin{aligned} \phi_n &= \beta_n + \frac{1}{2}\delta_n \\ \theta_n &= \beta_n - \frac{1}{2}\delta_n \end{aligned} \tag{2.17}$$

where:

$$\begin{aligned} \beta_n &\sim \mathcal{N}\left(\bar{\beta}_n, \sigma^2_{\beta_n}\right) \\ &\sim \mathcal{N}\left(\frac{1}{2}\left(\bar{\phi}_n + \bar{\theta}_n\right),\ \frac{1}{4}\left(\sigma^2_{\theta_n} + \sigma^2_{\phi_n} + 2\sigma_{\theta_n,\phi_n}\right)\right) \\ \delta_n &\sim \mathcal{N}\left(\bar{\delta}_n, \sigma^2_{\delta_n}\right) \\ &\sim \mathcal{N}\left(\bar{\phi}_n - \bar{\theta}_n, \sigma^2_{\theta_n} + \sigma^2_{\phi_n} + 2\sigma_{\theta_n,\phi_n}\right) \end{aligned} \tag{2.18}$$

and $(\bar{\cdot})$ denotes the mean of the underlying variable. The resulting 1D likelihood, $p(\beta_n|\delta_n, O_n = 1)$, is modeled using the form in (2.16) by optimizing the model parameters: number of mixands, N^B_{mix}, and the weight, mean, and covariance of each mixand, $\left[w^B_j, \bar{\beta}^B_j, \sigma^B_j\right] \forall\, j \in \left\{1, \ldots, N^B_{\text{mix}}\right\}$, using existing methods such as expectation-maximization (EM):

$$p(\beta_n|\delta_n, O_n = 1) \approx \sum_{j=1}^{N^B_{\text{mix}}} w^B_j \cdot \exp\left[-\frac{1}{2}\left(\frac{\beta - \bar{\beta}^B_j}{\sigma^B_j}\right)^2\right] \tag{2.19}$$

Range Measurement

The negative range likelihood function represents the distribution over the range, ρ, to the closest point on object n given the negative information, $p(\rho_n|O_n = 1)$. This 1D likelihood can be approximated directly using existing methods such as EM. Alternatively, an accurate and much more efficient closed-form approximation can be achieved by realizing that the shape of the likeli-

hood is dominated by a single occluding object, defined here as m^*, such that:

$$p\left(\rho_n|O_n=1\right) \approx p\left(\rho_n|O_{n,m^*}=1\right) \tag{2.20}$$

where:

$$m^* = \underset{m}{\operatorname{argmax}}\left[p\left(O_{n,m}|D_n=0\right)\right]$$

In this way, m^* is the object most likely to be occluding object n, and (2.20) is the range component of the single-occluder likelihood model, provided in (2.5) as a Gaussian CDF. The task at hand is then to model a Gaussian CDF, (2.20), with an unnormalized Gaussian Mixture, (2.16), for which a closed-form solution is available. Specifically, the range likelihood is then written as:

$$p\left(\rho_n|O_n=1\right) \approx \frac{1}{\sqrt{2\pi}} \sum_{j=1}^{N_{\text{mix}}^R} \exp\left\{-\frac{1}{2}\left[\frac{\bar{\rho}_{m^*} + \left(j-\frac{1}{2}\right)\cdot\sigma^R - \rho}{\sigma^R}\right]^2\right\} \tag{2.21}$$

where:

$$\sigma^R = 1.0451\cdot\sqrt{\sigma_{\rho_n}^2 + \sigma_{\rho_{m^*}}^2} \tag{2.22}$$

The number of mixands, N_{mix}^R, is found by realizing that the posterior object state belief is negligible anywhere the prior state belief is negligible, regardless of the likelihood; therefore, an accurate model of (2.20) for ranges with negligible prior is not required. Given the range prior $\sim \mathcal{N}\left(\bar{\rho}_n, \sigma_{\rho_n}^2\right)$, and a negligible probability mass assumed for $\rho > \left(\bar{\rho}_n + 3\sigma_{\rho_n}\right)$, the optimal number of mixands is the minimum integer that satisfies $p\left(\bar{\rho}_n + 3\sigma_{\rho_n}|O_n=1\right) \approx 1$. Specifically:

$$N_{\text{mix}}^R = \left\lceil \frac{\bar{\rho}_n + 3\sigma_{\rho_n} - \bar{\rho}_{m^*}}{\sigma^R} \right\rceil + 3 \tag{2.23}$$

where the ceiling function, $\lceil\cdot\rceil$, forces an integer number of mixands while continuing to satisfy the aforementioned constraint.

2.3.2 Track Existence Model

When the true number of objects in the scene is unknown, there is the potential for spurious measurements to spawn object tracks that do not truly exist in the local scene. These *spurious tracks* waste computation and result in inaccurate beliefs of the local environment; a detriment to dependent processes, such as planning algorithms. The proposed track existence model serves as a principled method for identifying and discarding spurious tracks, while allowing genuine tracks to persist; this collective effort, coupled with track birth, is referred to as *track maintenance*.

Sensor information, positive and negative, naturally informs track existence. Specifically, in agreement with the negative information in (2.14), tracks undetected in sensor data, $D_n = 0$, with a low likelihood of occlusion, are less likely to exist than those under likely occlusion. In this way, the occlusion model provided in (2.12) can be leveraged to identify spurious tracks. Further, tracks consistently and confidently detected in sensor data, $D_n = 1$, are more likely to exist than those undetected, or assigned measurements from a relatively ambiguous data association. In this case, the *positive* information is the assumption that traditional measurements imply existence of the detected objects:

$$(D_n = 1) \rightarrow (E_n = 1) \tag{2.24}$$

To this end, the following measurement of track existence, $z_n^{\text{exist}} \in [0, 1]$, is defined:

$$z_n^{\text{exist}} = \begin{cases} p\left(a_{n,\ell^*} | z_{\ell^*}\right) & \text{if } D_n = 1 \\ p\left(O_n | D_n = 0\right) & \text{if } D_n = 0 \end{cases} \tag{2.25}$$

where $p\left(O_n | D_n = 0\right)$ is the occlusion likelihood defined in (2.12), and $p\left(a_{n,\ell^*} | z_{\ell^*}\right)$ is the posterior probability that the measurement assigned to track n, defined as

z_{ℓ^*}, actually originated from track n, as computed in the data association routine. Note that obvious data assignments for detected tracks, $D_n = 1$, and likely occlusion for undetected tracks, $D_n = 0$, result in existence measurements, $z_n^{\text{exist}} \approx 1$, corresponding to strong evidence that the tracked object exists. Conversely, ambiguous data assignments for detected tracks, and unlikely occlusion for undetected tracks, result in existence measurements, $z_n^{\text{exist}} \approx 0$.

The measurement of track existence, z_n^{exist}, is used to update the track existence probability, $p(E_n)$, using Bayesian inference:

$$p\left(E_n \middle| z_n^{\text{exist}}\right) = \frac{p\left(z_n^{\text{exist}} \middle| E_n\right) \cdot p(E_n)}{p\left(z_n^{\text{exist}}\right)} \tag{2.26}$$

where $p(E_n)$ and $p\left(E_n \middle| z_n^{\text{exist}}\right)$ are the prior and posterior existence probabilities for track n, and $p\left(z_n^{\text{exist}}\right)$ is the marginal probability of the existence measurement, defined as:

$$p\left(z_n^{\text{exist}}\right) = \begin{cases} p\left(z_n^{\text{exist}} \middle| E_n\right) \cdot p(E_n) + (1 - p(E_n)) & \text{if } D_n = 1 \\ 1 & \text{if } D_n = 0 \end{cases} \tag{2.27}$$

$p\left(z_n^{\text{exist}} \middle| E_n\right)$ is the existence measurement likelihood chosen to be an exponential distribution:

$$p\left(z_n^{\text{exist}} \middle| E_n\right) = \exp\left(\frac{1}{f_s} \mathbf{\Gamma}\left(z_n^{\text{exist}}\right)\right) \tag{2.28}$$

where f_s is sensor frequency, and $\mathbf{\Gamma}(\cdot)$ is the rate parameter function, which determines how quickly the existence probability changes with each sensor frame. The rate parameter function, $\mathbf{\Gamma}(\cdot)$, is designed such that the existence probability of undetected tracks, $D_n = 0$, decays with rate inversely proportional to their occlusion likelihood, in agreement with the negative information in (2.14); likewise, the existence probability of detected tracks, $D_n = 1$, grows with rate proportional to the confidence in the detection, in agreement with the positive

information in (2.24). Specifically,

$$\boldsymbol{\Gamma}(x) = \begin{cases} \gamma_g \cdot x & \text{if } D_n = 1 \\ \frac{\overline{\gamma} - \underline{\gamma}}{1 + \exp(\boldsymbol{f_d}(x))} - \overline{\gamma} & \text{if } D_n = 0 \end{cases} \tag{2.29}$$

where γ_g is a growth parameter, and $\underline{\gamma}$, $\overline{\gamma}$ are decay parameters described in Appendix C. The $D_n = 0$ case of (2.29) is a scaled and translated logistic function, whose argument, $\boldsymbol{f_d}(\cdot)$, is a monotonic function mapping an input on the interval $(0, 1)$ to an output on the interval $(-\infty, \infty)$. These characteristics ensure that $\boldsymbol{\Gamma}\left(z_n^{\text{exist}}\right)$ achieves its extreme values at the extreme values of z_n^{exist} for all valid choices of the model parameters:

$$\boldsymbol{f_d}(x) = \alpha \cdot \ln\left(\frac{x\left(1 - \tilde{z}^{\text{exist}}\right)}{(1 - x)\,\tilde{z}^{\text{exist}}}\right) \tag{2.30}$$

where $\tilde{z}^{\text{exist}}$ and α are additional decay parameters, described in Appendix C, that can be used to tailor the character of the track existence model.

It is important to note that any valid set of parameters in the rate parameter function (2.29-2.30) removes spurious tracks faster than genuine tracks, which is the goal of the track existence model, and a unique feature of this work; the parameters simply afford the user some control over the assertiveness of the filter in pursuing this goal. The valid ranges, suggested values, and intuition for these model parameters are provided in Appendix C. In this work, a track is removed from the filter when its existence probability drops below 13.53% (two equivalent decay time constants).

Fig. 2.7 demonstrates the rate parameter function, $\boldsymbol{\Gamma}(\cdot)$, over the full range of z^{exist}. A family of curves are shown for the undetected case, $D_n = 0$, to demonstrate the influence of α on the shape of the curve, as it is the least intuitive model parameter. Further, $\underline{z}_{0.99}^{\text{exist}}$ and $\overline{z}_{0.99}^{\text{exist}}$ denote the left and right boundaries of

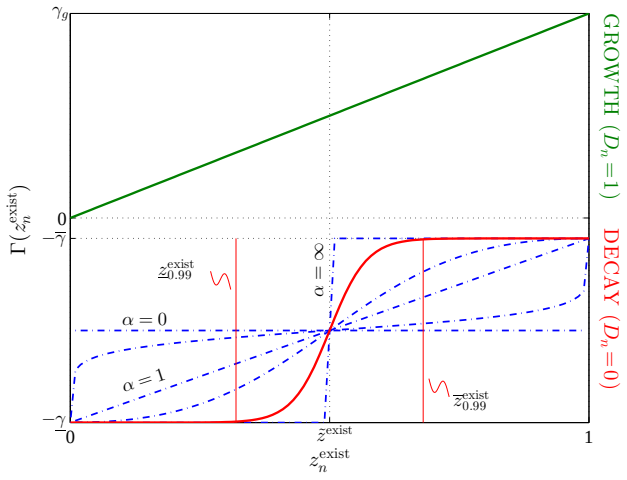

Figure 2.7: Rate parameter curve, $\mathbf{\Gamma}\left(z_n^{\text{exist}}\right)$. The vertical red lines in the decay region correspond to the left and right boundaries of the 99% of the rate parameter transient for the red α curve. The blue dashed curves show the effect of sweeping α between 0 (a horizontal line) and ∞ (a step function). $\alpha = 0$ is analogous to a traditional tracker, decaying the existence probability of all undetected obstacles at the same rate, independent from the occlusion information. The green line denotes the rate parameter for detected tracks, $D_n = 1$, which is positive for all z^{exist} to yield growth of the existence probability.

99% of the rate parameter transient for the solid red α curve, which can be computed using the equations derived in Appendix C. Note that when $\alpha = 0$, all objects are assumed to be visible to the sensor, and expected, but missing, measurements imply that the undetected objects do not exist, $(D_n = 0) \rightarrow (E_n = 0)$. Notice that if a track is undetected in sensor data, $D_n = 0$, its existence probability, $p(E_n)$, decays exponentially with rate inversely proportional to the negative information in (2.14). Conversely, when a track is detected in sensor data, $D_n = 1$, its existence probability grows with rate directly proportional to the positive information in (2.24).

Fig. 2.8 demonstrates the track existence model, (2.26), for detected tracks, $D_n = 1$, using the growth parameter for a 15 Hz sensor provided in Appendix

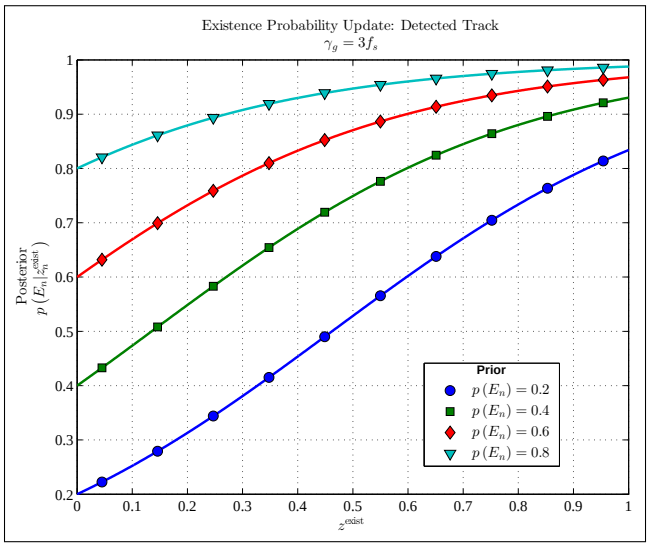

Figure 2.8: Demonstration of the existence probability update for detected tracks, $D_n = 1$, over a range of existence measurement, z^{exist}, and existence prior, $P(E_n)$.

C. Notice that confident data assignments (large z_n^{exist}) yield larger growth of the existence probability than ambiguous assignments (small z_n^{exist}) with the same existence prior. Further, the existence probability is a monotonically increasing function, reflecting the fact that positive information (receiving a measurement) congenitally cannot degrade confidence in the detected object's existence.

Fig. 2.9 demonstrates the track existence model for undetected tracks, $D_n = 0$, for the same scene depicted in Figs. 2.3 and 2.6, using the decay parameters for a 15 Hz sensor provided in Appendix C. Note that for small existence measurements, $z^{\text{exist}} \leq \underline{z}_{0.99}^{\text{exist}}$, the magnitude of the rate parameter, $\mathbf{\Gamma}(\cdot)$, approaches its maximum value, $\underline{\gamma}$, and the existence probability, $p(E_n)$, is approximately halved, corresponding to a strong belief that the track is spurious. Conversely, for large existence measurements, $z^{\text{exist}} \geq \overline{z}_{0.99}^{\text{exist}}$, the magnitude of the rate parameter approaches its minimum value, $\overline{\gamma}$, and the existence probability is preserved, corresponding to the belief that the missing measurement is reasonably

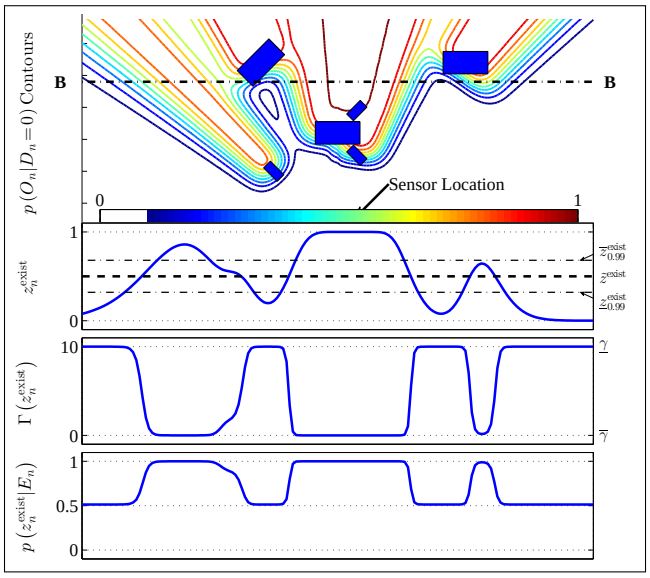

Figure 2.9: Demonstration of the existence model for undetected tracks. **Top:** Occlusion likelihood contours over a complex 2D scene from the point of view of the indicated sensor. **Second from top:** Existence measurement for undetected tracks, (2.25), corresponding to cross section B in the top pane. **Second from bottom:** Corresponding rate parameter, (2.29), using the suggested parameters for a 15 Hz sensor given in Appendix C. **Bottom:** Resulting existence measurement likelihood, (2.28).

explained by occlusion, and thus does not contradict object existence.

2.3.3 Data Association

In common probabilistic data association approaches, the discrete likelihood that measurement z_ℓ originated from object n, $a_{n,\ell}$, is computed as:

$$\begin{aligned} p(a_{n,\ell}|z_\ell) &= \frac{p(a_{n,\ell}, z_\ell)}{p(z_\ell)} \\ &= \frac{p(z_\ell|a_{n,\ell}) \cdot p(a_{n,\ell})}{p(z_\ell)} \end{aligned} \tag{2.31}$$

where $p(z_\ell|a_{n,\ell})$ is the likelihood of measurement z_ℓ given that it originated from object n, and $p(a_{n,\ell})$ represents *a priori* assignment information available before

the measurement is received. $p(a_{n,\ell})$ is typically chosen to be uniform over the assignment candidates, but other suitable options exist, such as the *a priori* probability that object n lies within the field of view of the sensor resolving measurement z_ℓ. Track birth is handled naturally by defining object index $n = 0$ as a new track with measurement likelihood, $p(z_\ell|a_{0,\ell})$, typically chosen to be uniform over the sensor field of view. The marginal probability of z_ℓ under this model is defined as:

$$p(z_\ell) = \sum_{i=0}^{N} p(z_\ell|a_{i,\ell}) \cdot p(a_{i,\ell}) \tag{2.32}$$

The traditional assignment likelihood in (2.31) is augmented to capture the following two data assignment constraints, referred to as the *visibility* and *existence* constraints, respectively, throughout this section:

1. All *genuine* measurements must be visible to the sensor given the correct assignment.
2. All objects emitting measurements must exist.

Augmenting the assignment likelihood in (2.31) to jointly consider the visibility and existence constraints stated above, gives:

$$p\left(a_{n,\ell}, \hat{D}_\ell, E_n \middle| z_\ell\right) = \frac{p\left(a_{n,\ell}, \hat{D}_\ell, E_n, z_\ell\right)}{p(z_\ell)} \tag{2.33}$$

where $\hat{D}_\ell$ denotes the predicted event that measurement z_ℓ is detected by the sensor. Note the distinction between the *observed* event that measurement z_ℓ is detected by the sensor, $D_\ell = 1 \; \forall \ell$, and the occlusion model *prediction* of the event, $\hat{D}_\ell$, the details of which are provided in the coming paragraphs.

The full joint probability in the numerator of (2.33) can be factored under the assumption that measurement z_ℓ and its predicted detection, $\hat{D}_\ell$, are both con-

ditionally independent of the existence of object n, E_n, given that measurement z_ℓ originated from object n, $a_{n,\ell}$:

$$\left\{z_\ell, \hat{D}_\ell\right\} \perp\!\!\!\perp E_n | a_{n,\ell} \tag{2.34}$$

This conditional independence rule is a consequence of the existence constraint. Specifically, implicit in the knowledge of object n having generated measurement z_ℓ, $a_{n,\ell}$, is that object n must exist, therefore, explicit knowledge of the existence variable, E_n, does not provide any additional information in the event that $a_{n,\ell}$ is known. Employing the conditional independence rule in (2.34) allows the following factorization:

$$p\left(a_{n,\ell}, \hat{D}_\ell, E_n, z_\ell\right) = p\left(z_\ell \middle| a_{n,\ell}\right) \cdot p\left(a_{n,\ell} \middle| E_n\right) \cdot p\left(\hat{D}_\ell \middle| a_{n,\ell}, z_\ell\right) \cdot p\left(E_n\right) \tag{2.35}$$

where $p\left(z_\ell | a_{n,\ell}\right)$ is the measurement likelihood and $p\left(a_{n,\ell} | E_n\right)$ represents any *a priori* assignment information available before the measurement is received, both introduced in (2.31). $p\left(E_n\right)$ is the existence probability of object n, for which an inference strategy was proposed in (2.26); this term favors assignments to objects that are likely to exist, in agreement with the existence constraint. Lastly, $p\left(\hat{D}_\ell \middle| a_{n,\ell}, z_\ell\right)$ is the probability that measurement z_ℓ is visible to the sensor, given that z_ℓ originated from object n. This visibility term is computed by substituting the metadata of z_ℓ, defined in (2.1), for that of object n, $\hat{z}_n$, in the detection likelihood, defined as the complement of the occlusion likelihood (2.12):

$$p\left(\hat{D}_\ell \middle| a_{n,\ell}, z_\ell\right) = 1 - p\left(O_n | D_n = 0, \hat{z}_n = z_\ell\right) \tag{2.36}$$

where $\hat{z}_n = z_\ell$ denotes the substitution of the measurement metadata for that of object n, to account for the candidate assignment, $a_{n,\ell}$. In the case of the new track candidate, $n = 0$, the measurement metadata is appended to the list of all tracked objects in the union calculation in (2.11) and (2.12). Note that (2.36) is

conditioned on object n being undetected in sensor data, $D_n = 0$, to reflect that the calculation is performed prior to an assignment decision; thus, detections at the object level have not yet been observed. This is a critical point for *predicting* visibility, as the occlusion model in (2.12) is designed such that $p\left(O_n | D_n = 1\right) = 0$ via (2.4). The marginal probability of z_ℓ under this assignment model is defined as:

$$p\left(z_\ell\right) = \sum_{i=0}^{N} p\Big(z_\ell \Big| a_{n,\ell}\Big) \cdot p\Big(a_{n,\ell} \Big| E_n\Big) \cdot p\Big(\hat{D}_\ell \Big| a_{n,\ell}, z_\ell\Big) \cdot p\Big(E_n\Big) \tag{2.37}$$

Let $\mathbb{S}_\ell \subseteq \{0, \ldots, N\}$ denote the set of all objects satisfying $p\left(\hat{D}_\ell \middle| a_{n,\ell}, z_\ell\right) \approx 1$ for measurement z_ℓ. Then, in agreement with the visibility constraint, for all *genuine* measurements, it must be true that $|\mathbb{S}_\ell| \geq 1$, to reflect that measurement z_ℓ was, indeed, detected, and therefore must be visible to the sensor under at least one candidate assignment. In these cases, the assignment likelihood in (2.33) favors assignments that satisfy this condition, $a_{n,\ell} \forall\, n \in \mathbb{S}_\ell$. In the event $\mathbb{S}_\ell = \{0, \ldots, N\}\ \forall\, \ell$, the assignment likelihood in (2.33) approaches the traditional likelihood in (2.31) with respect to visibility.

Conversely, in the event $|\mathbb{S}_\ell| = 0$, it is likely that measurement z_ℓ is spurious, as all genuine measurements must be visible to the sensor under the correct assignment, in agreement with the visibility constraint. Spurious measurements with this characteristic can arise due to unmodeled sensing anomalies or faulty processing of raw sensor data. While this type of spurious measurement may be rare with lidar sensors, it can be quite common with sensors that require heavy processing of the raw data to extract consumable measurements, such as vision sensors. In these cases, low unnormalized assignment likelihoods, (2.35), result for all candidates (including the new track candidate, $n = 0$). In order to avoid an arbitrary, incorrect, and costly assignment after normalization (i.e.

assignment to the best of a set of poor candidates), a null assignment candidate, $n = \varnothing$, is introduced as:

$$\begin{aligned} p\left(z_\ell | a_{\varnothing,\ell}\right) &= p\left(z_\ell | a_{0,\ell}\right) \\ p\left(\hat{D}_\ell \middle| a_{\varnothing,\ell}, z_\ell\right) &= 1 - p\left(\hat{D}_\ell \middle| a_{0,\ell}, z_\ell\right) \\ &= p\left(O_0 | D_0 = 0, \hat{z}_0 = z_\ell\right) \end{aligned} \tag{2.38}$$

Notice that the visibility component of the null candidate, $n = \varnothing$, favors areas of likely occlusion. All measurements assigned to the null candidate are discarded as spurious.

2.4 Practical Considerations

2.4.1 Ordering of Negative Measurements

As discussed in Section 2.3.1, the bearing and range components of the occlusion likelihood function are uncoupled, enabling sequential fusion of independent negative bearing and range measurements. Thus, the ordering of these measurements should not matter. Practically, however, the filter is utilizing parameterized mixture model approximations of all probability distributions. In the limit of infinite mixands, perfect approximations are achieved and the theoretical irrelevance of the measurement ordering is realized. Short of this, changing the order of the measurement fusion can impact the resulting posterior. Active management of mixands can be employed to mitigate these (and other) approximation artifacts, such as those presented in [23, 42].

2.4.2 Computational Complexity

The additional computational complexity for the negative information tracking filter supplements proposed in Section 2.3 are negligible in the event that all tracked objects are detected in the sensor scan. For each undetected object, the increased computation is dominated by the '*Negative* Measurement Update' block depicted in Fig. 2.5, which requires fitting a Gaussian Mixture to the 1D negative bearing measurement likelihood (a closed-form mixture is provided for the range likelihood in (2.21)), followed by $N_{\text{prior}} \times N_{\text{mix}}$ filter update steps for each negative measurement (bearing and range), where N_{prior} and N_{mix} denote the number of Gaussian mixands in the object state prior and negative measurement likelihood, respectively. This increased computation can be mitigated with judicious choices for mixture management and compression algorithms, for which many methods exist, e.g. [2, 23, 42, 45].

A significant computation savings can be realized by only fusing negative measurements that have an appreciable impact on the object state. That is, if the bearing component of the occlusion likelihood ≈ 1, the state prior is in agreement with the negative bearing information, and fusing a negative bearing measurement has a negligible impact on the object state (likewise for the range component). In practice, fusing negative measurements only when their corresponding likelihood components fall below a threshold saves computation with a minor impact on the posterior.

Moreover, a substantial computation savings can be achieved by only fusing negative measurements for undetected object tracks that are sufficiently likely to be occluded. Specifically, according to the negative information in (2.14), undetected tracks arise as a consequence of either: 1) the object exists and is

occluded (genuine case), or 2) the object does not exist and is the consequence of a spurious measurement (spurious case). The track existence model, (2.26), is designed to distinguish between these cases by considering the occlusion likelihood developed in Section 2.2 as a *measurement* of the feasibility of the genuine case, z^{exist}. If the genuine case is sufficiently feasible, the existence model allows the track to persist, and the negative measurements act to maintain a sensible belief of the object state in the absence of traditional measurements. If the genuine case is not sufficiently feasible, the existence model quickly discards the track as spurious; updating the state estimate of spurious tracks that will soon be discarded is unproductive. This is depicted in the tracking overview block diagram in Fig. 2.5 by the left most condition exiting the 'Data Association' block and bypassing measurement update.

Lastly, capturing the true shape of the rising edge of the negative range likelihood function, (2.21) using a large set of mixands, is not critical to the tracking filter, as the true likelihood is inherently conservative, as demonstrated in Fig. 2.4. Further, the range likelihood models each object as its closest point, ignoring the fact that the occluding object, being of non-negligible size, has some depth that also cannot be occupied by the occluded object. Therefore, in practice, a small approximation error is acceptable, and a compromise can be made in choosing the common standard deviation, σ^R, to limit the number of mixands required in the model.

2.4.3 Unbiased estimates

In a general tracking problem, the filter is tasked with tracking maneuvering objects without knowledge of their control inputs or intent. The classical approach is to choose a steady-state motion model with enough process noise to account for maneuvers, and then rely on frequent and precise sensor measurements to maintain consistent and unbiased state estimates throughout the set of conceivable maneuvers.

While negative measurements are useful in maintaining sensible beliefs of the local scene, they are generally less precise than positive measurements. Therefore, maneuvers not explicitly included in the dynamics model could result in biased state estimates (consistency should be preserved, barring any linearization artifacts that result from the biased estimate). In some applications, it may be useful to utilize a multiple model framework for undetected tracks to better capture the possible maneuvers of the object, and use the occlusion likelihood developed here to evaluate the probability of each model, as is done in [49]. That is, as long as the track is undetected in sensor data, models that predict a history of likely occlusion are more likely to represent ground truth than those that do not.

2.5 Baseline Tracker For Simulation and Experiments

An elegant approach for multi-object tracking from a dynamic sensory vehicle was developed in [37] and [38] for Cornell University's submission to the DUC. Specifically, a Rao-Blackwellized particle filter (RBPF) is utilized to factor

the joint estimation problem into a data assignment problem solved via particle filter, and a dynamic obstacle tracking problem solved with extended Kalman filters (EKF). Within each EKF the object state is modeled as:

$$X = [x, y, s, h, \psi]^T \tag{2.39}$$

where x, y, and ψ describe the 2D position and orientation, in ego-coordinates, of an arbitrary object-fixed coordinate frame, while s and h describe the objects absolute ground speed and heading relative to the ego robot. The states evolve according to the following nonlinear dynamics model:

$$\begin{aligned} \dot{x} &= s \cdot \cos(h) - v_x + y \cdot \omega_z + e_x \\ \dot{y} &= s \cdot \sin(h) - v_y - x \cdot \omega_z + e_y \\ \dot{s} &= e_s \\ \dot{h} &= -\omega_z + e_h \\ \dot{\psi} &= -\omega_z + e_\psi \end{aligned} \tag{2.40}$$

where v_x, v_y, and ω_z describe the ego robot velocity, and $\left[e_x, e_y, e_s, e_h, e_\phi\right]^T \sim \mathcal{N}(0, \boldsymbol{Q})$ denote white process noise variables.

In addition to the Gaussian belief of the object states (typical of most tracking filters), each EKF stores the most recent lidar cluster assigned to each object in the associated object-fixed coordinate frame; this stored cluster is used to predict the measurement distribution for the subsequent sensor scan from the *a priori* state distribution using the sigma point transform [28]. The objects are assumed to be rigid bodies, thus maintaining this cluster does not increase the dimension of the state vector; simply estimating the evolution of the object-centric coordinate origin using (2.40) is enough to reproduce the entire cluster in sensor coordinates at future times. Further, the set of stable metadata defined

in (2.1) is extracted from each cluster, $i \in \{1, \ldots, N\}$, and used to form the filter innovations driving the EKF measurement update.

The filter summarized above, and presented in detail in [38], is chosen as the baseline filter for evaluating the uses of negative information proposed in this paper. While the proposed uses of negative information can supplement a variety of tracking filters, the distributions over the object metadata in (2.1) are already computed in [38] for data association and measurement update, deeming it a synergistic and elegant choice.

For the remainder of this paper, F_A denotes the baseline filter [38], and F_0 denotes F_A adapted to consider the negative information, as outline in the block diagram of Fig. 2.5. In all studies F_0 utilizes the existence model parameters defined in Appendix C, and the negative measurement GMM posterior is compressed to a single Gaussian after each negative measurement update by preserving first and second moments. F_A maintains undetected tracks satisfying some occlusion reasoning for up to 1 min in the tracker under pure dynamics prediction. For the purposes of this evaluation, tracks satisfying $p(O_n|D_n = 0) \geq 0.05$, as determined by the occlusion likelihood model in (2.12), are considered to have satisfied this occlusion reasoning.

2.6 Simulation Results

The negative measurements proposed in Section 2.3.1 are evaluated for the complex scenario depicted in Fig. 2.10. The scenario involves three 0.5 m radius circular static objects, and three 5 m $\times$ 2.5 m rectangular dynamic objects traveling in tight formation along a sinusoidal path of increasing amplitude. Object IDs

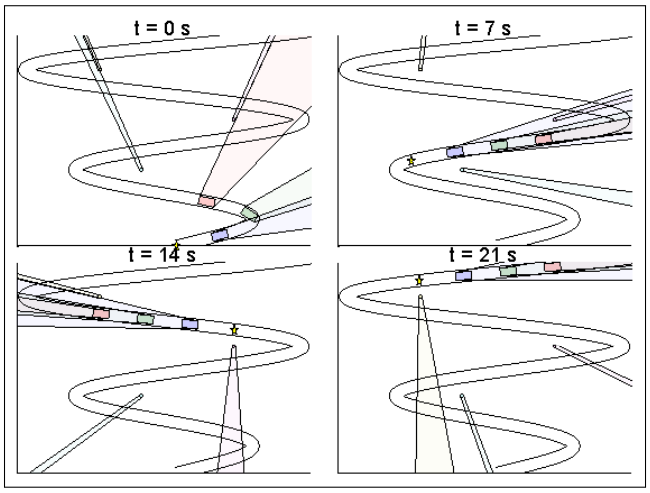

Figure 2.10: Snapshots of the simulation involving 3 rectangular dynamic objects, 3 circular static objects, and a dynamic ego sensory vehicle denoted by the yellow star. The occlusion shadow cast by each object, as seen from the sensory vehicle, is depicted with a semitransparent shadow of the same color as the object casting the shadow.

and labels are provided in Table 2.1, where the parenthetical descriptions refer to the object representations in Fig. 2.10. The dynamic objects are traveling with

Table 2.1: Objects to be tracked

Object ID	Object Label
1	Lead Vehicle (red rectangle)
2	Middle Vehicle (green rectangle)
3	Trailing Vehicle (blue rectangle)
4	Closest Static Object (cyan circle)
5	Middle Static Object (magenta circle)
6	Farthest Static Object (yellow circle)

6 m spacing at 13 m/s, which is far more aggressive than would be considered safe in a true driving scenario; this was chosen to make tracking more difficult by increasing the number, duration, and complexity of occlusion events which

are summarized in Table 2.2. Notice that the lead and middle vehicles (objects 2 and 3) are occluded for approximately 80% of the simulation.

Table 2.2: Occlusion event summary

Object ID	Total # of occlusion events	Max # of concurrent occluders	Occlusion event duration statistics		
			Max	Mean	Total
1	0	0	0	0	0
2	5	1	6.08	3.70	18.48
3	7	2	5.92	2.58	18.08
4	25	1	1.04	0.24	6.00
5	11	3	2.72	0.58	6.40
6	14	3	3.12	0.63	8.88

The ego sensory vehicle trails the dynamic objects with equal spacing, while tracking all six objects in the scene. Data is simulated for a 360° FOV single-scan planar lidar firing at 12.5 Hz with 0.5° bearing resolution. Random sensor noise is drawn independently for each beam in each scan from $\mathcal{N}\left(0, 1\,\text{cm}^2\right)$ and added to the lidar range returns to simulate the accuracy of realistic lidar sensors. Both filters, F_A and F_0, are run for comparison in terms of tracking *precision* and *accuracy*.

The *precision* of the ℓth object state estimate at time k, $\epsilon_{\ell,k}$, is defined as the inverse of the area spanned by the the n-σ object position marginal covariance ellipse, $\boldsymbol{P}_{\ell,k}$:

$$\epsilon_{\ell,k} = \left(\pi \cdot n^2 \cdot \sqrt{\left|\boldsymbol{P}_{\ell,k}\right|}\right)^{-1} \tag{2.41}$$

Therefore, the overall precision improvement is computed as the mean precision

ratio of filter F_0 to filter F_A:

$$\begin{aligned}\bar{\epsilon}_{F_{0:A}} &= \frac{1}{N_o \cdot N_k} \sum_{k=1}^{N_k} \sum_{\ell=1}^{N_O} \frac{\epsilon_{\ell,k}^{F_0}}{\epsilon_{\ell,k}^{F_A}} \\ &= \frac{1}{N_o \cdot N_k} \sum_{k=1}^{N_k} \sum_{\ell=1}^{N_O} \sqrt{\frac{\left|\boldsymbol{P}_{\ell,k}^{F_A}\right|}{\left|\boldsymbol{P}_{\ell,k}^{F_0}\right|}}\end{aligned} \tag{2.42}$$

where N_O and N_k denote the number of tracked objects and number of time steps in the data, respectively. Notice that (2.42) is independent of the choice of pdf level set, n. Conceptually, the metric in (2.42) is interpreted as the posterior object position belief having, on average, $\bar{\epsilon}_{F_{0:A}}$ times less uncertainty when negative measurements are utilized; thus filter F_0 is considered to be $\bar{\epsilon}_{F_{0:A}}$ times more precise than filter F_A for the given scenario.

The *accuracy* of the ℓth object state estimate at time k is defined as the joint likelihood of the true object metadata given the filter belief, $p\left(\theta_{\ell,k}, \phi_{\ell,k}, \rho_{\ell,k} \middle| F_A/F_0\right)$, where the object metadata are defined in (2.1). Therefore, the overall accuracy improvement is computed as the mean likelihood ratio of filter F_0 to filter F_A:

$$\bar{\lambda}_{F_{0:A}} = \frac{1}{N_o \cdot N_k} \sum_{k=1}^{N_k} \sum_{\ell=1}^{N_O} \frac{p\left(\theta_{\ell,k}, \phi_{\ell,k}, \rho_{\ell,k} \middle| F_0\right)}{p\left(\theta_{\ell,k}, \phi_{\ell,k}, \rho_{\ell,k} \middle| F_A\right)} \tag{2.43}$$

Conceptually, the metric in (2.43) is interpreted as the true scenario being, on average, $\bar{\lambda}_{F_{0:A}}$ times more likely when the negative measurements are utilized; thus filter F_0 is considered $\bar{\lambda}_{F_{0:A}}$ times more accurate than filter F_A for the given scenario. Note, to avoid numerical underflow when evaluating likelihoods, (2.43) is computed directly from the metadata log-likelihoods using the *log-of-sums trick* provided in Appendix D.

The precision and accuracy metrics for the simulation are provided in Table 2.3, computed over all objects at all times, and then again for only the occluded objects. The results demonstrate that the negative measurements proposed in this paper enable a more accurate and precise belief of the local scene.

Table 2.3: Precision and Accuracy Results

Data Used	$\bar{\epsilon}_{F_{0:A}}$	$\bar{\lambda}_{F_{0:A}}$
All objects	1.46	10.6
Occluded objects	2.22	16.56

Notice that the performance improvement is diluted when computed over all objects due to the fact that the filters are identical in tracking objects detected in sensor data; therefore, computing the metrics for occluded tracks only provides more insight into the value of the negative measurements proposed in this work.

To provide insight into the qualitative benefits of negative measurement fusion, Figs. 2.11 and 2.12 present four time snapshots of the tracking results for filter F_A and F_0, respectively, as seen from a global reference frame[1]. The snapshots are distributed over approximately half a period of the sinusoidal path, and similar tracking characteristics arise periodically with the trajectory as demonstrated in the video accompanying this paper[1]. The true scenario for each frame is depicted as it is in Fig. 2.10, and the tracking belief of each object is overlaid as the stored lidar point cloud (representing the most recently visible object surface) centered in the 1-σ ellipse of its position belief. Note that the tracked object beliefs are colored to match their corresponding true object, with the point clouds colored black for objects believed to be occluded.

The first time frame ($t = 3.6$ sec) shows the scenario as the lead vehicle rounds the second peak of the sinusoidal path, and is about to emerge from

[1]An annotated video of the full tracking results from which these snapshots are taken is provided by the author as supplementary downloadable material available at http://ieeexplore.ieee.org. The video size is 38.2 MB.

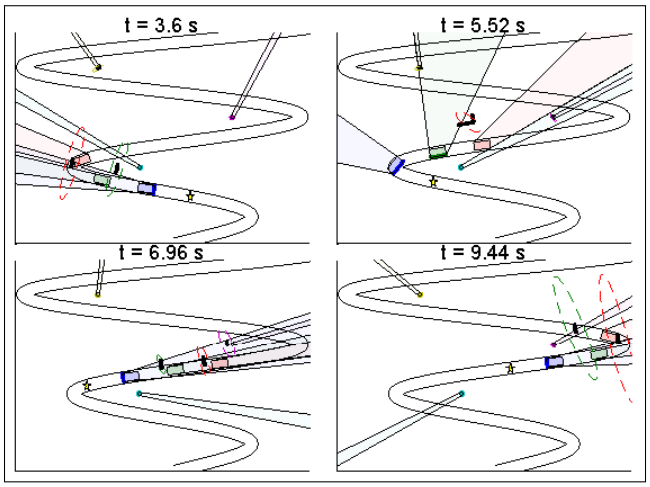

Figure 2.11: Time snapshots of the baseline filter, F_A, tracking results overlayed with the true scenario as seen from a global reference frame.

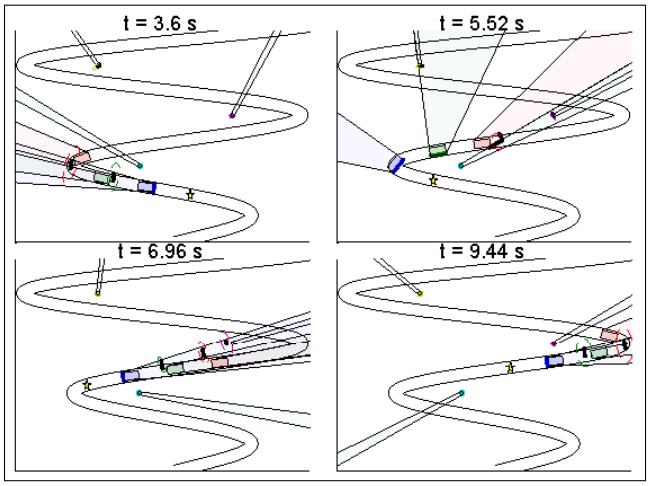

Figure 2.12: Time snapshots of the negative information filter, F_0, tracking results overlayed with the true scenario as seen from a global reference frame.

a 2.16 sec occlusion event. Notice that the F_A belief reflects a substantial probability that the occluded objects (the lead and middle vehicles) lie in space visible to the sensor, at odds with the negative information and human intuition; conversely, filter F_0 provides a more precise, accurate, and intuitive belief of the occluded object states.

The second time frame (t = 5.52 sec) demonstrates the ability of filter F_0 to recover from unmodeled object maneuvers despite the absence of traditional measurements. Specifically, the periodic sharp turns at the peaks of the sinusoidal trajectory are strong violations of the constant heading model assumed in both filters, (2.40), giving rise to a spike in heading estimate error. Subsequent occlusion of the lead vehicle prevents a full recovery of the heading estimate due to missing positive measurements, and results in an increasingly inaccurate filter F_A estimate during the occlusion event. Conversely, the negative measurements update the filter F_0 position belief to coincide with the occlusion shadow cast by the occluding object, and, through coupling in the dynamics model, the filter N heading estimate recovers from the sharp unmodeled turn maneuver. Again, the negative measurements result in a more precise, accurate, and intuitive belief of the occluded object states, while the baseline filter reflects a strong belief that the occluded object lies in space visible to the sensor, at odds with the negative information and human intuition.

The third time frame (t = 6.69 sec) shows one of the many examples of complicated occlusion scenarios present in this simulation. The scenario involves multiple simultaneously occluded objects, some of which have multiple simultaneous occluders, and some of which have occluders that are, themselves, occluded. Complicated cases such as this one are handled naturally and rigorously by the occlusion model defined in (2.12).

The final time frame (t = 9.44 sec) is approximately half a trajectory period after the first frame (t = 3.6 sec) and depicts a similar scenario, except that the lead vehicle has just completed a sharper turn in the opposite direction. Further, the concluding occlusion event is longer at 3.44 sec, giving rise to increased

uncertainty in the F_A belief. The increased uncertainty of F_A makes the benefits of negative measurements more pronounced; specifically, the filter F_A belief of the middle and lead objects reflects a substantial probability that the objects lie within space visible to the sensor, at odds with the negative information present in the sensor data.

2.7 Experimental results

The multiple uses of negative information proposed here are validated using experimental data collected from three Ibeo ALASCA XT range finders on Cornell University's autonomous vehicle platform [37], each recording at 12.5 Hz. The scenario consists of a lead vehicle, an occluding vehicle, and the autonomous ego vehicle.

Fig. 2.13 is an annotated photograph of the experiment site taken from a location near the ego vehicle at the start of the experiment. Notice that the road is lined with vegetation, which creates challenges in the filter due to the increase in spurious lidar clusters. The scenario begins with the lead vehicle departing a 3-way intersection, and the ego vehicle stopped at the intersection in the same lane. The occluding vehicle executes a right turn from the perpendicular lane, occluding the lead vehicle from the ego vehicle. The ego vehicle then proceeds straight through the intersection, in pursuit of the now-occluded lead vehicle.

Data was collected for this scenario, and filter F_0 is analyzed in the following three studies:

- **A**: Belief sensibility study

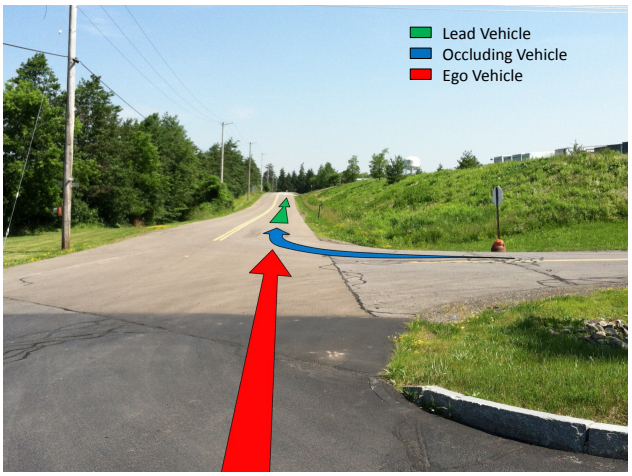

Figure 2.13: Photograph of the experiment environment, annotated with the high-level scenario. The trajectories shown are synchronized such that the vehicles continue down the road in the formation implied by the figure.

- **B**: Track maintenance study
- **C**: Model accuracy study

A fourth study was run to analyze the negative information data association outlined in Section 2.3.3. However, sensor anomalies such as laser reflection and absorption were not experienced during the run, and the data assignment problem is relatively benign; therefore, the benefits of negative information for data association were not apparent, and this study was omitted.

2.7.1 Belief Sensibility Study

The sensibility of the occluded object belief of F_0 is evaluated against filter F_A, the baseline parametric filter [38] (no negative information). Fig. 2.14 presents five time snapshots of the tracking results for F_0 (green) and F_A (cyan) overlayed during the 27.5 sec occlusion event, as seen from the ego vehicle; the global ego

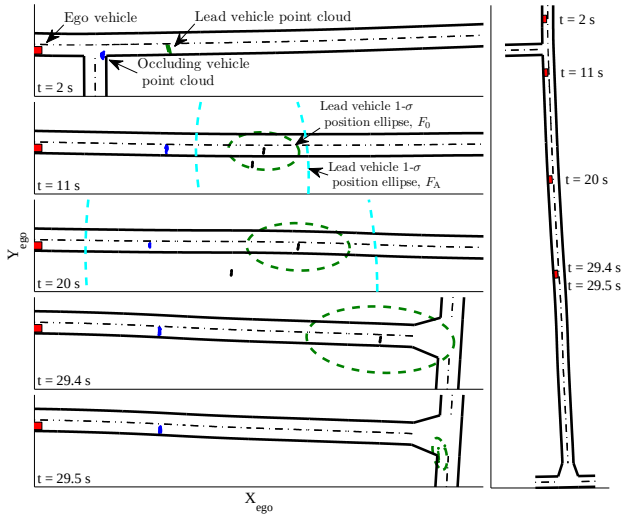

Figure 2.14: **Left:** Time snapshots of the experimental tracking results for F_0 and F_A overlayed, as seen from the ego vehicle. **Right:** The global position of the ego vehicle for each of the time snapshots.

vehicle pose at each time snapshot is depicted to the right of the figure. The first time frame (t = 2 sec) shows the last positive evidence of the lead vehicle, prior to occlusion, as a view of the rear bumper. This lidar point cloud is stored in an object-centric reference frame, propagated forward in time with the object state, and depicted in future time frames as a black point cloud centered in its 1-σ position ellipse.

Three subsequent time frames (t = 11, 20, and 29.4 sec) highlight the evolution of the lead vehicle belief over the remainder of the occlusion event. Note that F_A incorrectly assigns a spurious measurement to the lead vehicle track at t = 24.5 sec, resulting in $p(O_n|D_n=0) < 0.05$ for the updated state, and deletes the lead vehicle track; thus F_A has an incomplete belief of the local scene for $t > 24.5$ sec. Prior to deletion, a majority of the F_A belief is well within view of the sensors, at odds with the negative information (and human intuition). Meanwhile, the 1-σ ellipse for the F_0 belief is consistently contained within the shadow cast

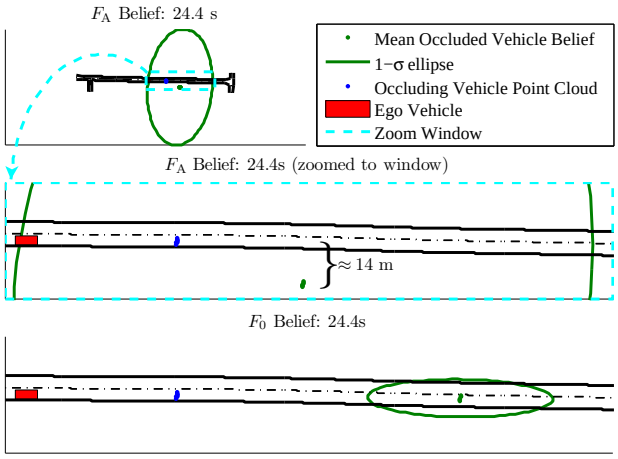

Figure 2.15: Comparison of the F_0 and F_A belief of the occluded object state at $t = 24.4$ sec, just prior to F_0 misassociating a spurious measurement to the occluded object. **Top:** Full-scale time snapshot of the F_A belief. **Middle:** Top pane zoomed to the window denoted by the dashed cyan rectangle. **Bottom:** Time snapshot of the F_0 belief, at the same scale as the middle pane.

by the occluding vehicle; also, the mean is consistently beyond $\approx 3\sigma$ of the visible edge of the occluding vehicle. This belief is consistent with the assertion that the undetected object is not likely to be located in areas visible to the sensors.

The final time frame of Fig. 2.14 ($t = 29.5$ sec) demonstrates that F_0 is able to recognize the lead vehicle emerging from the occlusion shadow during a right turn, enabling the ego vehicle to continue accurately tracking the lead vehicle.

Fig. 2.15 shows a time snapshot of the tracking results at $t = 24.4$ sec; the time step just prior to F_0 misassociating a spurious measurement to, and losing track of, the occluded lead vehicle (between the third and fourth frames of Fig. 2.14). The peak precision ratio of 89.94 occurred at this time step, and the mean precision ratio defined in (2.42) was computed as 30.50 for the lead vehicle during the first 22.4 sec of the occlusion event through this time step; this is to say, on average, the F_0 belief was 30.5 times more precise than that of F_A.

The occluded vehicle accelerated from the first intersection while undetected, resulting in a bias in the F_A velocity estimate. However, through coupling in the dynamics model, the negative measurements enabled F_0 to increase its ground speed estimate from 7.1 m/s at the onset of occlusion to $\approx$ 12 m/s; much closer to the true steady-state vehicle speed of $\approx$ 13 m/s. This recovery by F_0 is apparent in Figs. 2.14 and 2.15 as the growing distance between the means of the F_0 and F_A beliefs. Further, notice the F_0 estimate is consistently centered in the correct road lane despite a slight bend in the road just after the onset of occlusion; the bend results in $\approx$ 14 m of accumulated off-track error (distance to the closest point on the road lane centerline) in the mean of the F_A estimate, the peak of which is witnessed in the middle pane of Fig. 2.15.

The belief sensibility study demonstrates that negative information enables the filter to maintain a sensible, intuitive, and relatively precise belief of the occluded object state, despite the absence of traditional measurements.

2.7.2 Track Maintenance Study

Track maintenance derived from the track existence model of F_0 is evaluated against filter F_B; the baseline parametric filter [38] (no negative information), modified to delete all undetected tracks quickly. F_B uses the track existence model, (2.26), with $\alpha = 0$ to achieve the uniform rate parameter depicted in Fig. 2.7, and $\gamma_g \approx \infty$ such that the existence probability updates to $p(E_n) \approx 1$ for all objects after being detected in a single sensor scan. Note that, with these settings, F_B will only delete tracks that are undetected in 6 consecutive sensor frames (0.5 sec).

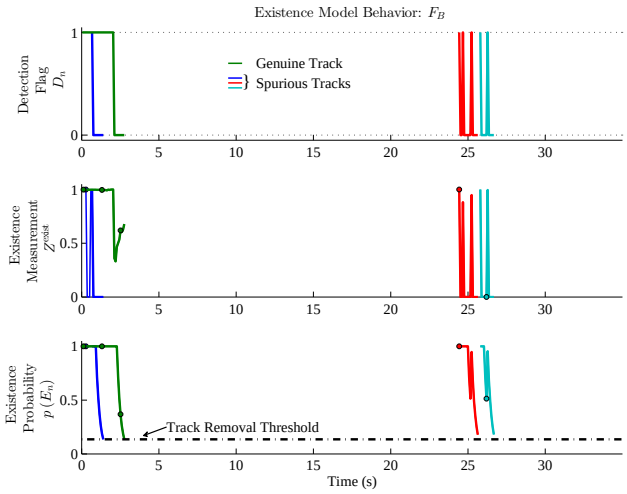

Figure 2.16: Behavior of the existence model for filter F_B. **Top:** Time history of the detection flag, D_n, for each undetected track. **Middle:** Existence measurement, z^{exist}, for each undetected track. **Bottom:** Time history of the existence probability for each undetected track.

Figs. 2.16 and 2.17 demonstrate the behavior of the existence model for all undetected tracks in F_B and F_0, respectively. Note that the existence measurement, z^{exist}, is plotted in the middle pane of Fig. 2.16 for reference only; the existence probability in F_B was made independent of the existence measurement by setting $\alpha = 0$ and $\gamma_g \approx \infty$. During the experiment there was a single occluded genuine track (the lead vehicle), and six spurious measurements generated from the vegetation depicted in Fig. 2.13, giving rise to five spurious tracks in F_0, and three spurious tracks in F_B.

Fig. 2.17 shows that F_0 maintained the lead vehicle track (green) for the entire 27.5 sec occlusion event, as also demonstrated in Fig. 2.14, while Fig. 2.16 shows F_B deleted the lead vehicle track 6 sensor frames (0.5 sec) after becoming occluded. Further, F_0 correctly identified and deleted five tracks born from spurious measurements quicker than F_B; four of five were deleted in 3 sensor frames (0.25 sec), and the fifth (red in Fig. 2.17) was assigned a second spurious

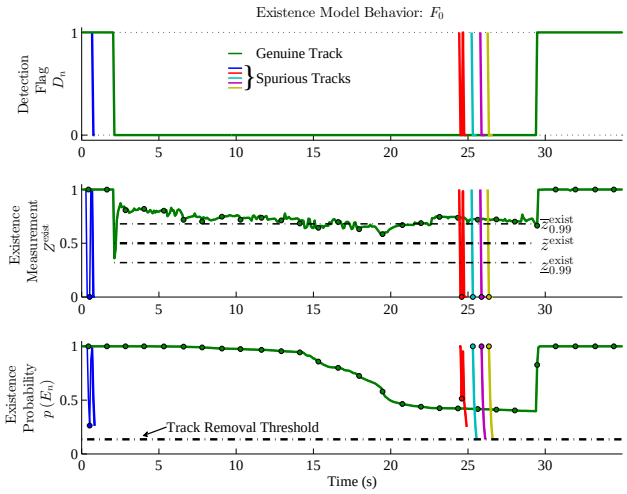

Figure 2.17: Behavior of the existence model for filter F_0. **Top:** Time history of the detection flag, D_n, for each undetected track. **Middle:** Existence measurement, z^{exist}, for each undetected track. **Bottom:** Time history of the existence probability for each undetected track.

measurement, extending its life to 5 frames, (0.42 sec). Alternatively, F_B deleted its three spurious tracks in 6 frames (0.5 sec), 10 frames (0.83 sec, assigned two additional spurious measurements), and 9 frames (0.75 sec, assigned one additional spurious measurment), listed chronologically.

Note that, in Fig. 2.17, when the lead vehicle is undetected, $D_n = 0$, and $z^{\text{exist}} > \bar{z}^{\text{exist}}_{0.99}$ ($2.1 < t < 14.2$ and $22.7 < t < 29.5$), the existence probability remains approximately constant to reflect the belief that the missing measurements are suitably explained by occlusion. Conversely, the existence probability decays at close to the maximum rate allowed by the model for all spurious tracks, as they regularly receive existence measurements of $z^{\text{exist}} < \underline{z}^{\text{exist}}_{0.99}$. Lastly, when the lead vehicle is detected, its existence probability climbs above 97% within 4 sensor scans due to consistent and confident data assignment.

The track maintenance study demonstrates that negative information en-

ables the tracker to distinguish between genuine and spurious tracks, maintaining the former and deleting the latter, resulting in a more complete and accurate belief of the local scene.

2.7.3 Model Accuracy Study

The accuracy of the occluded object state belief of F_0 is evaluated against filter F_C; a particle implementation of the baseline parametric filter [38] with added negative measurements. Given that F_C is a non-parametric filter, the state belief is free from linearization errors and information loss associated with modeling the object state posterior as a single Gaussian. Further, the negative measurements can be fused directly without modeling their likelihood in a form consumable by a parametric filter, as outlined in Section 2.3.1 for F_0. Specifically, the weight of the ith particle representing the belief of object n at time k, $w_n^{i,k}$, is updated via the following equation:

$$w_n^{i,k} = \frac{p\left(O_n|D_n = 0, x_n = x_n^i\right) \cdot w_n^{i,k-1}}{\sum_i p\left(O_n|D_n = 0, x_n = x_n^i\right) \cdot w_n^{i,k-1}} \tag{2.44}$$

where $p\left(O_n|D_n = 0, x_n = x_n^i\right)$ is the occlusion likelihood of object n given in (2.12), assuming its state is represented by particle i. A total of $100,000$ particles were used to approach the optimal occluded object state belief given the modeled negative information. Note that the object posterior in F_0 was reduced to a single Gaussian at each time step, representing the worst-case information loss, and best-case computational efficiency. However, dynamic mixand reduction methods are available that offer a trade-off between computation and accuracy of the object posterior at each time step [42], [23]. It should also be noted that F_C is significantly more expensive than F_0, and thus not suitable for real-time

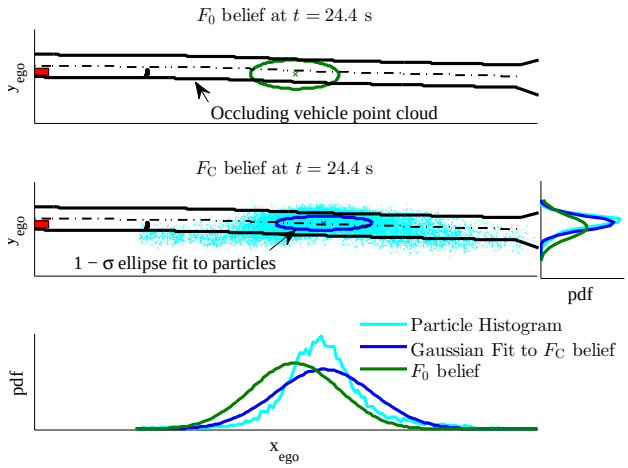

Figure 2.18: **Top:** Time snapshot of the F_0 belief at $t = 24.4$ sec. **Middle:** Time snapshot of the F_C belief at $t = 24.4$ sec. **Bottom:** Marginal distributions over the x_{ego} dimension (major principal axis of the position uncertainty) of the normalized histogram of: the F_C particle set, the Gaussian fit to the F_C particle set, and the F_0 Gaussian. **Right:** Same as bottom plot but for y_{ego} dimension (minor principal axis of the position uncertainty).

implementation.

Fig. 2.18 compares the F_0 and F_C beliefs for a single time step, which is typical of the entire run. A Gaussian is extracted from F_C by matching the first and second moments of the particle set, and overlays the particle set in the middle pane. Further, the x_{ego} and y_{ego} position marginals are overlayed with a normalized histogram of the particle set in the bottom and right panes, respectively.

Note that F_0 approximates the F_C belief well despite the accumulation of linearization errors, modeling errors, and worst-case information loss (reduction to a single Gaussian mixand at each time step) over the duration of the experiment.

The model accuracy study demonstrates that the parametric formulation of F_0, particularly the Gaussian mixture approximation of the negative measurement fusion process and subsequent compression to a single Gaussian posterior, has comparable accuracy to much more expensive non-parametric filters that avoid modeling and linearization errors altogether.

2.8 Conclusion

This paper introduced a novel method to model occlusion, and presented three specific approaches to using the model in dynamic multi-object tracking of extended objects from a moving vehicle, by using negative information. Experimental results demonstrate that the use of negative information improves the precision and accuracy of the tracking filter, as well as the sensibility of occluded object state estimates when traditional sensor measurements are unavailable. The ability to maintain sensible beliefs of occluded objects acts to decrease situational risk by increasing scene awareness, and enables high level behaviors that might otherwise be thwarted by occlusion. The proposed methods were demonstrated in the context of autonomous urban driving, however, the concepts presented can be formulated to fit a variety of extended multi-object tracking applications.

CHAPTER 3

PRIORITY-BASED TRACKING OF ARBITRARILY SHAPED EXTENDED OBJECTS

3.1 Introduction

Humans consistently outperform robots in perceptual tasks despite certain hardware advantages favoring robots over humans; this suggests that human cognition is superior to analogous robotic algorithms in these areas, and potentially worth emulating. For instance, sensors providing metric information over a wide field-of-view (FOV) are readily available for robots, while human sensors provide ordinal information, at best, over a limited FOV. In fact, empirical studies have concluded that human vision provides ordinal information via a variety of visual cues, such as occlusion, binocular disparities, and motion parallax, which the human brain fuses into a single cohesive belief of the perceptual space [16,17,33,35]. As distance from the observer decreases, the number, type, and quality of available ordinal cues increases, and the human belief quickly converges from an imprecise ordinal representation to a precise metrical one, despite purely ordinal sensor information.

As with many other biological phenomena, the characteristic of human perception described above serves as a complement to most human ventures, in that humans typically only require detailed, metrical representations of the nearby scene in which they are currently an active participant; therefore, anything more than a general ordinal awareness of objects at greater distances is, at the very least, a misallocation of limited cognitive resources, and a precursor for distraction.

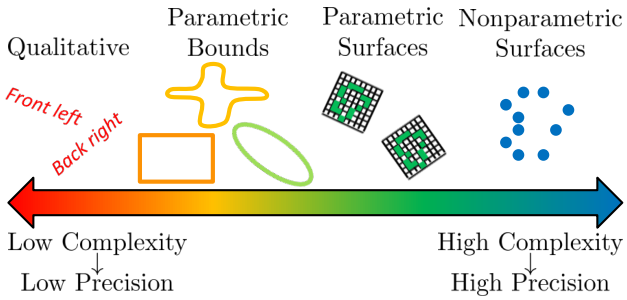

Figure 3.1: Distilled spectrum of extent model complexity and tracking precision, which tend to be directly correlated.

Given that robots with finite computational resources are often employed to perform human tasks such as driving, many of the requirements of human perception discussed above apply equally well to robotic perception. However, analogs to the complementary human perception characteristics are largely absent from robotic algorithms. Inspired by human cognition, this work proposes a priority-based framework for allocating computational resources in extended object tracking.

Object tracking is a perception application in which the states (e.g. kinematics) of objects present in the local environment are estimated from sensor data. Extended objects are defined as objects of non-negligible size relative to the sensor resolution, such that they cannot be accurately modeled as a mathematical point. Extended object tracking (EOT) differs from traditional object tracking in that it violates the foundational assumption that each object can return, at most, a single measurement per sensor query. Further, extended objects cast shadows in sensor data, known as occlusion shadows, which result in incomplete or missing measurements of the objects of interest; this includes self-occlusion, in which the object surface nearest the sensor occludes its remaining self. This work is particularly focused on applications in which the extended object size

and shape is unknown *a priori*, precluding tracking approaches that rely on this information, such as spatial distribution models [19].

Various successful approaches to extended object tracking in the absence of *a priori* shape information have been proposed, spanning a spectrum of computational complexity and tracking precision. Two central and related factors determining an EOT algorithm's position on this spectrum, are the detail and accuracy of the object shape/extent model. Specifically, detailed and accurate knowledge of object shape/extent enables high fidelity sensor models that offer a precise and detailed interpretation of the sensor data, and its relationship to the object state; in this way, detailed and accurate extent models engender high precision tracking, generally at the expense of increased computational complexity. Fig. 3.1 depicts a generalization of this trade-off, patently distilled for the following discussion of prior work.

At the low complexity end of the spectrum, qualitative/topological object tracking approaches exist, where the object state definition itself is imprecise by nature. For example, the state could be defined as a single discrete random variable representing the region of space, or topological node, that a dynamic object occupies over time. Qualitative state representations have gained interest in robotic applications due to their efficiency, scalability, and natural synergy with inexpensive ordinal sensor information, such as monocular cameras or human input [36]. In a conceptually related approach, traditional static occupancy grid mapping concepts, which are relatively simple, have been extended to characterize dynamic scenes [4, 44]. These approaches tend to be extremely efficient, but their qualitative/topological state representation is too imprecise for applications requiring robots to safely navigate in close proximity to, or interact with,

other dynamic objects.

Moving along the spectrum toward higher precision, simple parametric object shape models are prescribed *a priori*, and the parameters of the model are jointly estimated as states in the object tracker. Common simple parametric shapes include circular discs [8], ellipses [7, 9, 12, 30, 31], and rectangles [10,34,40]. These simple shape models generally represent a tight enclosing bound, or circumscription of the true, more complex, underlying shape, rather than the shape itself; therefore, they are sufficient for tracking a variety of arbitrarily shaped objects without *a priori* knowledge of the object shape or size. However, the inherent, uncharacterized mismatch between the true underlying object shape and the simple prescribed circumscription restrict the sensor model fidelity, thereby degrading tracking precision. This degradation is mitigated somewhat by increasing the complexity and flexibility of the shape bound; this is the goal of star convex random hypersurface models (RHM) [11]. These models are appropriate when detailed *a priori* information about object shape and size is unavailable, and computational efficiency is at a premium.

A simple and intuitive method for tracking arbitrarily shaped extended objects involves the use of occupancy grids anchored to an object-centric coordinate frame, dubbed Object Local Grid Maps (OLG) [5]. OLG shape models can be rigorous and flexible, however, the precision and complexity depend on an appropriate choice of grid extent and resolution, which requires some *a priori* knowledge of the size and shape complexity of objects to be tracked.

Finally, at the high precision end of the spectrum, very detailed, non-parametric point cloud models are employed [24, 25, 27, 37, 38, 48–51]. These models are extremely flexible in providing rich, 2 or 3-dimensional (3D) ren-

derings of the true underlying object surfaces for arbitrary shapes, and often do not require *a priori* information about the object shape and size. These detailed surface representations enable high precision sensor models, which, in turn, contribute to high precision kinematic state estimates; all at the expense of high computational complexity. Therefore, these methods are appropriate when tracking precision is at a premium, *a priori* information about object shape and size is unavailable, and computational resources are abundant.

The appropriate approach for tracking arbitrarily shaped extended objects is application dependent, and may also be time dependent within an individual application; for example, if any of the following items are time dependent: environment characteristics; robot task; quality, quantity, or type of available sensor information; availability of computational resources; or object relevance to the ego robot task. This work addresses the final exemplified scenario by proposing a *priority*-based tracking framework for arbitrarily shaped extended objects in the absence of *a priori* knowledge of object shape or size; here, *priority* refers to the object's relevance to the ego-robot task, as defined by a chosen set of object relevancy metrics. For this work, probabilistic object relevancy metrics are developed to dynamically allocate computational resources by informing a hybrid system switching strategy; each mode in the hybrid system represents a different EOT method with unique characteristics on the complexity-precision spectrum of Fig. 3.1. In this way, objects relevant to the ego vehicle task are allocated more resources and tracked with higher precision, while objects of peripheral relevance are efficiently accounted for with low precision methods.

Section 3.2 formally defines the general object tracking problem, Section 3.3 introduces the proposed hybrid system model implementation of the priority-

based EOT framework, Section 3.4 provides the details of the existing, constituent tracking algorithms chosen to make up each mode of the hybrid framework for this work, Section 3.5 defines the hybrid system mode transitions, Section 3.6 presents a discussion of simulation results, and finally 3.7 provides some concluding remarks.

3.2 Object Tracking Problem Formulation

The goal of general multi-object tracking is to estimate the full latent object state history, $X_{1:K}$, of an unknown number of maneuvering objects, N_O, from a history of noisy observations, $Z_{1:K}$, without knowledge of object controls or intent. Probabilistic inference provides a rigorous means for accounting for the many sources of uncertainty in the problem, deeming it a valuable tool for multi-object tracking. Specifically, rather than estimating the latent variables directly, inference methods estimate the joint posterior probability distribution over the latent variables conditioned on the observations,

$$p\left(X_{1:K}^{1:N_O}|Z_{1:K}\right) = \prod_{n=1}^{N_O} p\left(X_{1:K}^{n}|Z_{1:K}, X_{1:K}^{1:n\text{-}1}\right) \tag{3.1}$$

from which optimal estimates of state trajectories, $\hat{X}_{1:K}^{1:N_O}$, can be computed via existing techniques, such as Minimum Mean Square Error (MMSE) or Maximum-a-Posteriori (MAP).

In most practical applications, it is accurate to model the objects as being mutually independent, $p\left(X_{1:K}^{n}|X_{1:K}^{m}\right) = p\left(X_{1:K}^{n}\right) \forall n \neq m$, which is convenient in that it simplifies the full joint multi-object tracking problem into the product of

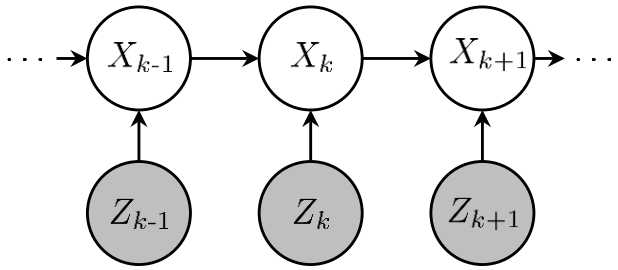

Figure 3.2: Graphical representation of the Hidden Markov model (HMM) used to represent the single-object tracking problem. Shaded nodes denote observed variables, and unshaded nodes denote the hidden, i.e. latent, variables to be estimated.

single-object marginals that can be studied independently in parallel:

$$p\left(X_{1:K}^{1:N_O}|Z_{1:K}\right) = \prod_{n=1}^{N_O} p\left(X_{1:K}^{n}|Z_{1:K}\right) \tag{3.2}$$

Therefore, theoretical development in object tracking is commonly focused on the single-object tracking problem, i.e. estimating:

$$p\left(X_{1:K}|Z_{1:K}\right) \tag{3.3}$$

Further, the computation of the posterior in (3.3) is made tractable, efficient, and deterministic via the following conditional independence assumptions, which gives rise to the Hidden Markov model (HMM) depicted graphically in Fig. 3.2:

$$\begin{aligned} p\left(Z_k|X_k, Z_\ell\right) &= p\left(Z_k|X_k\right) \;\; \forall \ell \neq k \\ p\left(X_k|X_{k\text{-}1}, X_{k\text{-}\ell}\right) &= p\left(X_k|X_{k\text{-}1}\right) \;\; \forall \ell > 1 \end{aligned} \tag{3.4}$$

Lastly, online tracking applications require state estimates in real-time as data is received, and are often principally concerned with the current object state, rather than the full time-history. Therefore, the inference problem is further simplified to estimating:

$$p\left(X_k|Z_{1:k}\right) \forall k \in \{1, \ldots, K\} \tag{3.5}$$

That is, the marginal distribution over the states at any time, k, is conditioned only on the observation history through k.

Given the HMM of Fig. 3.2, (3.5) can be computed recursively at each time step via the following two step process:

1. *Prediction step*: the posterior state belief at the previous time step, $p(X_{k\text{-}1}|Z_{1:k\text{-}1})$, is propagated forward in time via the prescribed stochastic object dynamics model represented by the transition density, $p(X_{k\text{-}1}, X_k)$. The result is the prior state belief at the current time step: $p(X_k|Z_{1:k\text{-}1})$.
2. *Update step*: the prior state belief at the current time step, $p(X_k|Z_{1:k\text{-}1})$, is updated with the current observation via the prescribed stochastic measurement model represented by the observation density, $p(X_k, Z_k)$. The result is the posterior state belief at the current time step: $p(X_k|Z_{1:k})$.

Estimating (3.5) is commonly referred to as *filtering*, which can be supplemented with *smoothing* to estimate the full posterior distribution over the state history in (3.3).

3.3 Estimation Framework

As discussed in the introduction, human visual perception degrades from a precise metrical representation to a rough ordinal one, as distance from the observer increases; this is a direct consequence of the diminishing quality and availability of ordinal visual cues, and an evolutionary advantage given finite cognitive resources, and significant relative importance of close objects compared to distant ones. In characterizing this phenomenon, cognitive scientists

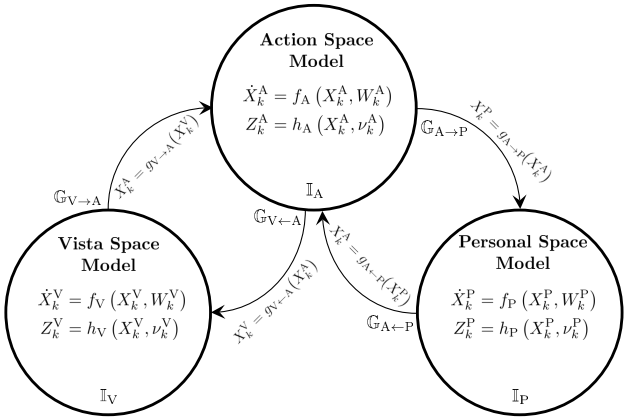

Figure 3.3: Hybrid model of the proposed tracking framework, where $\mathbb{G}$ and $\mathbb{I}$ denote the guards and invariants governing the discrete mode transitions.

have discretized perceptual space into three distinct regions defined by distance from the human observer. Specifically, in order of decreasing distance and improved convergence to a metrical representation: *Vista* space, *Action* space, and *Personal* space. In human trials, the specific distances to the boundaries dividing these regions were found to depend on a variety of variables, including the quality of the observer's vision, and characteristics of the particular scene, e.g. clutter, object familiarity, and scene geometry [17] [16].

Given that robots are often designed to perform human tasks, such as navigation or driving, and there exists natural analogies between finite human cognition and robot computation, as well as the degradation of human and robot sensor information with distance from the observer, a priority-based tracking framework is proposed here, inspired by the aforementioned characteristics of human perception. Fig. 3.3 depicts the hybrid system model chosen to implement the estimation framework, where each discrete mode represents a unique EOT approach, and mode transitions are driven by object relevancy metrics; the specific modal EOT tracking approaches and transition metrics are chosen

such that, as an object becomes increasingly *relevant* to the ego robot task, the tracker transitions along the path: Vista → Action → Personal, causing the overall tracker to move from left to right on the spectrum depicted in Fig. 3.1.

All measurement models discussed in this work correspond to sensors providing (potentially multiple) position/distance returns per query, such as lidar, radar, or binocular/RGB-D cameras. Throughout the paper, an unadorned Z_k denotes the raw position measurement, or set of measurements, at time step k, while superscripted variables, $Z_k^{\mathrm{V/A/P}}$, denote a particular interpretation of the raw data (e.g. metadata, or summary statistics) leveraged by the sensor model associated with the hybrid mode identified in the superscript.

The parameters of the hybrid system model of Fig. 3.3, i.e. the modal tracking methods and object relevancy metrics driving mode transitions, should be chosen to reflect the specific tracking application and goals motivating the use of the proposed priority-based framework. This work focuses on robust tracking of arbitrarily shaped extended objects without *a priori* knowledge of their shape or size for autonomous driving applications; therefore, the chosen hybrid model parameters outlined in the coming sections are chosen for their efficacy in achieving this tracking goal. However, the framework can be easily tailored for other applications or goals via judicial hybrid system parameter selection.

3.4 Modal Tracking Approaches

The modal EOT approaches should be chosen with regard to the particular tracking application, and the goals motivating the use of the proposed priority-based framework. The motivation of this work is to invoke high precision EOT

Table 3.1: Suggested modal tracking methods

Vista Space	**Action Space**	**Personal Space**
Qualitative	*Parametric Bound*	*Nonparametric Surface*
· Temporal occupancy grid [4]	· Circular Disk [8]	·2D Point Cloud [37,38,48–51]
· Markov chain occupancy grid [44]	· Ellipse [7,9,12,30,31]	· 3D Point Cloud [24,25,27]
· Topological	· Rectangle [10,34,40]	
	· Star Convex RHM [11]	

methods for objects that are *relevant* to the ego robot task, and inexpensive EOT methods for objects that are of peripheral relevance. Therefore, the Vista, Action, and Personal space models are chosen from the left, center, and right of the tracking spectrum presented in Fig. 3.1, respectively; a list of suggested tracking approaches for each mode is provided in Table 3.1. The modal tracking approaches chosen from Table 3.1 for this work are presented in detail in the coming sections.

Note that, while beyond the scope of this paper, the hybrid system framework depicted in Fig. 3.3 can also be outfit to address alternative tracking motivations. For instance, consider the example goal of achieving robustness in the tracker. An EOT approach robust to occlusion could be selected when driving in a cluttered environment, e.g. [3, 20, 21, 48, 49, 51], while an alternative approach may prove prudent when the clutter subsides. In heavy traffic, cars could be tracked in groups rather than individually, e.g. [31], or EOT methods

that account for the inherent correlations in the behavior of the traffic participants could be developed; i.e. by omitting the independence assumption leading to (3.2).

3.4.1 Vista Space Model

Vista mode is reserved for objects with the most peripheral significance to the ego robot task; a general awareness of objects in Vista space is useful, but the computational resources required for detailed object tracking are better spent elsewhere. Further, similar to human sensors, robot sensor precision/resolution often degrades with distance from the sensor (e.g. the spatial resolution of a spinning lidar); in these cases, the tracking precision also degrades with distance from the sensor regardless of the chosen tracking algorithm, and thus the benefits of 'high precision' methods are limited. To this end, the perceptual space surrounding the ego-vehicle is abstracted into the eight disjoint discrete qualitative states depicted in Fig. 3.4 (left), the topology of which is encoded in the state transition diagram in Fig. 3.4 (right). Each qualitative state, X_k^{V}, is parameterized by a bearing and a range interval, $\mathbb{B}_{X_k^{\mathrm{V}}}$ and $\mathbb{R}_{X_k^{\mathrm{V}}}$, respectively, defined as:

$$\begin{aligned} \mathbb{B}_{X_k^{\mathrm{V}}} &= \frac{\pi}{2} \cdot \left\{ X_k^{\mathrm{V}} + \left[\; -\tfrac{1}{2}, \quad \tfrac{1}{2} \; \right) \right\} \\ \mathbb{R}_{X_k^{\mathrm{V}}} &= \begin{cases} \left[\; 0, \quad \overline{\rho} \; \right) & \text{if } X_k^{\mathrm{V}} \in \text{Near} \\ \left[\; \overline{\rho}, \quad \infty \right) & \text{if } X_k^{\mathrm{V}} \in \text{Far} \end{cases} \end{aligned} \tag{3.6}$$

where $\overline{\rho}$ is the user-defined range boundary between 'Near' and 'Far'; i.e. the circle in Fig. 3.4 (left).

Qualitative state representations have gained interest in robotic applications,

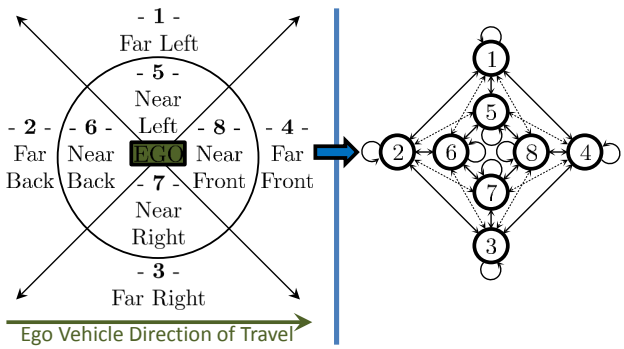

Figure 3.4: Qualitative abstraction of the perceptual space of the ego robot. **Left:** Eight qualitative discrete states comprised of two range sets and four bearing quadrants. **Right:** Graphical model representation of the available transitions between the qualitative states of the discrete abstraction. Dashed arrows denote transitions enabled by the discrete time nature of the filter driven by the finite temporal resolution of the sensor.

such as relational mapping [36], due to their efficiency, scalability, and natural synergy with inexpensive ordinal sensor information, such as monocular cameras or human input. The qualitative states depicted in Fig. 3.4 are chosen because of their similarity to human accounting of objects in the Vista space of human perception; common robotic sensors provide this information directly (i.e. bearing and range), eliminating the need for intricate interpretations of the data, such as reasoning about object shapes and surfaces. The object state in Vista space, X_k^{V}, is then an integer denoting the qualitative state of the object at time k, the belief of which, $p\left(X_k^{\mathrm{V}}|Z_{1:k}\right)$, is distributed according to a categorical distribution.

Belief Prediction

Minding the conditional independence rules defined in (3.4), the posterior categorical distribution over the object state at time $k-1$ is predicted forward to

time k as follows:

$$\begin{aligned} p\left(X_k^{\mathrm{V}}|Z_{1:k\text{-}1}\right) &= \sum_{X_{k\text{-}1}^{\mathrm{V}}} p\left(X_{k\text{-}1}^{\mathrm{V}}, X_k^{\mathrm{V}}|Z_{1:k\text{-}1}\right) \\ &= \sum_{X_{k\text{-}1}^{\mathrm{V}}} p\left(X_k^{\mathrm{V}}|X_{k\text{-}1}^{\mathrm{V}}\right) p\left(X_{k\text{-}1}^{\mathrm{V}}|Z_{1:k\text{-}1}\right) \end{aligned} \tag{3.7}$$

where $p\left(X_{k\text{-}1}^{\mathrm{V}}|Z_{1:k\text{-}1}\right)$ is the posterior state distribution at time $k-1$, and $p\left(X_k^{\mathrm{V}}|X_{k\text{-}1}^{\mathrm{V}}\right)$ is the symmetric discrete state transition density defined over the transition graph on the right of Fig. 3.4:

$$p\left(X_k^{\mathrm{V}}|X_{k\text{-}1}^{\mathrm{V}}\right) = \frac{\mathcal{L}\left(X_k^{\mathrm{V}}|X_{k\text{-}1}^{\mathrm{V}}\right)}{\sum\limits_{X_k^{\mathrm{V}}} \mathcal{L}\left(X_k^{\mathrm{V}}|X_{k\text{-}1}^{\mathrm{V}}\right)} \tag{3.8}$$

where the conditional likelihood function is defined as:

$$\mathcal{L}\left(X_k^{\mathrm{V}} = \iota|X_{k\text{-}1}^{\mathrm{V}} = \jmath\right) = \begin{cases} \mathcal{L}_{\boldsymbol{I}} & \text{if } \iota = \jmath \\ \mathcal{L}_{\bowtie} & \text{if } \iota \bowtie \jmath \\ \mathcal{L}_{\tilde{\bowtie}} & \text{if } \iota \,\tilde{\bowtie}\, \jmath \\ 0 & \text{if } \iota \neg\bowtie \jmath \end{cases} \tag{3.9}$$

and $\bowtie$, $\neg\bowtie$, and $\tilde{\bowtie}$ denote adjacency, non-adjacency, and diagonal adjacency of qualitative states (Fig. 3.4 left), and appear as solid, missing, and dashed edges between graph nodes in Fig. 3.4 (right), respectively. When applied to the graph in Fig. 3.4, the conditional distribution in (3.8) is depicted as the following symmetric, positive definite matrix:

$$p\left(X_k^{\mathrm{V}}|X_{k\text{-}1}^{\mathrm{V}}\right) = \overbrace{\left.\begin{bmatrix} p_{\boldsymbol{I}} & p_{\bowtie} & 0 & p_{\bowtie} & p_{\bowtie} & p_{\tilde{\bowtie}} & 0 & p_{\tilde{\bowtie}} \\ p_{\bowtie} & p_{\boldsymbol{I}} & p_{\bowtie} & 0 & p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\tilde{\bowtie}} & 0 \\ 0 & p_{\bowtie} & p_{\boldsymbol{I}} & p_{\bowtie} & 0 & p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\tilde{\bowtie}} \\ p_{\bowtie} & 0 & p_{\bowtie} & p_{\boldsymbol{I}} & p_{\tilde{\bowtie}} & 0 & p_{\tilde{\bowtie}} & p_{\bowtie} \\ p_{\bowtie} & p_{\tilde{\bowtie}} & 0 & p_{\tilde{\bowtie}} & p_{\boldsymbol{I}} & p_{\bowtie} & 0 & p_{\bowtie} \\ p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\tilde{\bowtie}} & 0 & p_{\bowtie} & p_{\boldsymbol{I}} & p_{\bowtie} & 0 \\ 0 & p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\tilde{\bowtie}} & 0 & p_{\bowtie} & p_{\boldsymbol{I}} & p_{\bowtie} \\ p_{\tilde{\bowtie}} & 0 & p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\bowtie} & 0 & p_{\bowtie} & p_{\boldsymbol{I}} \end{bmatrix}\right\} X_{k\text{-}1}^{\mathrm{V}}}^{X_k^{\mathrm{V}}} \tag{3.10}$$

where:

$$p_{(\cdot)} = \frac{\mathcal{L}_{(\cdot)}}{\mathcal{L}_{\boldsymbol{I}} + 3\mathcal{L}_{\bowtie} + 2\mathcal{L}_{\tilde{\bowtie}}} \tag{3.11}$$

Conceptually, the likelihoods, $\mathcal{L}_{(\cdot)}$, can be set according to the relative area of the boundary associated with each type of transition, giving: $\mathcal{L}_{\boldsymbol{I}} > \mathcal{L}_{\bowtie} > \mathcal{L}_{\tilde{\bowtie}}$.

Belief Update

Minding the conditional independence rules defined in (3.4), the prior categorical distribution at time k, (3.7), is updated to reflect the observation at time k via the following equation:

$$\begin{aligned} p\left(X_k^{\mathrm{V}}|Z_{1:k}\right) &= \frac{p\left(X_k^{\mathrm{V}}, Z_k|Z_{1:k\text{-}1}\right)}{p\left(Z_k|Z_{1:k\text{-}1}\right)} \\ &= \frac{p\left(Z_k|X_k^{\mathrm{V}}\right) \cdot p\left(X_k^{\mathrm{V}}|Z_{1:k\text{-}1}\right)}{\sum_{X_k^{\mathrm{V}}=1}^{8} p\left(Z_k|X_k^{\mathrm{V}}\right) \cdot p\left(X_k^{\mathrm{V}}|Z_{1:k\text{-}1}\right)} \end{aligned} \tag{3.12}$$

where $p\left(X_k^{\mathrm{V}}|Z_{1:k\text{-}1}\right)$ is the prior computed in (3.7). The conditional measurement likelihood, $p\left(Z_k|X_k^{\mathrm{V}}\right)$, is found by counting the sensor returns from the discrete

region of space corresponding to X_k^{V}, parameterized by the bearing and range intervals defined in (3.6):

$$p\left(Z_k|X_k^{\mathrm{V}}\right) = \sum_{\ell=1}^{n_k^z} \left(\beta_{z_\ell} \in \mathbb{B}_{X_k^{\mathrm{V}}}\right) \cap \left(\rho_{z_\ell} \in \mathbb{R}_{X_k^{\mathrm{V}}}\right) \tag{3.13}$$

where β_{z_ℓ} and ρ_{z_ℓ} denote the bearing and range to sensor return $z_\ell \in Z_k \ \forall \ell \in \left\{1, \ldots, n_k^z\right\}$.

3.4.2 Action Space Model

Action mode is reserved for objects of increasing significance to the ego robot task, i.e. they have a substantial impact on perception related routines such as anticipatory path planning [22, 23]. Therefore, a reasonable estimate of the object's position, velocity, and approximate size are desired. To this end, the extended object tracking approach chosen for the action space is the random matrix method introduced in [30] and studied further in [31] and [7]. For the random matrix approach, the object state in action space at time k, X_k^{A}, is defined as a random vector representing the objects position and velocity in the motion plane:

$$X_k^{\mathrm{A}} = \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{bmatrix}_k \tag{3.14}$$

and the object extent in the motion plane is modeled as an ellipse by way of a symmetric positive definite random matrix, $\boldsymbol{E}_k$:

$$\boldsymbol{E}_k = \begin{bmatrix} e_x & e_{x,y} \\ e_{y,x} & e_y \end{bmatrix}_k \tag{3.15}$$

The tracking problem in (3.5) is then to estimate the joint distribution over the object state and elliptical extent given the history of measurements, $Z_{1:k}$:

$$p\left(X_k^{\mathrm{A}}, \boldsymbol{E}_k | Z_{1:k}\right) \tag{3.16}$$

The joint distribution in (3.16) can be factored exactly into the product of vector and matrix valued distributions:

$$p\left(X_k^{\mathrm{A}}, \boldsymbol{E}_k | Z_{1:k}\right) = p\left(X_k^{\mathrm{A}} | \boldsymbol{E}_k, Z_{1:k}\right) p\left(\boldsymbol{E}_k | Z_{1:k}\right) \tag{3.17}$$

where $p\left(X_k^{\mathrm{A}} | \boldsymbol{E}_k, Z_{1:k}\right)$ is the vector valued distribution over the object state, modeled as a multivariate Gaussian, and $p\left(\boldsymbol{E}_k | Z_{1:k}\right)$ is the matrix valued distribution over the elliptical object extent, modeled to be inverse Wishart:

$$\begin{aligned} p\left(X_k^{\mathrm{A}} | \boldsymbol{E}_k, Z_{1:k}\right) &= \mathcal{N}\left(\bar{X}_{k|k}^{\mathrm{A}}, \boldsymbol{P}_{k|k}^{\mathrm{A}}\right) \\ p\left(\boldsymbol{E}_k | Z_{1:k}\right) &= \mathcal{W}^{-1}(\boldsymbol{\Psi}_{k|k}, \alpha_{k|k}) \end{aligned} \tag{3.18}$$

Thus, the posterior distribution in (3.16) is fully specified by the Gaussian mean, $\bar{X}_k^{\mathrm{A}}$, and covariance, $\boldsymbol{P}_k^{\mathrm{A}}$, coupled with the inverse Wishart scale matrix, $\boldsymbol{\Psi}_k$, and degrees of freedom, α_k.

The inverse Wishart distribution serves as the conjugate prior for the covariance matrix of a multivariate Gaussian. Further, the mean, $\bar{\boldsymbol{E}}_k$, variance of the $(\imath, \jmath)^{\text{th}}$ element, $\left(\sigma_{k|k}^{\imath,\jmath}\right)^2$, and covariance between the $(\imath, \jmath)^{\text{th}}$ and $(\ell, m)^{\text{th}}$ elements, $\sigma_{k|k}^{(\imath,\jmath),(\ell,m)}$, of the extent matrix belief, $p\left(\boldsymbol{E}_k | Z_{1:k}\right)$, are computed from the inverse Wishart parameters as:

$$\bar{\boldsymbol{E}}_{k|k} = \frac{\boldsymbol{\Psi}_{k|k}}{\alpha_{k|k} + d - 1} \tag{3.19}$$

$$\begin{aligned} \left(\sigma_{k|k}^{\imath,\jmath}\right)^2 &= \frac{\left(\alpha_{k|k} - d + 1\right)\left(\psi_{k|k}^{\imath,\jmath}\right)^2 + \left(\alpha_{k|k} - d - 1\right)\psi_{k|k}^{\imath,\imath}\psi_{k|k}^{\jmath,\jmath}}{\left(\alpha_{k|k} - d\right)\left(\alpha_{k|k} - d - 1\right)^2\left(\alpha_{k|k} - d - 3\right)} \\ \sigma_{k|k}^{(\imath,\jmath),(\ell,m)} &= \frac{2\psi_{k|k}^{\imath,\jmath}\psi_{k|k}^{\ell,m} + \left(\alpha_{k|k} - d - 1\right)\left(\psi_{k|k}^{\imath,\ell}\psi_{k|k}^{\jmath,m} + \psi_{k|k}^{\imath,m}\psi_{k|k}^{\jmath,\ell}\right)}{\left(\alpha_{k|k} - d\right)\left(\alpha_{k|k} - d - 1\right)^2\left(\alpha_{k|k} - d - 3\right)} \end{aligned} \tag{3.20}$$

where d is the dimension of $\boldsymbol{E}_k$ ($d = 2$ in this case).

Belief Prediction

The kinematic states evolve according to a stochastic continuous time differential equation model of the form in (D.6) provided in Appendix D.1. Therefore, the Kalman filter prediction equations in (D.9) are used to compute the mean and covariance of the prior state belief at time k:

$$p\left(X_k^{\mathrm{A}}|Z_{1:k\text{-}1}\right) = \mathcal{N}\left(\bar{X}_{k|k\text{-}1}^{\mathrm{A}}, \boldsymbol{P}_{k|k\text{-}1}^{\mathrm{A}}\right) \tag{3.21}$$

The dynamics model matrices, $\boldsymbol{F}_k^{\mathrm{A}}$ and $\boldsymbol{G}_k^{\mathrm{A}}$, are defined as a function of the ego vehicle rotation matrix, $\boldsymbol{M}_{\mathrm{g},k}^{\mathrm{ego}}$, and its time-derivatives:

$$\begin{aligned}
\boldsymbol{F}_k^{\mathrm{A}} &= \begin{bmatrix} \mathbf{0}_{2\times 2} & \boldsymbol{I}_{2\times 2} \\ \left(\boldsymbol{M}_{\mathrm{g},k}^{\mathrm{ego}}\right)^T \ddot{\boldsymbol{M}}_{\mathrm{g},k}^{\mathrm{ego}} & -2\left(\boldsymbol{M}_{\mathrm{g},k}^{\mathrm{ego}}\right)^T \dot{\boldsymbol{M}}_{\mathrm{g},k}^{\mathrm{ego}} \end{bmatrix} \\
\boldsymbol{G}_k^{\mathrm{A}} &= \begin{bmatrix} \mathbf{0}_{2\times 2} \\ -\left(\boldsymbol{M}_{\mathrm{g},k}^{\mathrm{ego}}\right)^T \end{bmatrix}
\end{aligned} \tag{3.22}$$

where subscript 'g' indicates that the variable is described in a global coordinate frame, and superscript 'ego' denotes that the variable pertains to the ego vehicle dynamics, and is provided by an independent localization routine on-board the ego vehicle. The ego vehicle rotation matrix and its time-derivatives are defined as:

$$\begin{aligned}
\boldsymbol{M}_{\mathrm{g},k}^{\mathrm{ego}} &= \begin{bmatrix} \cos\phi_{\mathrm{g},k}^{\mathrm{ego}} & -\sin\phi_{\mathrm{g},k}^{\mathrm{ego}} \\ \sin\phi_{\mathrm{g},k}^{\mathrm{ego}} & \cos\phi_{\mathrm{g},k}^{\mathrm{ego}} \end{bmatrix} \\
\dot{\boldsymbol{M}}_{\mathrm{g},k}^{\mathrm{ego}} &= \frac{\partial \boldsymbol{M}_{\mathrm{g},k}^{\mathrm{ego}}}{\partial \phi_{\mathrm{g},k}^{\mathrm{ego}}} \dot{\phi}_{\mathrm{g},k}^{\mathrm{ego}} \\
\ddot{\boldsymbol{M}}_{\mathrm{g},k}^{\mathrm{ego}} &= \frac{\partial \boldsymbol{M}_{\mathrm{g},k}^{\mathrm{ego}}}{\partial \phi_{\mathrm{g},k}^{\mathrm{ego}}} \ddot{\phi}_{\mathrm{g},k}^{\mathrm{ego}} - \boldsymbol{M}_{\mathrm{g},k}^{\mathrm{ego}} \left(\dot{\phi}_{\mathrm{g},k}^{\mathrm{ego}}\right)^2
\end{aligned} \tag{3.23}$$

The process noise represents the object acceleration in the ego reference frame, which is assumed to be driven by the following Gaussian white noise process:

$$W_k^{\mathrm{A}} \sim \mathcal{N}\left(\begin{bmatrix} \ddot{x}_{\mathrm{g},k}^{\mathrm{ego}} \\ \ddot{y}_{\mathrm{g},k}^{\mathrm{ego}} \end{bmatrix}, \boldsymbol{Q}_k^{\mathrm{A}}\right) \tag{3.24}$$

The elliptical object extent evolves according to the following rotation, accounting for the changing perspective of the ego-vehicle:

$$\boldsymbol{E}_k = \boldsymbol{M}_{\mathrm{g},\Delta k}^{\mathrm{ego}} \boldsymbol{E}_{k-1} \left(\boldsymbol{M}_{\mathrm{g},\Delta k}^{\mathrm{ego}}\right)^T \tag{3.25}$$

where $\boldsymbol{M}_{\mathrm{g},\Delta k}^{\mathrm{ego}}$ is a rotation matrix accounting for the change in orientation of the ego-vehicle from $k-1$ to k:

$$\boldsymbol{M}_{\mathrm{g},\Delta k}^{\mathrm{ego}} = \begin{bmatrix} \cos\left(\phi_{\mathrm{g},k}^{\mathrm{ego}} - \phi_{\mathrm{g},k\text{-}1}^{\mathrm{ego}}\right) & -\sin\left(\phi_{\mathrm{g},k}^{\mathrm{ego}} - \phi_{\mathrm{g},k\text{-}1}^{\mathrm{ego}}\right) \\ \sin\left(\phi_{\mathrm{g},k}^{\mathrm{ego}} - \phi_{\mathrm{g},k\text{-}1}^{\mathrm{ego}}\right) & \cos\left(\phi_{\mathrm{g},k}^{\mathrm{ego}} - \phi_{\mathrm{g},k\text{-}1}^{\mathrm{ego}}\right) \end{bmatrix} \tag{3.26}$$

Therefore, the parameters of the inverse Wishart distribution are predicted over the time interval, δt, using the following equations:

$$\begin{aligned} \boldsymbol{\Psi}_{k|k\text{-}1} &= \boldsymbol{M}_{\mathrm{g},\Delta k}^{\mathrm{ego}} \boldsymbol{\Psi}_{k\text{-}1|k\text{-}1} \left(\boldsymbol{M}_{\mathrm{g},\Delta k}^{\mathrm{ego}}\right)^T \\ \alpha_{k|k\text{-}1} &= \exp\left(\frac{-\delta t}{\tau}\right)(\alpha_{k\text{-}1|k\text{-}1} - 2) + 2 \end{aligned} \tag{3.27}$$

where τ is the user-defined time constant governing the rate of change of the object extent.

Belief Update

The measurement of objects in action space is defined as an observation of the object centroid:

$$Z_k^{\mathrm{A}} = \frac{1}{n_k^z} \sum_{\ell=1}^{n_k^z} z_k^{\ell} \tag{3.28}$$

where $z_k^\ell \ \forall \ell \in \left\{1, \ldots, n_k^z\right\}$ are the individual raw sensor returns at time k. The measurement in (3.28) is modeled as:

$$\begin{aligned} Z_k^{\mathrm{A}} &= \boldsymbol{H} X_k^{\mathrm{A}} + \nu_k \\ \boldsymbol{H} &= [\boldsymbol{I}_2, \boldsymbol{0}_2] \end{aligned} \tag{3.29}$$

where the measurement noise is defined as:

$$\begin{aligned} \nu_k &\sim \mathcal{N}\left(0_2, \bar{\boldsymbol{R}}_k\right) \\ \bar{\boldsymbol{R}}_{k|k\text{-}1} &= \boldsymbol{\Psi}_{k|k\text{-}1} + \boldsymbol{R}_k^{\mathrm{A}} \end{aligned} \tag{3.30}$$

where $\boldsymbol{R}_k^{\mathrm{A}}$ is the measurement noise covariance of each individual sensor return, which is typically provided by the sensor specification. Notice that the sensor uncertainty reflected in (3.30) is bloated by the object extent scale matrix, $\boldsymbol{\Psi}_{k|k\text{-}1}$, thus tracking precision degrades for large objects.

Given the sensor model in (3.29), the measurement related parameters of the joint Gaussian distribution over the object state and measurement in (D.11) are given as:

$$\begin{aligned} \bar{Z}_k^{\mathrm{A}} &= \boldsymbol{H} \bar{X}_{k|k\text{-}1}^{\mathrm{A}} \\ \boldsymbol{P}_{Z_k}^{\mathrm{A}} &= \boldsymbol{H} \boldsymbol{P}_{k|k\text{-}1}^{\mathrm{A}} \boldsymbol{H}^T + \frac{1}{n_k^z} \bar{\boldsymbol{R}}_{k|k\text{-}1} \\ \boldsymbol{P}_{X_{k|k\text{-}1} Z_k}^{\mathrm{A}} &= \boldsymbol{P}_{k|k\text{-}1}^{\mathrm{A}} \boldsymbol{H}^T \end{aligned} \tag{3.31}$$

and the Kalman filter equations in (D.13), provided in Appendix D.1, are used to compute the posterior distribution over the object state in (3.17) and (3.18):

$$p\left(X_k^{\mathrm{A}} | \boldsymbol{E}_k, Z_{1:k}\right) = \mathcal{N}\left(\bar{X}_{k|k}^{\mathrm{A}}, \boldsymbol{P}_{k|k}^{\mathrm{A}}\right) \tag{3.32}$$

The parameters of the inverse Wishart distribution are updated to get the posterior distribution over the object extent in (3.17) and (3.18),

$$p\left(\boldsymbol{E}_k | Z_{1:k}\right) = \mathcal{W}^{-1}\left(\boldsymbol{\Psi}_{k|k}, \alpha_{k|k}\right) \tag{3.33}$$

using the following equations:

$$\begin{aligned} \boldsymbol{\Psi}_{k|k} &= \frac{1}{\alpha_{k|k}} \left(\alpha_{k|k\text{-}1} \boldsymbol{\Psi}_{k|k\text{-}1} + \hat{\boldsymbol{N}}_{k|k\text{-}1} + \hat{\boldsymbol{\Sigma}}_{k|k\text{-}1} \right) \\ \alpha_{k|k} &= \alpha_{k|k\text{-}1} + n_k^z \end{aligned} \tag{3.34}$$

where:

$$\begin{aligned} \hat{\boldsymbol{N}}_{k|k\text{-}1} &= \Upsilon_{k|k\text{-}1} \boldsymbol{N}_{k|k\text{-}1} \Upsilon_{k|k\text{-}1}^T \\ \hat{\boldsymbol{\Sigma}}_{k|k\text{-}1} &= \Xi_{k|k\text{-}1} \boldsymbol{\Sigma}_k \Xi_{k|k\text{-}1}^T \end{aligned} \tag{3.35}$$

and:

$$\begin{aligned} \boldsymbol{N}_{k|k\text{-}1} &= \left(Z_k^{\mathrm{A}} - \bar{Z}_{k|k\text{-}1}^{\mathrm{A}} \right) \left(Z_k^{\mathrm{A}} - \bar{Z}_{k|k\text{-}1}^{\mathrm{A}} \right)^T \\ \boldsymbol{\Sigma}_k &= \sum_{\ell=1}^{n_k^z} \left(z_k^\ell - Z_k^{\mathrm{A}} \right) \left(z_k^\ell - Z_k^{\mathrm{A}} \right)^T \\ \Upsilon_{k|k\text{-}1} &= \boldsymbol{\Psi}_{k|k-1}^{\frac{1}{2}} \left(\boldsymbol{P}_{Z_k}^{\mathrm{A}} \right)^{-\frac{1}{2}} \\ \Xi_{k|k\text{-}1} &= \boldsymbol{\Psi}_{k|k-1}^{\frac{1}{2}} \bar{\boldsymbol{R}}_{k|k\text{-}1}^{-\frac{1}{2}} \end{aligned} \tag{3.36}$$

Matrix square roots when computing $\Upsilon_{k|k\text{-}1}$ and $\Xi_{k|k\text{-}1}$ in (3.36) are computed via the Cholesky factorization.

3.4.3 Personal Space Model

Personal space is reserved for objects with extreme relevance to the ego robot task. For this reason, precise estimates of object kinematics and occupied space are critical for safely interacting with objects in personal space, deeming the tracking approaches chosen for Vista and Actions spaces insufficient.

The personal space object state is defined as the position, velocity, and orientation of the object relative to some arbitrary initial orientation, described in

a coordinate frame fixed to the ego-vehicle centroid:

$$X_k^{\mathrm{P}} = \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \\ \phi \end{bmatrix}_k \tag{3.37}$$

Note that the orientation, ϕ, is decoupled from the object heading, $\tan^{-1}(\dot{x}/\dot{y})$, to accommodate arbitrary objects with a variety of latent motion constraints.

The object extent, χ_k, is modeled as the most recent lidar scan returned from the object at each time step,

$$p(\chi_k|Z_{1:k}) = \mathcal{N}(Z_k, \boldsymbol{R}_k) \tag{3.38}$$

effectively maintaining a detailed, non-parametric, representation of the *immediately visible* object surface over a single time step.

Belief Prediction

Given that the first four object states in (3.37) are identical to X_k^{A} given in (3.14), they evolve according to the same model. The additional state, the orientation of the object, evolves according to the following scalar differential equation:

$$\dot{\phi}_k = \dot{\phi}_{\mathrm{g},k} - \dot{\phi}_{\mathrm{g},k}^{\mathrm{ego}} \tag{3.39}$$

Therefore, the parameters of the prior distribution at k,

$$p\left(X_k^{\mathrm{P}}|Z_{1:k\text{-}1}\right) = \mathcal{N}\left(\bar{X}_{k|k\text{-}1}^{\mathrm{P}}, \boldsymbol{P}_{k|k\text{-}1}^{\mathrm{P}}\right) \tag{3.40}$$

can be computed using the Kalman filter equations in (D.9), provided in Appendix D.1, using the following model:

$$\begin{aligned} \boldsymbol{F}_k^{\mathrm{P}} &= \begin{bmatrix} \boldsymbol{F}_k^{\mathrm{A}} & \mathbf{0}_{4\times 1} \\ \mathbf{0}_{1\times 4} & 0 \end{bmatrix} \\ \boldsymbol{G}_k^{\mathrm{P}} &= \begin{bmatrix} \boldsymbol{G}_k^{\mathrm{A}} & \mathbf{0}_{2\times 1} \\ \mathbf{0}_{1\times 2} & -1 \end{bmatrix} \\ W_k^{\mathrm{P}} &\sim \mathcal{N}\left(\bar{W}_k^{\mathrm{P}}, \boldsymbol{Q}_k^{\mathrm{P}}\right) \end{aligned} \tag{3.41}$$

where:

$$\begin{aligned} \bar{W}_k^{\mathrm{P}} &= \begin{bmatrix} \bar{W}_k^{\mathrm{A}} \\ \dot{\phi}_{\mathrm{g},k}^{\mathrm{ego}} \end{bmatrix} \\ \boldsymbol{Q}_k^{\mathrm{P}} &= \begin{bmatrix} \boldsymbol{Q}_k^{\mathrm{A}} & \mathbf{0}_{2\times 1} \\ \mathbf{0}_{1\times 2} & q_{\dot{\phi}_{\mathrm{g},k}} \end{bmatrix} \end{aligned} \tag{3.42}$$

The object is assumed to be a rigid body, thus the object extent, χ_k, is propagated forward in time via the following rigid body transform:

$$\chi_k = \boldsymbol{M}_{\Delta k}^{\mathrm{Block}}\left(\chi_{k\text{-}1} - T_{k\text{-}1}^{\mathrm{Block}}\right) + T_k^{\mathrm{Block}} \tag{3.43}$$

where:

$$\begin{aligned} \boldsymbol{M}_{\Delta k}^{\mathrm{Block}} &= \boldsymbol{I}_{n_{k-1}^z} \otimes \begin{bmatrix} \cos\left(\phi_k - \phi_{k-1}\right) & -\sin\left(\phi_k - \phi_{k-1}\right) \\ \sin\left(\phi_k - \phi_{k-1}\right) & \cos\left(\phi_k - \phi_{k-1}\right) \end{bmatrix} \\ T_k^{\mathrm{Block}} &= \mathbf{1}_{n_{k-1}^z \times 1} \otimes \begin{bmatrix} x_k \\ y_k \end{bmatrix} \end{aligned} \tag{3.44}$$

$\otimes$ denotes the kronecker product, and (x_k, y_k, ϕ_k) refer to the position and orientation object states in (3.37). The parameters of the prior distribution over the object extent at k,

$$p\left(\chi_k | Z_{1:k\text{-}1}\right) = \mathcal{N}\left(\bar{\chi}_{k|k\text{-}1}, \boldsymbol{P}_{\chi_{k|k\text{-}1}}\right) \tag{3.45}$$

can be computed with the Sigma Point Transform [28].

Belief Update

The measurement of objects in personal space is inspired by [37,38], and defined as an observation of the extremities of the object:

$$Z_k^{\mathrm{P}} = \begin{bmatrix} \beta_k^{\mathrm{cw}} \\ \beta_k^{\mathrm{ccw}} \\ \rho_k \end{bmatrix} \tag{3.46}$$

where β_k^{cw}, β_k^{ccw}, and ρ_k denote the clockwise and counterclockwise most bearings, and the minimum range to the object. The measurement model corresponding to (3.46) is defined as:

$$Z_k^{\mathrm{P}} = h\left(\chi_k\right) + \nu_k \tag{3.47}$$

where χ_k is the extent model, and $h\left(\cdot\right)$ is a function extracting the measurement metadata in (3.46) from χ_k.

The measurement related parameters of the joint distribution over the measurement metadata, Z_k^{P}, and object state, X_k^{P}, in (D.11), specifically, $\bar{Z}_k^{\mathrm{P}}$, $\boldsymbol{P}_{Z_k}^{\mathrm{P}}$, and $\boldsymbol{P}_{X_{k|k\text{-}1}Z_k}^{\mathrm{P}}$, are computed from the prior distributions over the state, (3.40), and object extent, (3.45), using the Sigma Point Transform [28]. Finally, the parameters of the posterior state distribution,

$$p\left(X_k^{\mathrm{P}}|Z_{1:k}\right) = \mathcal{N}\left(\bar{X}_{k|k}^{\mathrm{P}}, \boldsymbol{P}_{k|k}^{\mathrm{P}}\right) \tag{3.48}$$

are computed with the Kalman filter update equations in (D.13), and the distribution over the object extent is updated with (3.38); i.e. replacing the prior extent belief with the most recent lidar scan.

Table 3.2: Example relevance-based metrics

***Relevance* definition**	**Associated Metric**
· Proximity	· Distance to object
· Danger	· Probability of collision
· Anomalous/erratic behavior	· χ^2 test on tracker innovations
· Object of interest	· Object recognition probability

3.5 Mode Transitions

The mode transitions among, Vista, Action, and Personal spaces, depicted in Fig. 3.3, are fully defined by their guards, $\mathbb{G}$, invariants, $\mathbb{I}$, and state transition functions of the form, $X_k^{\iota} = g_{\jmath\to\iota}\left(X_k^{\jmath}\right)$, which transform the state belief from the source mode representation, $\jmath$, to that of the destination mode, ι. Generally, metrics informing the guards and invariants should be chosen with regard to the tracking application, as well as the specific goals motivating the use of the proposed priority-based framework. The motivation of this work is to invoke high precision EOT methods for objects that are *relevant* to the ego robot task, and inexpensive EOT methods for objects of peripheral relevance. Therefore, the metrics informing mode transitions must serve as a measure of the chosen definition of *object relevance*; a list of example *relevance* definitions and suggestions for their associated metrics is provided in Table 3.2. The definitions and metrics chosen for this work are presented in the following sections along with their associated guards, invariants, and mode transition functions.

Note that, while beyond the scope of this paper, the hybrid system framework depicted in Fig. 3.3 can also be outfit to transition according to *ability*-based (or other) metrics, rather than object relevance. For instance, general ob-

ject trackers, such as those discussed in the introduction, can be leveraged at object track initialization when specific static attributes of the object, such as object type or class, are unavailable; then, as estimates of the static object attributes converge, the system can transition to more specific, ad hoc, trackers designed to leverage information inferred from the estimated object attribute. Alternatively, a unique synergy may exist between EOT approaches and available senor types; for instance, dense 3D colored point cloud approaches, e.g. [24, 25, 27], perform well in regions of space where the field-of-view (fov) of a color camera intersects the fov of one (or more) lidar sensor(s). In cases such as this, transitions can be triggered as objects enter and exit the fov of different sensors, or areas of multi-sensor overlap, leveraging the identified synergistic sensor-tracker pairings.

3.5.1 Vista ↔ Action

Guards

The relevancy metric governing the transitions among Vista and Action modes is chosen as the probability that the object is 'Far' from the sensor, $p\left(\text{Far}_k|Z_{1:k}\right)$, defined in (3.50). Specifically, the guards are defined as:

$$\begin{aligned} \mathbb{G}_{\text{V}\rightarrow\text{A}} &\triangleq p\left(\text{Far}_k|Z_{1:k}\right) \leq p_{\text{V,A}} \\ \mathbb{G}_{\text{V}\leftarrow\text{A}} &\triangleq p\left(\text{Far}_k|Z_{1:k}\right) > p_{\text{A,V}} \end{aligned} \tag{3.49}$$

where $p_{\text{A,V}} \geq p_{\text{V,A}}$ are user-defined probability thresholds.

For the purposes of this work, 'Far' is defined by the discrete abstraction in Fig. 3.4, as the object occupying any of the first four qualitative states. Therefore,

the far probability metric is computed as:

$$\begin{aligned} p\left(\text{Far}_k|Z_{1:k}\right) &= p\left(\bigcup_{X_k^{\text{V}}=1}^{4} X_k^{\text{V}} \middle| Z_{1:k}\right) \\ &= \sum_{X_k^{\text{V}}=1}^{4} p\left(X_k^{\text{V}}|Z_{1:k}\right) \end{aligned} \tag{3.50}$$

where the equality of the first and second lines of (3.50) is conditioned on the fact that the qualitative states are disjoint. For the transition from Vista to Action, V→A, $p\left(X_k^{\text{V}}|Z_{1:k}\right)$ is the current state belief posterior already computed for the measurement update step. For the reverse transition, V←A, $p\left(X_k^{\text{V}}|Z_{1:k}\right)$ is computed from (3.12) with a uniform prior over the qualitative states, X_k^{V}. Specifically,

$$p\left(X_k^{\text{V}}|Z_{1:k}\right) = \frac{p\left(Z_k|X_k^{\text{V}}\right)}{\sum\limits_{X_k^{\text{V}}} p\left(Z_k|X_k^{\text{V}}\right)} \tag{3.51}$$

where $p\left(Z_k|X_k^{\text{V}}\right)$ is the vista mode measurement likelihood defined in (3.13).

Transition functions

Given that the Vista model, by design, reflects only low fidelity qualitative information about the object position, it does not have much to offer the continuous metrical Action model in terms of a transition function; therefore, at the V→A transition, the Action model is initialized directly from the lidar scan. Specifically, the inverse Wishart parameters of the extent model distribution are computed as:

$$\begin{aligned} \boldsymbol{\Psi}_k &= \frac{1}{n_k^z}\boldsymbol{\Sigma}_k \\ \alpha_k &= n_k^z \end{aligned} \tag{3.52}$$

where $\mathbf{\Sigma}_k$ is defined in (3.36), and n_k^z is the number of measurement returns at time k. The Gaussian parameters of the state distribution are computed as:

$$\bar{X}_k^{\mathrm{A}} = \begin{bmatrix} Z_k^{\mathrm{A}} \\ -\dot{x}_{\mathrm{g},k}^{\mathrm{ego}} \\ -\dot{y}_{\mathrm{g},k}^{\mathrm{ego}} \end{bmatrix} \qquad \boldsymbol{P}_k^{\mathrm{A}} = \begin{bmatrix} \frac{1}{n_k^z}\bar{\boldsymbol{R}}_k & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \boldsymbol{P}_{V_0}^{\mathrm{A}} \end{bmatrix} \tag{3.53}$$

where Z_k^{A} and $\bar{\boldsymbol{R}}_k$ are the mean and covariance of the Action space centroid measurement, defined in (3.28) and (3.30), respectively, and $\boldsymbol{P}_{V_0}^{\mathrm{A}}$ is set to a large diagonal matrix reflecting the large amount of uncertainty in the velocity initialization. Note that the velocity state initialization is naive in assuming the object is static in the global reference frame. However, the linear properties of the dynamics and measurement model in (3.22) and (3.28) allow for a quick estimate convergence from a potentially poor initialization, as demonstrated in the results section. More elaborate initialization schemes can be implemented without loss of generality.

At the V←A transition, the Vista model is initialized from (3.51), which was already computed to evaluate the guard, $\mathbb{G}_{\mathrm{V}\leftarrow\mathrm{A}}$, in (3.49).

3.5.2 Action ↔ Personal

Guards

The relevancy metric governing the transitions among Action and Personal modes is chosen as the anticipated probability of collision with the object over

a defined time horizon, h, $p\left(C_{k:k+h}|Z_{1:k}\right)$. Specifically, the guards are defined as:

$$\begin{aligned}\mathbb{G}_{\mathrm{A}\rightarrow\mathrm{P}} &\triangleq p\left(C_{k:k+h}|Z_{1:k}\right) \geq p_{\mathrm{A,P}} \\ \mathbb{G}_{\mathrm{A}\leftarrow\mathrm{P}} &\triangleq p\left(C_{k:k+h}|Z_{1:k}\right) < p_{\mathrm{P,A}}\end{aligned} \tag{3.54}$$

where $p_{\mathrm{A,P}} \geq p_{\mathrm{P,A}}$ are user-defined probability thresholds.

The anticipated probability of collision, $p\left(C_{k:k+h}|Z_{1:k}\right)$, is taken as the maximum *instantaneous* collision probability over each time step in a horizon, h:

$$p\left(C_{k:k+h}|Z_{1:k}\right) = \max_{\ell\in\{1,\ldots,h\}}\left[p\left(C_{k+\ell}|Z_{1:k}\right)\right] \tag{3.55}$$

Conceptually, the instantaneous collision probabilities, $p\left(C_{k+\ell}|Z_{1:k}\right) \quad \forall\ell \in \{1,\ldots,h\}$, are computed as the probability that the space occupied by the object intersects that of the ego vehicle at each future instant, $k+\ell$. As demonstrated in Fig. 3.5, mathematically this is equivalent to the probability that the ego-vehicle centroid (i.e. the origin of the tracking coordinate frame) lies within the anticipated *collision region*, $\mathcal{O}_{k+\ell}$, defined as the dilation of the uncertain object extent at time $k+\ell$ by the known ego vehicle extent.

Thus the anticipated instantaneous collision probability is calculated as:

$$p\left(C_{k+\ell}|Z_{1:k}\right) = \exp\left(-\frac{1}{2}\left(\bar{D}^{\min}_{\mathcal{O}_{k+\ell|k}}\right)^{T}\boldsymbol{P}^{-1}_{D^{\min}_{\mathcal{O}_{k+\ell|k}}}\bar{D}^{\min}_{\mathcal{O}_{k+\ell|k}}\right) \tag{3.56}$$

where:

$$D^{\min}_{\mathcal{O}_{k+\ell|k}} \sim \mathcal{N}\left(\bar{D}^{\min}_{\mathcal{O}_{k+\ell|k}}, \boldsymbol{P}_{D^{\min}_{\mathcal{O}_{k+\ell|k}}}\right) \tag{3.57}$$

is the vector from the ego vehicle to the closest point in the collision region, $\mathcal{O}_{k+\ell}$.

In the simplest case, object state anticipation over the time horizon is accomplished by iterating over the usual filter prediction step. However, for highly dynamic scenes or large time horizons, the state uncertainty can quickly explode

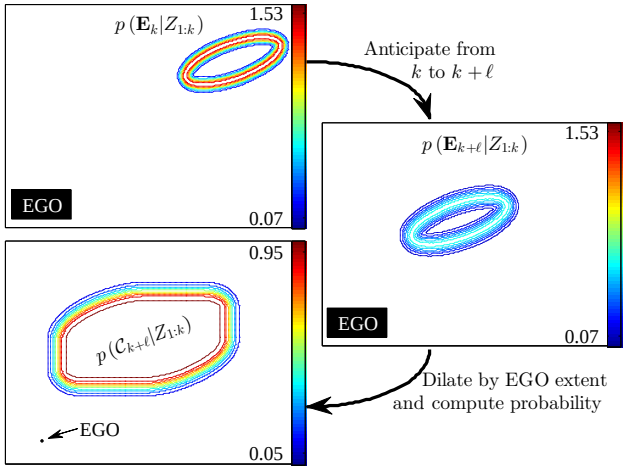

Figure 3.5: Demonstration of the instantaneous collision probability calculation for an object in Action space. **Top:** Belief of the elliptical object extent at time k. **Right:** Anticipated belief of the elliptical object extent at time $k + \ell$. **Bottom:** Probability of collision. Note that the ego-vehicle centroid was swept over the space to generate the probability contours for demonstration purposes, however, in practice, the collision probability only needs to be evaluated at the ego vehicle centroid labeled EGO in the figure.

to produce an uninformative belief. In these cases, it is recommended to leverage more intelligent, specialized anticipation methods that integrate advanced features such as traffic lane following controllers, traffic laws, etc. [22, 23].

Transition functions

Given the similarity of the state representations in (3.14) and (3.37), the state transition functions for A→P is:

$$\begin{aligned} \bar{X}_k^{\mathrm{P}} &= \begin{bmatrix} \bar{X}_k^{\mathrm{A}} \\ 0 \end{bmatrix} \\ \boldsymbol{P}_k^{\mathrm{P}} &= \begin{bmatrix} \boldsymbol{P}_k^{\mathrm{A}} & 0_{4\times 1} \\ 0_{1\times 4} & \epsilon \end{bmatrix} \end{aligned} \tag{3.58}$$

where ϵ is a small positive number indicating perfect knowledge of the initial relative orientation, ϕ_k, while maintaining the positive definite requirement of $\boldsymbol{P}_k^{\mathrm{P}}$; since the extent models do not identify the *front* of the object, ϕ_k is defined as the orientation relative to some arbitrary initialization, and thus can be initialized with absolute certainty to any numerical value. The distribution over the object extent in Personal space, $p\left(\chi_k|Z_{1:k}\right)$, is initialized from the current lidar scan returned from the object via (3.38).

The state transition function from A←P is defined as:

$$\begin{aligned} \bar{X}_k^{\mathrm{A}} &= \bar{X}_k^{\mathrm{P}}\left(1:4\right) \\ \boldsymbol{P}_k^{\mathrm{A}} &= \boldsymbol{P}_k^{\mathrm{P}}\left(1:4,1:4\right) \end{aligned} \tag{3.59}$$

and the inverse Wishart parameters of the distribution over the object extent are initialized from the lidar scan as in (3.52).

3.5.3 Invariants

The hybrid model invariants come directly from the guards already defined for each transition in (3.49) and (3.54). Specifically,

$$\begin{aligned} \mathbb{I}_{\mathrm{V}} &\triangleq p\left(\mathrm{Far}_k|Z_{1:k}\right) > p_{\mathrm{A,V}} \\ \mathbb{I}_{\mathrm{A}} &\triangleq \left(p\left(\mathrm{Far}_k|Z_{1:k}\right) \leq p_{\mathrm{V,A}}\right) \cap \left(p\left(C_{k:k+h}|Z_{1:k}\right) < p_{\mathrm{A,P}}\right) \\ \mathbb{I}_{\mathrm{P}} &\triangleq p_{\mathrm{P,A}} \leq p\left(C_{k:k+h}|Z_{1:k}\right) \end{aligned} \tag{3.60}$$

3.6 Simulation Results

To demonstrate the performance and computational savings enabled by the tracking framework depicted in Fig. 3.3, the proposed hybrid priority-based

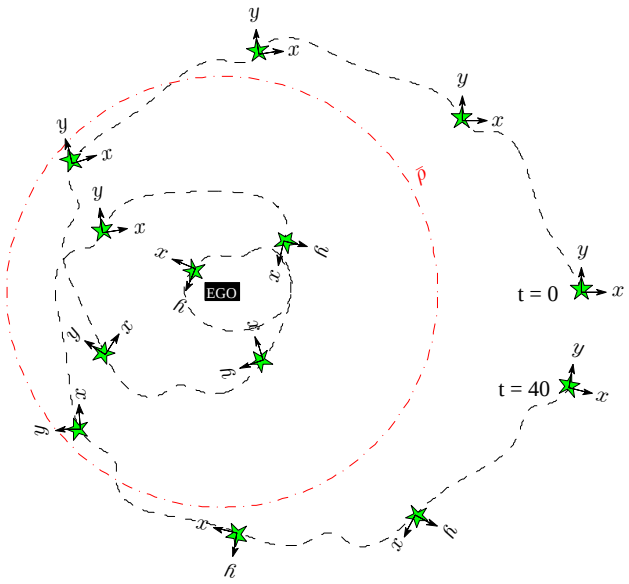

Figure 3.6: Simulated scenario of a star shaped object spiraling in toward the ego-vehicle and then back out over a period of 40 s; 13 time steps are shown.

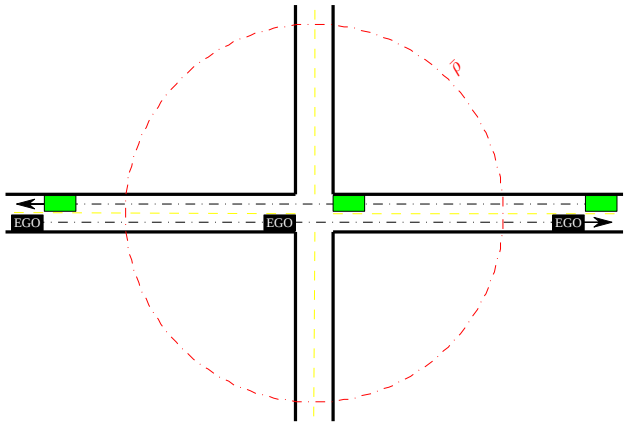

Figure 3.7: Simulated scenario of a rectangular object passing the ego vehicle head-on in close proximity at an intersection; three time steps are shown.

tracker is evaluated for two simulated scenarios depicted in Figs. 3.6 and 3.7.

The scenario depicted in Fig. 3.6 involves a star shaped object maneuvering with continuously variable orientation and velocity along a spiral trajectory centered on the stationary ego vehicle; this scenario is intended to represent

a somewhat arbitrary, unstructured, and challenging tracking application. The scenario depicted in Fig. 3.7 represents a common autonomous driving scenario in which the ego-vehicle and object (modeled as rectangles) pass each other with less than 0.5 m clearance in a four-way controlled intersection. The vehicles initially approach the intersection at a constant cruising speed of 12 m/s ($\approx$ 26.8 mph), decelerate to a full stop at the edge of the intersection, pause for 3 s, then accelerate straight through the intersection until they reach their initial cruising speed.

Data is simulated for a 360° field-of-view planar lidar firing at 12.5 Hz with 0.5° bearing resolution, fixed to the centroid of the ego vehicle. Random sensor noise is sampled independently for each beam in each scan from $\mathcal{N}\left(0, 1\text{cm}^2\right)$ and added to the lidar range returns to simulate the accuracy of realistic lidar sensors. The filter parameters used for both simulations are defined in Table 3.3. All simulations were coded in Matlab with known feature accelerators and code optimizers turned off, and run on a single thread of an Intel® Core™ i7-4770 CPU @ 3.40 GHz. Given that the code is written in an interpreted language and has not been optimized, only discussion about *relative* computational effort among the tracker modes is meaningful.

For this evaluation, *precision*, $\gamma_k^{n\text{-}\sigma}$, is defined as:

$$\gamma_k^{n\text{-}\sigma} = \bar{\mathcal{A}}^{-1}\left(p\left(X_k|Z_{1:k}\right), n\right) \tag{3.61}$$

where $\bar{\mathcal{A}}\left(p\left(X_k|Z_{1:k}\right), n\right)$ denotes the expected value of the area enclosed by the n-σ confidence bound. To obtain a fair comparison with objects in Vista space, the object position belief is taken to be uniformly distributed over the qualitative region represented by the state, and the n-σ confidence bound is interpreted as the area required to enclose the same probability, $p(n)$, as the n-σ confidence

Table 3.3: Hybrid tracking parameters

Mode Transitions		Vista		Action		Personal	
$p_{\mathrm{V,A}}$	0.45	$\overline{\rho}$	30 m	τ	60 s		
$p_{\mathrm{A,V}}$	0.55	$p_{\boldsymbol{I}}$	0.2				
$p_{\mathrm{A,P}}$	0.5	$p_{\bowtie}$	0.1				
$p_{\mathrm{P,A}}$	0.1	$p_{\tilde{\bowtie}}$	0.01				
h	1 s						

bound of a Gaussian, where n-σ refers to the Mahalanobis distance from the mean of the distribution:

$$p(n) = 1 - \exp\left(-\frac{n^2}{2}\right) \tag{3.62}$$

Specifically, for the discrete abstraction in Fig. 3.4:

$$\bar{\mathcal{A}}\Big(p\left(X_k|Z_{1:k}\right), n\Big) = \begin{cases} p(n)\cdot\pi\cdot\frac{\mathbb{E}\left[\rho_{\mathrm{o}}^2-\rho_{\mathrm{i}}^2\right]}{4} & \text{if Vista} \\ n^2\cdot\pi\cdot\sqrt{\left|\boldsymbol{P}_{k|k}^{xy}\right|} & \text{otherwise} \end{cases} \tag{3.63}$$

where ρ_{i} and ρ_{o} denote the inner and outer radii of the qualitative regions associated with the Vista states. The expected value in the numerator of the Vista case of (3.63) becomes:

$$\mathbb{E}\left[\rho_{\mathrm{o}}^2 - \rho_{\mathrm{i}}^2\right] = p\left(X_k \in \mathrm{Far}|Z_{1:k}\right)\cdot\rho_{\max}^2 - \left[2\cdot p\left(X_k \in \mathrm{Far}|Z_{1:k}\right) - 1\right]\cdot\overline{\rho}^2 \tag{3.64}$$

where $\rho_{\max}$ denotes the maximum range of the sensor, $p\left(\mathrm{Far}|Z_{1:k}\right)$ is defined in (3.50). Note that the expression in (3.64) reflects that 'Far' states in Vista space are bounded at the sensor range, $\rho_{\max} = 80$ m; while this is not technically an attribute of the abstraction in Fig. 3.4, it is a sensible bound to avoid infinite area (and infinitesimal precision) given that, inherent in the event that the object returns a sensor measurement, is the fact that the object must be within the range of the sensor.

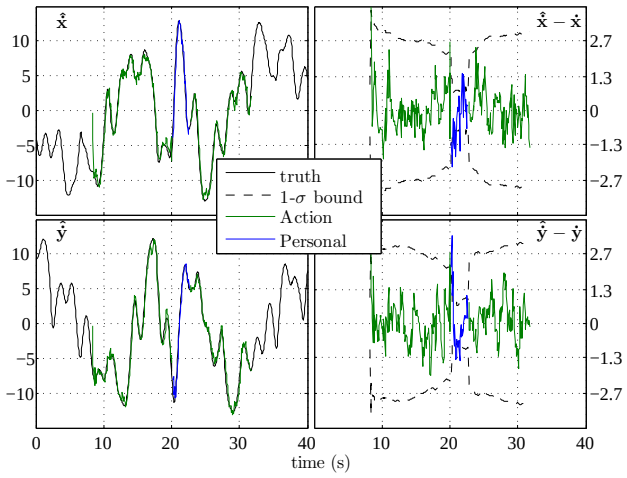

Figure 3.8: Performance in tracking the velocity states for the star scenario: **Left:** MAP estimates overlaid on ground truth, **Right:** tracking error and 1-σ bounds. Note that there is not a concept of velocity in Vista space.

For the purposes of this evaluation, computational *effort*, ϵ_k, is defined as the clock time required to compute each filter recursion, $\delta t_{\text{computation}}$, normalized by the filter time step dictated by the sensor frequency, f_{sensor}:

$$\epsilon_k = \delta t_{\text{computation}} \cdot f_{\text{sensor}} \tag{3.65}$$

Figs. 3.8 and 3.9 demonstrate the tracking performance for the star and intersection scenarios, respectively, by comparing the *maximum-a-posteriori* (MAP) velocity estimates to the simulated truth values. In both scenarios, the filter appears inconsistent (under-confident) when in Action space, i.e. it is overestimating the filter uncertainty. This is an artifact of some over-simplifying assumptions in the object extent model limiting the amount of information that can be extracted from the lidar scan. Specifically, despite the centimeter-level precision of the lidar sensor, the measurement model in (3.29) reflects the naive and highly uncertain expectation that the origin of each lidar return is the centroid of the elliptical extent; a direct consequence of the extent model lacking a concept of

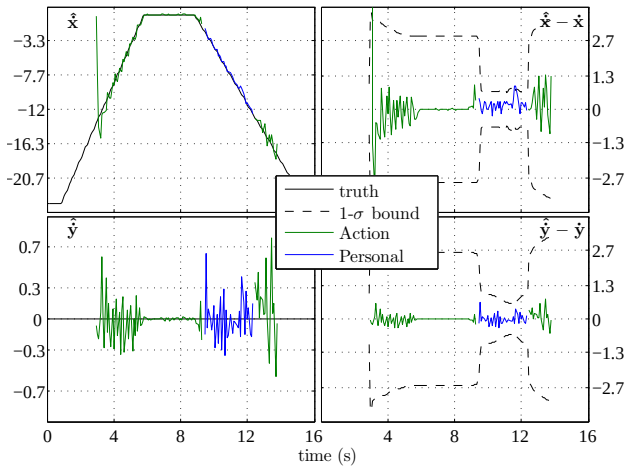

Figure 3.9: Performance in tracking the velocity states for the intersection scenario: **Left:** MAP estimates overlaid on ground truth, **Right:** tracking error and 1-σ bounds. Note that there is not a concept of velocity in Vista space.

object *surface*. Note that this naiveté is also a subtle but critical feature enabling the random matrix approach (Action space) to be computationally simple and efficient while simultaneously remaining flexible and robust in tracking objects from a variety of classes and applications. This measurement origin uncertainty is reflected in the measurement noise model of (3.30), which scales directly with the size of the object extent, and is ultimately the source of the degraded tracking precision in Action space.

Also apparent in Figs. 3.8 and 3.9 is that, as objects approach the ego vehicle, the likelihood that the ego-vehicle may interact with the object increases and the tracker transitions to Personal mode. This transition triggers a dramatic improvement in the estimate uncertainty, which is a critical feature enabling the ego vehicle to safely maneuver in close proximity with uncooperative dynamic objects. In both scenarios, the filter quickly recovers from the naive velocity initialization defined in (3.53) within two time steps (0.16 s) of the filter transition

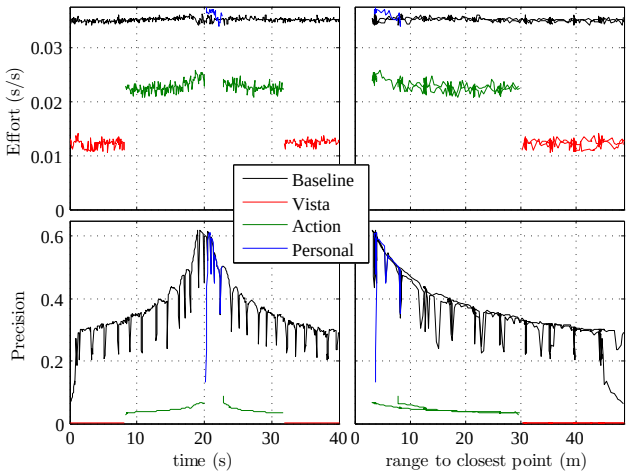

Figure 3.10: Computational effort, ϵ, and precision, γ, as a function of time (left) and range to the closest point on the object (right) for the star scenario.

Table 3.4: Efficiency vs. Precision

	Vista	Action	Personal
Mean Effort, $\bar{\epsilon}$	0.01203	0.02180	0.03653
% of max Effort	32.9 %	59.7 %	100 %
Mean Precision, $\bar{\gamma}$	0.00126	0.04257	0.52372
% of max precision	0.24 %	8.1 %	100 %

from Vista to Action mode. Further, the filter seamlessly transitions between Action and Personal modes in both directions, mitigating the need for elaborate initialization schemes; a direct consequence of the synergy between models.

Figs. 3.10 and 3.11 plot the precision, ϵ defined in (3.61), and computational effort, γ defined in (3.65), over time, for both the star and intersection scenarios, respectively; the combined summary of these metrics is provided in Table 3.4. *Baseline* refers to a tracker that operates solely in Personal mode to emphasize the contribution of the hybrid framework depicted in Fig. 3.3. The Personal

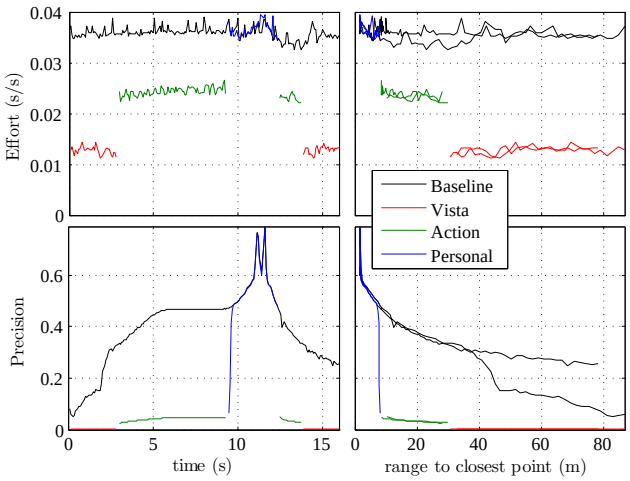

Figure 3.11: Computational effort, ϵ, and precision, γ, as a function of time (left) and range to the closest point on the object (right) for the intersection scenario.

model is used for comparison, as it is the only model of the three that achieves the tracking precision required for interacting with objects in close proximity - a requirement of many robotics applications, including autonomous driving. As designed, both the computational effort and the tracking precision increase as the filter transitions from Vista through Action to Personal mode, and the reciprocal trend exists for transitions in the opposite direction. Specifically, in terms of computational effort, roughly 3 objects can be tracked in Vista mode for every 2 in Action mode, and every 1 in Personal mode, at the cost of decreased tracking precision. The periodic spike in the Personal mode precision for the star scenario in Fig. 3.10 is a direct consequence of the measurement model in (3.47) reflecting latent characteristics of the object shape, which invokes a relatively strong viewpoint-dependence for the state observability compared to the other tracking modes. This characteristic is not as apparent in Fig. 3.11 for the intersection scenario, due to the relatively simple object shape, and slow, acyclic, viewpoint changes compared to the star scenario; however, it is briefly apparent as the vehicles pass each other in close proximity at $t \approx 13$ s, when the

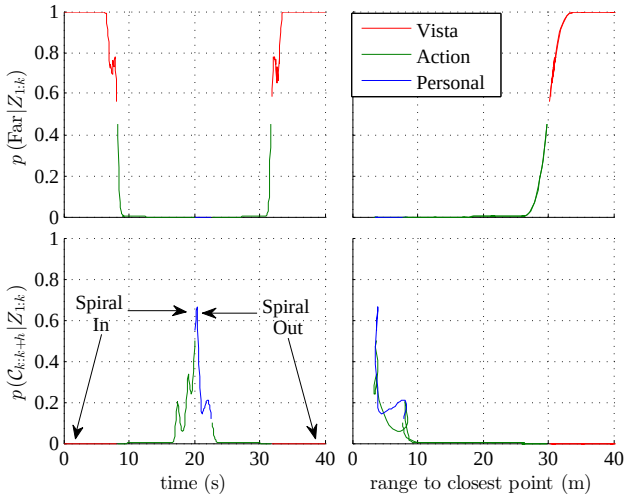

Figure 3.12: Probabilistic object relevancy metric trajectories for the star scenario. Annotations referring to events in the ground truth scenario are provided for perspective. **Top Row:** Probability that the object is 'Far' away, governing the transitions between Vista and Action modes. **Bottom Row:** Anticipated collision probability for a $h = 1$ s time horizon, governing the transitions between Action and Personal modes. **Left Column:** Variables plotted against time. **Right Column:** Variables plotted against distance to the closest point on the object.

viewpoint is changing most rapidly.

Lastly, Figs. 3.12 and 3.13 demonstrate the trajectories of the probabilistic object relevancy metrics for both scenarios. The flat region in the bottom left of Fig. 3.13 in the approximate range, 5.5 s$< t <$ 8.5 s, corresponds to the 3 second pause of both vehicles before proceeding through the intersection. Notice that, while range to the closest point on the object inherently factors into the collision probability, it is not an accurate predictor in itself. This is most apparent in the bottom right of Fig. 3.13, in that the collision probability is strictly higher as the vehicles approach each other at the center of the intersection (portion of the curve labeled 'On coming' in Fig. 3.13) than it is after they depart the inter-

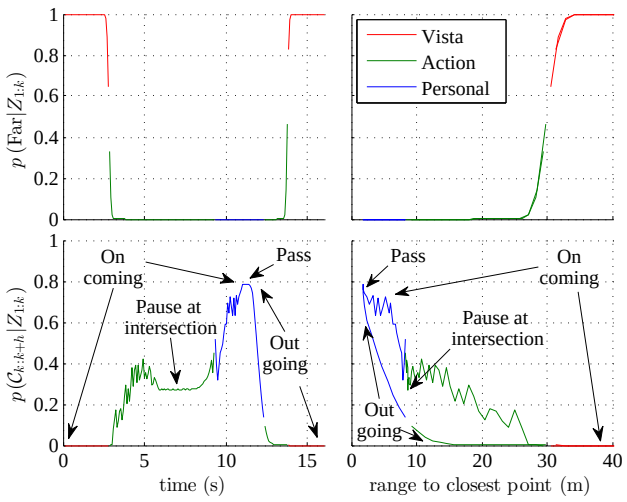

Figure 3.13: Probabilistic object relevancy metric trajectories for the intersection scenario. Annotations referring to events in the ground truth scenario are provided for perspective. **Top Row:** Probability that the object is 'Far' away, governing the transitions between Vista and Action modes. **Bottom Row:** Anticipated collision probability for a $h = 1$ s time horizon, governing the transitions between Action and Personal modes. **Left Column:** Variables plotted against time. **Right Column:** Variables plotted against distance to the closest point on the object.

section in opposing directions (portion of the curve labeled 'Out going' in Fig. 3.13). This is a direct result of the anticipatory nature of the collision probability. Specifically, as the vehicles approach, the algorithm anticipates that the distance between them continues to narrow, increasing the likelihood of an impending collision; conversely, as the vehicles depart, the algorithm anticipates that the distance between the objects continues to grow, decreasing the likelihood of an impending collision. This characteristic is not as apparent in Fig. 3.12 due to the spiral object trajectory. Specifically, given that the object approaches the ego vehicle without ever driving directly at it, the anticipation routine predicts that this behavior continues, and the probability of impending collision is relatively small until the object is within approximately 3 m of the ego vehicle. Given

this attribute, and the exact symmetry of the spiral trajectory about $t = 20$ s, the minor asymmetries in the collision probability in Fig. 3.12 (bottom) can be predominately attributed to the increase in the precision of the state belief in the latter half of the scenario (labeled 'Sprial Out' in Fig. 3.12) - a direct consequence of the hybrid mode transition to Personal space.

3.7 Conclusion

Inspired by human perception, this paper introduces a novel method to dynamically allocate computational resources to achieve variable precision tracking of arbitrary extended objects, without *a priori* knowledge of the environment or object extent characteristics (e.g. shape and size). Many sensible extended object tracking (EOT) methods exist, with the main distinction being the model chosen to represent the object extent. In general, simple extent models result in computationally efficient EOT, but engender low precision tracking by way of imprecise sensor models (i.e. large measurement source uncertainty); conversely, detailed and complex extent models tend to be computationally expensive, but engender high precision tracking by enabling complementary detailed and precise sensor models.

With the assertion that objects in a given scene are often of variable importance to the ego robot task, a priority-based hybrid tracking framework is proposed, enabling the robot to track objects that are important to the ego robot task with relatively expensive, high precision methods, and objects of peripheral importance with relatively efficient, low precision methods. The proposed hybrid approach is a direct analog to the human perception concepts of *atten-*

tion and *focus*. Probabilistic object relevancy metrics are derived to inform the hybrid mode transitions to achieve the intended priority-based tracker for arbitrary extended objects.

Autonomous driving simulation results are presented for the proposed priority based tracking framework, and evaluated over two different scenarios. Compared to a baseline high precision EOT algorithm, the results demonstrate that the priority-based framework enables a significant computational savings by relaxing its precision requirements for objects deemed to be of peripheral importance, while maintaining high precision tracks for objects regarded as essential to the ego robot task. The priority distinction is informed by the proposed set of probabilistic relevancy metrics, which are shown to accurately reflect their respective definitions of object *relevance* to the ego robot task, which is defined as collision free operation.

CHAPTER 4

PRECISION TRACKING VIA JOINT DETAILED SHAPE ESTIMATION OF ARBITRARY EXTENDED OBJECTS FROM LIDAR

4.1 Introduction

Extended objects, objects of non-negligible size relative to the sensor resolution, and *group objects*, clusters of closely-spaced objects relative to the sensor resolution such that they cannot be uniquely resolved in each sensor query, both violate the traditional tracking assumption that each object can produce, at most, a single measurement per sensor query [6]. Given their common distinction from classical object tracking, extended and group objects are often addressed together in a single comprehensive framework [7,12,18,30,31].

Comprehensive approaches to extended and group object tracking typically prescribe a parameterized structure of the shape/extent, e.g. circular disc [8], or ellipse [9,12,18,21,30,31], and estimate the shape parameters along with object kinematics in a recursive filter. Two popular methods include the use of random matrices [18,21,30,31] to represent the sample mean and covariance of the object extent, and the use of elliptic Random Hypersurface Models (RHM) [9,12] to represent the tightest enclosing ellipse of the object. A detailed comparison of Random Matrices to elliptic RHMs is provided in [7]. These methods tend to be efficient and rigorous in providing reasonable tracking performance for a variety of practical applications, but are insufficient for applications that require detailed shape estimates, or precise estimates of extended object kinematics. Among the applications demanding precise estimates of object kinematics are those requiring an autonomous agent to maneuver in close proximity to other

uncooperative objects/agents of unknown and arbitrary size and shape; for example, autonomous driving in traffic.

Alternatively, in [19], object measurements are modeled as random samples from a prescribed spatial distribution model (SDM) embodying the object shape and size. SDMs can represent more detailed sensor models than random matrices or elliptic RHMs, which contribute to increased tracking precision. However, as presented in [19], the SDM must be fully specified *a priori*, deeming these methods inappropriate when this information is unavailable.

For objects that are opaque to the sensor, self-occlusion (and occlusion in general) continuously morphs the true measurement distribution with changing viewpoint in ways that are unpredictable, as most algorithms lack a concept of object *surface* in the extent model. Therefore, many of the methods discussed thus far are better suited, or perhaps designed, for estimating the extent of the *visible* portion of the object in each sensor frame, rather than the extent of the entire object in space, over time.

Further, tracking approaches that comprehensively address both extended and group objects inherently suffer from degraded precision in tracking objects that are large relative to the sensor resolution. This is because, for group objects, sensor models are limited to predicting that measurements will arise from *anywhere* within the object extent; therefore, sensor model uncertainty scales with object size, and tracking precision degrades accordingly.

Two characteristics distinguishing extended objects from group objects, are that measurements from extended objects tend to originate solely from the boundary of the object (i.e. the object *surface*), and exhibit a rigid geometric

structure (i.e. the object *shape*). Conversely, measurements from group objects tend to be amorphous, and originate from the interior of the object extent. Leveraging these characteristics enables higher precision tracking of extended objects by way of refined sensor models [5, 10, 11, 24, 25, 27, 34, 37, 38, 40, 48–51].

Initial attempts to leverage the rigid geometric structure of measurements from extended objects generally involve prescribing a parameterized object shape structure, e.g. a rectangle [10, 34, 40], and estimating the parameters of the shape along with the object kinematics in a recursive filter. These methods differ from the comprehensive extended and group object tracking methods discussed above, in that they typically prescribe a more detailed shape structure, and assume measurements are generated solely from the object boundary. These methods tend to be efficient and rigorous in providing quality tracking performance, but only for the narrow set of objects accurately modeled by the prescribed structure. Therefore, *a priori* knowledge of the shape structure is required, deeming these methods inappropriate when this information is unavailable.

In [11], the RHM work of [12] and [9] evolves to include a more general model of object extent, the star-convex RHM, allowing for explicit estimation of the star-convex hull of the object shape. Specifically, the star-convex hull of the object is modeled as a Fourier series mapping angle to distance of the object extent from the star-convex center. The $2 \cdot N_f + 1$ parameters of the N_f degree Fourier series are estimated along with object kinematics in a recursive filter. In principle, this model can be proficient at tracking both extended and group objects, however, the performance and efficiency rely on the appropriate choice of N_f, which may not be apparent *a priori*. Further, reasonable initialization of

the Fourier parameter estimates at track instantiation appears critical for filter convergence, yet an appropriate choice is not clear without *a priori* knowledge of the approximate object shape and size.

A simple and intuitive method for tracking arbitrarily shaped extended objects involves the use of occupancy grids anchored to an object-centric coordinate frame, dubbed Object Local Grid Maps (OLG) [5]. OLG shape models can be rigorous and flexible, however, the accuracy and efficiency depend on an appropriate choice of grid extent and resolution, which requires some *a priori* knowledge of the size and shape complexity of objects to be tracked.

Given that lidar scan returns from extended objects tend to originate solely from the boundary of the object (i.e. the visible object *surfaces*) and exhibit a rigid geometric structure (i.e. the object *shape*), it is natural to define the object shape model as a set of points sampled from the object surface over time by a laser scanner. In this vein, non-parametric 3-dimensional (3D) point cloud models have been used to represent object shape by recursively aligning tracked object models with assigned segmented data, then appending the segmented data to the aligned shape model [24, 25, 27]. Point cloud models can be extremely flexible and provide rich 3D renderings of arbitrary object shapes, however, existing methods are fragile to uncertainty in sensing and the environment. Specifically, since only the first moment of the shape belief is maintained, the algorithms lack a concept of uncertainty in their estimate, and thus are not robust to common errors in the hard decisions on which they rely (e.g. data segmentation, data association, data alignment, etc.); in [24, 25] under-segmented tracks are omitted from the experimental evaluation. Further, measurement-model overlap is essential for model alignment and data association, and may be negligible for

under-developed models in highly dynamic scenes, or scenes with prevalent occlusion.

This work couples the flexibility and intuition of non-parametric point cloud models with the robustness and rigor of filtered parametric shape models to estimate the shape of arbitrary extended objects jointly with their kinematics. Preliminary work presented in [50] defines the shape model as a set of *salient* (viz. spatially distinct) points, dubbed *shape points*, sampled from the object surface over time, and infers the joint probability distribution over the shape model and kinematic states recursively from lidar data. Each lidar scan point is predicted as a probability-weighted combination of the visible shape points, yielding a precise sensor model founded on kernel-based regression.

Despite having many favorable qualities, defining saliency in terms of spatial distinction engenders a model complexity (number of shape points) that scales with object size, rather than the true complexity of the underlying object shape. As a remedy, this work builds upon the work presented in [50] by introducing a concept of *object surface* in the shape model, and redefining saliency in terms of the ability of a point to be predicted by its neighbors; the result is a polygonal representation of the hull of the object. The proposed approach for jointly estimating the object shape with its kinematics is dubbed *Polygonal Hull Shape Estimation and Tracking*, or PH-SET for short. This work develops and formalizes key algorithm components, discusses practical considerations, and provides both simulation and experimental analysis of the proposed method for complicated shaped objects maneuvering in natural, unstructured environments.

Section 4.2 introduces the proposed shape model, Section 4.3 describes the

extraction of an observation of the 2D object hull from a 3D lidar scan, Section 4.4 introduces the general object tracking problem, Sections 4.5, 4.6, and 4.7 derive the proposed PH-SET algorithm from the general object tracking problem, Sections 4.8, 4.9, and 4.10 present simulation and experimental results evaluating the PH-SET algorithm for an autonomous driving application, and finally Section 4.11 provides some concluding remarks.

4.2 Modeling the Shape of Arbitrary Extended Objects

The key to developing a shape model with complexity that scales only with the complexity of the true underlying object shape is the inclusion of a notion of *object surface*. With this in mind, the shape model at time k, χ_k, is defined as a polytope with vertices, V_k, representing the positions of salient lidar points returned from the object surface over time, dubbed *shape points*, and edges, E_k, encoding the number, N_v, and connectivity of the shape points in forming object surfaces:

$$\chi_k = (V_k, E_k) \tag{4.1}$$

In this model, *salient* shape points are defined as lidar scan points returned from the object surface that are not co-planar with their immediate neighbors. This work is indifferent to specific algorithms used to identify these points; any one of the many existing algorithms that extract features from sensor data could be used without loss of generality.

While all physical objects reside in 3-dimensional (3D) space, and *shape* is defined accordingly, this work is focused on the 2D tracking problem common with autonomous ground robotics; therefore a 2D abstraction of the full 3D ob-

ject shape is used as the shape model, in this case a 2D polytope (i.e. a polygon). For applications involving a flat ground plane and a single parallel lidar scan plane (e.g. common of indoor applications), the proposed 2D shape model represents the shape of the 2D slice through the object at the height of the scan plane. For more general applications, the proposed 2D shape model represents the hull of all such 2D slices over the height of the object, which is generally non-convex. Conceptually, the proposed shape model is a polygonal representation of the shadow cast by the object from a light source at infinite distance directly overhead.

4.3 Shape Observation

In the general, non-planar application of the proposed work, a 3D lidar sensor scans the visible surface of the object. Given that the shape model represents the 2D hull of the object projected onto the plane of motion, i.e. the horizontal *x*-*y* plane, an observation of the 2D hull must be extracted from each 3D scan.

Conceptually, the observation of the 2D hull, Z_k^{hull}, is computed by projecting all scan points, Z_k^{3D}, onto the *x*-*y* plane, and deleting those that lie on the interior of the hull, $\mathbb{D}_k$. Specifically,

$$Z_k^{\text{hull}} = \left(\boldsymbol{I}_{N_z} \otimes [\boldsymbol{I}_2, \boldsymbol{0}_{2\times 1}]\right)\left(Z_k^{\text{3D}} \backslash \hat{\mathbb{D}}_k\right) \tag{4.2}$$

where $\otimes$ denotes the Kronecker product, and pre-multiplying by $\left(\boldsymbol{I}_{N_z} \otimes [\boldsymbol{I}_2, \boldsymbol{0}_{2\times 1}]\right)$ projects the 3D points, $Z_k^{\text{3D}} \backslash \hat{\mathbb{D}}_k$, onto the *x*-*y* plane by extracting only the *x* and *y* coordinates of each of the N_z points contributing to the shape observation. $\hat{\mathbb{D}}_k$ denotes the estimate of the set of points that do not contribute to the 2D hull, $\mathbb{D}_k$, computed as the set of scan points projected onto the *x*-*y* plane with nearby

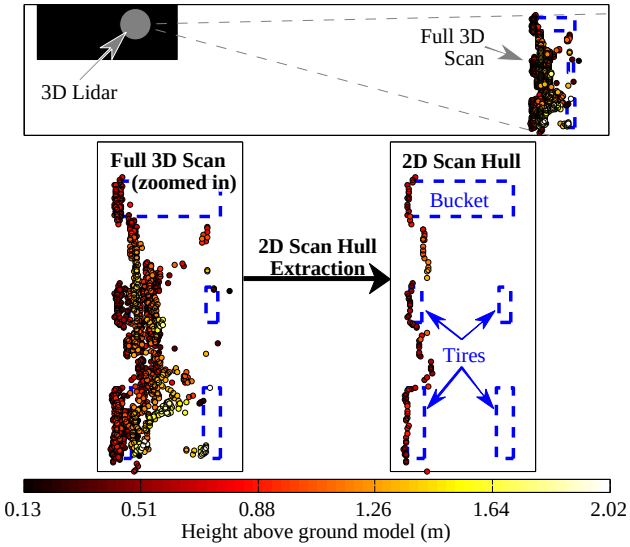

Figure 4.1: Example 2D hull extraction of a 3D Velodyne HDL-64 scan of a 1995 Kubota L2850 tractor using a bearing window of $\Delta\beta = 0.5°$. Scan points are pseudo-colored by height above the local ground model to demonstrate that scan points at varying heights contribute to the 2D hull observation. Tractor features (bucket and tires) were placed by hand to orient the viewer.

neighbors on the left (counterclockwise) and right (clockwise) at a shorter range, from the perspective of the sensor:

$$\hat{\mathbb{D}} = \left\{ z_\ell | z_\ell \in Z_k^{3D}, \mathbb{C}_{z_\ell}^{\text{ccw}} \neq \emptyset, \mathbb{C}_{z_\ell}^{\text{cw}} \neq \emptyset \right\} \tag{4.3}$$

where $\mathbb{C}_{z_\ell}^{\text{ccw}} \subset Z_k^{3D}$ and $\mathbb{C}_{z_\ell}^{\text{cw}} \subset Z_k^{3D}$ are the disjoint subsets of scan points that fall within counterclockwise and clockwise bearing windows relative to z_ℓ, respectively, at a shorter range:

$$\begin{aligned} \mathbb{C}_{z_\ell}^{\text{ccw}} &= \left\{ z_i \middle| z_i \in Z_k^{3D}, \rho_{z_i} < \rho_{z_\ell}, \beta_{z_\ell} < \beta_{z_i} < (\beta_{z_\ell} + \Delta\beta) \right\} \\ \mathbb{C}_{z_\ell}^{\text{cw}} &= \left\{ z_i \middle| z_i \in Z_k^{3D}, \rho_{z_i} < \rho_{z_\ell}, \beta_{z_\ell} > \beta_{z_i} > (\beta_{z_\ell} - \Delta\beta) \right\} \end{aligned} \tag{4.4}$$

where β_{z_ℓ} and ρ_{z_ℓ} denote the bearing and range in the x-y plane to the ℓth scan point, and $\Delta\beta$ is the width of the bearing window chosen to be slightly larger than the horizontal bearing resolution of the sensor.

Fig. 4.1 shows an example 2D hull observation computed from a 3D Velodyne HDL-64 scan of a 1995 Kubota L2850 tractor using a bearing window of $\Delta\beta = 0.5°$. Notice that the proposed method effectively removes returns from surfaces that do not contribute to the 2D hull of the object shape.

Note that, hereafter, the 'hull' superscript is dropped for brevity, and in general, non-planar, applications, Z_k denotes the 2D hull measurement in (4.2), $Z_k := Z_k^{\text{hull}}$.

4.4 Object Tracking Problem Formulation

The goal of general multi-object tracking is to estimate the full latent object state history, $X_{1:K}$, of an unknown number of maneuvering objects, N_O, from a history of noisy observations, $Z_{1:K}$, without knowledge of object controls or intent. Probabilistic inference provides a rigorous means for accounting for the many sources of uncertainty in the problem, deeming it a valuable tool for multi-object tracking. Specifically, rather than estimating the latent variables directly, inference methods estimate the joint posterior probability distribution over the latent variables conditioned on the observations,

$$p\left(X_{1:K}^{1:N_O}\middle|Z_{1:K}\right) = \prod_{n=1}^{N_O} p\left(X_{1:K}^{n}|Z_{1:K}, X_{1:K}^{1:n\text{-}1}\right) \tag{4.5}$$

from which optimal estimates of state trajectories, $\hat{X}_{1:K}^{1:N_O}$, can be computed via existing techniques, such as Minimum Mean Square Error (MMSE) or Maximum-a-Posteriori (MAP).

In most practical applications, it is accurate to model the objects as being mutually independent, $p\left(X_{1:K}^{n}|X_{1:K}^{m}\right) = p\left(X_{1:K}^{n}\right) \forall n \neq m$, which is convenient in

that it simplifies the full joint multi-object tracking problem into the product of single-object marginals that can be studied independently in parallel:

$$p\left(X_{1:K}^{1:N_O}\middle|Z_{1:K}\right) = \prod_{n=1}^{N_O} p\left(X_{1:K}^{n}|Z_{1:K}\right) \tag{4.6}$$

Therefore, theoretical development in object tracking is commonly focused on the single-object tracking problem, i.e. estimating:

$$p\left(X_{1:K}|Z_{1:K}\right) \tag{4.7}$$

Lastly, online tracking applications require state estimates in real-time as data is received, and are often principally concerned with the current object state, rather than the full time-history. Therefore, the inference problem is further simplified to estimating:

$$p\left(X_k|Z_{1:k}\right) \forall k \in \{1, \ldots, K\} \tag{4.8}$$

That is, the marginal distribution over the states at any time, k, is conditioned only on the observation history through k. Note that estimating (4.8) is commonly referred to as *filtering*, which can be supplemented with *smoothing* to estimate the full posterior distribution over the state history in (4.7). This work focuses solely on the filtering problem in (4.8), in accordance with the strict efficiency requirements for online tracking applications; however, the associated smoothing algorithms can be easily derived from the given models and equations.

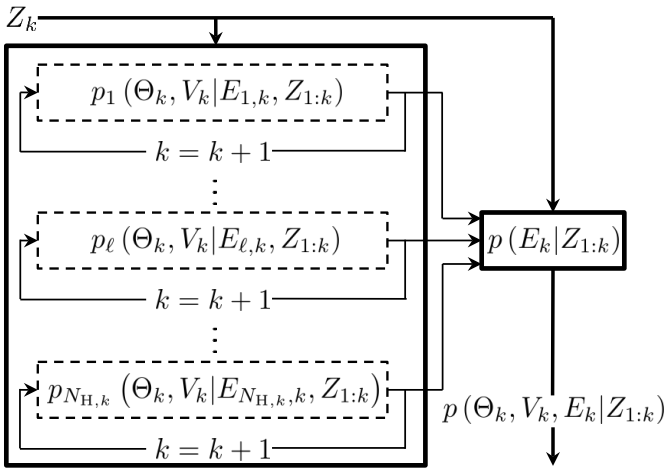

Figure 4.2: Block diagram of the multiple hypothesis framework for estimating the joint posterior in (4.10).

4.5 Estimating Shape Jointly with Kinematics

With the interest of rigorously estimating object shape, χ_k, jointly with object kinematics, Θ_k, the object state is defined as:

$$\begin{aligned} X_k &= (\Theta_k, \chi_k) \\ &= (\Theta_k, V_k, E_k) \end{aligned} \tag{4.9}$$

and the filtering problem in (4.8) is factored to partition the hybrid state into its continuous and discrete parts, as follows:

$$\begin{aligned} p\left(X_k | Z_{1:k}\right) &= p\left(\Theta_k, V_k, E_k | Z_{1:k}\right) \\ &= p\left(\Theta_k, V_k | E_k, Z_{1:k}\right) \cdot p\left(E_k | Z_{1:k}\right) \end{aligned} \tag{4.10}$$

The posterior distribution in (4.10) is estimated using the multiple-hypothesis approach depicted in Fig. 4.2. Specifically, a bank of $N_{\mathrm{H},k}$ parallel filters solve the filtering problem for the continuous partition of the object state, $X_k^{\mathrm{c}} = (\Theta_k, V_k)$, with each filter further conditioned on its own unique hypothesis

of the discrete state partition, E_k:

$$p_\ell\left(\Theta_k, V_k | E_{\ell,k}, Z_{1:k}\right) \forall \ell \in \{1, \ldots, N_{\mathrm{H},k}\} \tag{4.11}$$

Further, the $N_{\mathrm{H},k}$ constituent, mutually exclusive, posterior distributions over the continuous states, (4.11), are evaluated against the data to infer the posterior categorical distribution over hypotheses of the discrete states, E_k:

$$p\left(E_k | Z_{1:k}\right) \tag{4.12}$$

thereby completing the full inference problem in (4.10). The proposed approach is dubbed *Polygonal Hull Shape Estimation and Tracking*, or PH-SET for short. The details of both the continuous and discrete partitions of this approach are provided in the coming sections.

4.6 Continuous Problem Partition

The continuous problem partition involves estimating the distribution over the continuous states, $X_k^{\mathrm{c}} = (\Theta_k, V_k)$, for each of the $N_{\mathrm{H},k}$ parallel filters depicted in Fig. 4.2, conditioned on the history of measurements, $Z_{1:k}$, and their own unique hypothesis of the discrete states, $E_k = E_{\ell,k}$. Since the algorithm is identical for all constituent filters, it is presented here for only one filter, and the filter/hypothesis ID, ℓ depicted in Fig 4.2, is dropped from the notation for brevity.

With a goal of efficient and closed-form inference, the continuous problem is represented using the Hidden Markov model (HMM) depicted graphically in

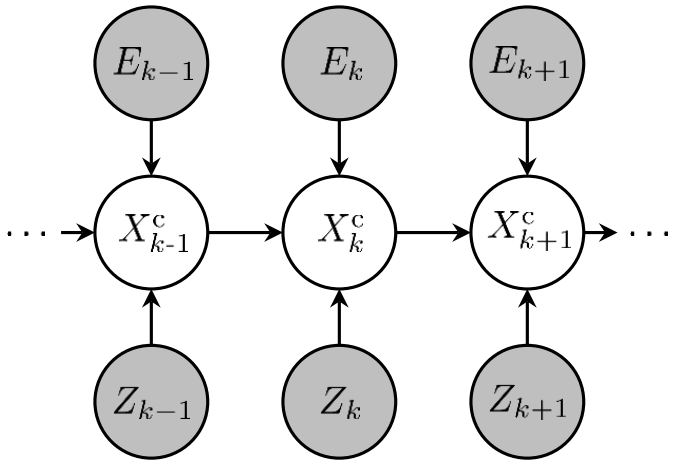

Figure 4.3: Graphical representation of the Hidden Markov model (HMM) used to represent the continuous partition of the PH-SET inference approach within each of the $N_{\mathrm{H},k}$ parallel filters.

Fig. 4.3, with each conditional distribution (node) modeled as Gaussian:

$$\begin{aligned} p\left(\Theta_k, V_k | E_k, Z_{1:k}\right) &= p\left(X^{\mathrm{c}}_k | E_k, Z_{1:k}\right) \\ &= \mathcal{N}\left(\bar{X}^{\mathrm{c}}_{k|Z_k}, \boldsymbol{P}_{X^{\mathrm{c}}_k|Z_k}\right) \end{aligned} \quad \forall\ k \tag{4.13}$$

where the belief distribution at time k is fully specified by the mean, $\bar{X}^{\mathrm{c}}_{k|Z_k}$, and covariance, $\boldsymbol{P}_{X^{\mathrm{c}}_k|Z_k}$, structured as:

$$\begin{aligned} \bar{X}^{\mathrm{c}}_{k|Z_k} &= \begin{bmatrix} \bar{\Theta}_{k|Z_k} \\ \bar{V}_{k|Z_k} \end{bmatrix} \\ \boldsymbol{P}_{X^{\mathrm{c}}_k|Z_k} &= \begin{bmatrix} \boldsymbol{P}_{\Theta_k|Z_k} & \boldsymbol{P}_{\Theta_k V_k|Z_k} \\ \boldsymbol{P}^T_{\Theta_k V_k|Z_k} & \boldsymbol{P}_{V_k|Z_k} \end{bmatrix} \end{aligned} \tag{4.14}$$

Kalman filtering techniques can be leveraged to perform the inference. Further, the MMSE and MAP estimates of the state are both identically equal to the mean of the distribution:

$$\hat{X}^{\mathrm{c,MAP}}_k = \hat{X}^{\mathrm{c,MMSE}}_k = \bar{X}^{\mathrm{c}}_{k|Z_k} \tag{4.15}$$

deeming the extraction of state estimates from the inferred posterior distributions trivial. The dynamics prediction and measurement update steps of the

filtering process are discussed in the following sections.

4.6.1 Dynamics Prediction

This work is largely indifferent to sensible choices for the kinematic states, Θ, and their dynamics model; therefore, the model chosen to validate the proposed approach is presented here, and is appropriate for applications lacking known motion constraints for the tracked objects. However, more detailed dynamics models can be substituted without loss of generality.

Employing a rigid-body assumption for the object shape greatly improves the efficiency of the PH-SET algorithm in that it enables the dynamics of the entire set of shape points, V_k, to be described collectively with a single discrete coordinate transformation. With this in mind, the belief of the full continuous partition of the object state, X_k^{c}, is propagated forward in time via a two-stage, hybrid prediction step; first, the distribution over the kinematic states, Θ_k, is propagated forward in time via a prescribed continuous differential equation model, then, the distribution over the shape points, V_k, is propagated via a discrete coordinate transformation between the kinematic states at both time steps. Finally, the full joint prior distribution over the continuous states, X_k^{c}, is built from the output parameters of the constituent prediction stages:

$$p\left(X_k^{\mathrm{c}} | E_k, Z_{1:k\text{-}1}\right) \sim \mathcal{N}\left(\bar{X}_{k|Z_{k\text{-}1}}^{\mathrm{c}}, \boldsymbol{P}_{X_k^{\mathrm{c}}|Z_{k\text{-}1}}\right) \tag{4.16}$$

where the mean, $\bar{X}_{k|Z_{k\text{-}1}}^{\mathrm{c}}$, and covariance, $\boldsymbol{P}_{X_k^{\mathrm{c}}|Z_{k\text{-}1}}$, follow the same block structure provided for the posterior in (4.14). Specifically, the outputs of the kinematics prediction stage are: $\bar{\Theta}_{k|Z_{k\text{-}1}}$ and $\boldsymbol{P}_{\Theta_k|Z_{k\text{-}1}}$, and the outputs of the shape point prediction stage are: $\bar{V}_{k|Z_{k\text{-}1}}$, $\boldsymbol{P}_{V_k|Z_{k\text{-}1}}$, and $\boldsymbol{P}_{\Theta_k V_k|Z_{k\text{-}1}}$.

The following subsections provide the state propagation equations (i.e. the dynamics model) for both the continuous and discrete stages of the hybrid prediction step. The state distributions can be pushed through these equations via any suitable existing method, such as the prediction step of the Extended Kalman Filter (EKF) [6] or Sigma Point Filter (SPF) [28].

Kinematic States

The kinematic states are chosen as the position of an arbitrary point fixed to the object, (x, y), the change in orientation of the object about that point, ϕ, relative to its arbitrary initial orientation (the shape model does not define the *front*, thus absolute orientation is undefined), and their time-derivatives, all described from the perspective of the ego-vehicle:

$$\Theta = \left[x, y, \phi, \dot{x}, \dot{y}, \dot{\phi}\right]^T \tag{4.17}$$

Note that the orientation of the object, ϕ, is decoupled from the object heading, $\tan^{-1}(\dot{y}/\dot{x})$, to accommodate arbitrary objects with a variety of latent motion constraints.

The kinematic states evolve according to the following continuous time differential equation:

$$\dot{\Theta}_k = \boldsymbol{F}_k \Theta_k + \boldsymbol{G}_k \left(W_{\mathrm{g},k} - \begin{bmatrix} \ddot{x}_{\mathrm{g},k}^{\mathrm{ego}} \\ \ddot{y}_{\mathrm{g},k}^{\mathrm{ego}} \\ \ddot{\phi}_{\mathrm{g},k}^{\mathrm{ego}} \end{bmatrix} \right) \tag{4.18}$$

where subscript 'g' indicates that the variable is described in a global coordinate frame, and superscript 'ego' denotes that the variable pertains to the ego vehicle dynamics, and is provided by an independent localization routine on-board

the ego vehicle. The object acceleration in the global reference frame, $W_{\mathrm{g},k}$, is assumed to be driven by a zero-mean Gaussian white noise process:

$$W_{\mathrm{g},k} = \begin{bmatrix} \ddot{x}_{\mathrm{g},k} \\ \ddot{y}_{\mathrm{g},k} \\ \ddot{\phi}_{\mathrm{g},k} \end{bmatrix} \sim \mathcal{N}(\boldsymbol{0}_{3\times 1}, \boldsymbol{Q}) \quad \forall\, k \tag{4.19}$$

The matrices $\boldsymbol{F}_k$ and $\boldsymbol{G}_k$ are defined as a function of the ego vehicle rotation matrix, $\boldsymbol{M}_{\mathrm{g},k}^{\mathrm{ego}}$, and its time-derivatives:

$$\begin{aligned} \boldsymbol{F}_k &= \begin{bmatrix} & \boldsymbol{0}_{3\times 3} & \boldsymbol{I}_{3\times 3} & \\ \left(\boldsymbol{M}_{\mathrm{g},k}^{\mathrm{ego}}\right)^T \dot{\boldsymbol{M}}_{\mathrm{g},k}^{\mathrm{ego}} & \boldsymbol{0}_{2\times 1} & -2\left(\boldsymbol{M}_{\mathrm{g},k}^{\mathrm{ego}}\right)^T \dot{\boldsymbol{M}}_{\mathrm{g},k}^{\mathrm{ego}} & \boldsymbol{0}_{2\times 1} \\ & \boldsymbol{0}_{1\times 6} & & \end{bmatrix} \\ \boldsymbol{G}_k &= \begin{bmatrix} \boldsymbol{0}_{3\times 3} & \\ \left(\boldsymbol{M}_{\mathrm{g},k}^{\mathrm{ego}}\right)^T & \boldsymbol{0}_{2\times 1} \\ \boldsymbol{0}_{1\times 2} & 1 \end{bmatrix} \end{aligned} \tag{4.20}$$

where:

$$\begin{aligned} \boldsymbol{M}_{\mathrm{g},k}^{\mathrm{ego}} &= \begin{bmatrix} \cos\phi_{\mathrm{g},k}^{\mathrm{ego}} & -\sin\phi_{\mathrm{g},k}^{\mathrm{ego}} \\ \sin\phi_{\mathrm{g},k}^{\mathrm{ego}} & \cos\phi_{\mathrm{g},k}^{\mathrm{ego}} \end{bmatrix} \\ \dot{\boldsymbol{M}}_{\mathrm{g},k}^{\mathrm{ego}} &= \frac{\partial \boldsymbol{M}_{\mathrm{g},k}^{\mathrm{ego}}}{\partial \phi_{\mathrm{g},k}^{\mathrm{ego}}} \dot{\phi}_{\mathrm{g},k}^{\mathrm{ego}} \\ \ddot{\boldsymbol{M}}_{\mathrm{g},k}^{\mathrm{ego}} &= \frac{\partial \boldsymbol{M}_{\mathrm{g},k}^{\mathrm{ego}}}{\partial \phi_{\mathrm{g},k}^{\mathrm{ego}}} \ddot{\phi}_{\mathrm{g},k}^{\mathrm{ego}} - \boldsymbol{M}_{\mathrm{g},k}^{\mathrm{ego}} \left(\dot{\phi}_{\mathrm{g},k}^{\mathrm{ego}}\right)^2 \end{aligned} \tag{4.21}$$

Shape Points

After the kinematic states have been propagated using the model in (4.18), the shape points are propagated via the following discrete coordinate transform between the ego coordinate frame at time k and $k + 1$:

$$V_k = \boldsymbol{M}_{k\text{-}1,k}^{\mathrm{Block}} \left(V_{k\text{-}1} - \boldsymbol{T}_{k\text{-}1}^{\mathrm{Block}}\right) + \boldsymbol{T}_k^{\mathrm{Block}} \tag{4.22}$$

where:

$$\begin{aligned} \boldsymbol{M}_{k\text{-}1,k}^{\text{Block}} &= \boldsymbol{I}_{N_V} \otimes \begin{bmatrix} \cos(\phi_k - \phi_{k\text{-}1}) & -\sin(\phi_k - \phi_{k\text{-}1}) \\ \sin(\phi_k - \phi_{k\text{-}1}) & \cos(\phi_k - \phi_{k\text{-}1}) \end{bmatrix} \\ \boldsymbol{T}_k^{\text{Block}} &= \mathbf{1}_{N_V \times 1} \otimes \begin{bmatrix} x_k \\ y_k \end{bmatrix} \end{aligned} \tag{4.23}$$

where N_V denotes the number of shape points, and superscript 'Block' denotes the stacking of the point-wise transformation terms to accommodate the full set of shape points; i.e. the effect of the Kronecker product, $\otimes$. Note that the kinematic states must be predicted first via (4.18) to obtain the coordinate transformation terms in (4.23). The coordinate transform in (4.22) projects the shape points from the ego coordinate frame at time k to the ego coordinate frame at time $k + 1$, by way of an intermediate coordinate frame fixed to the object described by the first three kinematic states (x, y, ϕ). Given the rigid body assumption, the shape points are static with respect to the intermediate coordinate frame, and the collective motion of the shape points over the time interval, as seen from the ego coordinate frame, is described by the coordinate transformation terms in (4.23).

4.6.2 Measurement Update

For the posterior distribution over the continuous partition of the object state, $X_k^{\text{c}} = (\Theta_k, V_k)$, to be Gaussian at each time step, as specified in (4.13), the joint pdf over X_k^{c} and the measurement, Z_k, must be modeled as being jointly Gaussian:

$$p(X_k^{\text{c}}, Z_k | E_k) = \mathcal{N}\left(\begin{bmatrix} \bar{X}_{k|Z_{k\text{-}1}}^{\text{c}} \\ \bar{Z}_k \end{bmatrix}, \begin{bmatrix} \boldsymbol{P}_{X_k^{\text{c}}|Z_{k\text{-}1}} & \boldsymbol{P}_{X_k^{\text{c}}|Z_{k\text{-}1},Z_k} \\ \boldsymbol{P}_{X_k^{\text{c}}|Z_{k\text{-}1},Z_k}^T & \boldsymbol{P}_{Z_k} \end{bmatrix} \right) \tag{4.24}$$

where $\bar{X}^{\mathrm{c}}_{k|Z_{k-1}}$, and $\boldsymbol{P}_{X^{\mathrm{c}}_k|Z_{k-1}}$ denote the mean and covariance of the *a priori* state belief at time k in (4.16), and $\bar{Z}_k$, $\boldsymbol{P}_{Z_k}$, and $\boldsymbol{P}_{X^{\mathrm{c}}_k|Z_{k-1},Z_k}$ denote the mean and covariance of the expected measurement and the measurement-state covariance, respectively, as derived from the measurement model. In this way, the parameters of the Gaussian posterior in (4.14) can be computed in closed-form:

$$\begin{aligned} \bar{X}^{\mathrm{c}}_{k|Z_k} &= \bar{X}^{\mathrm{c}}_{k|Z_{k-1}} + \boldsymbol{K}_k \left(Z_k - \bar{Z}_k \right) \\ \boldsymbol{P}_{X^{\mathrm{c}}_k|Z_k} &= \boldsymbol{P}_{X^{\mathrm{c}}_k|Z_{k-1}} - \boldsymbol{K}_k \boldsymbol{P}^T_{X^{\mathrm{c}}_k|Z_{k-1},Z_k} \\ \boldsymbol{K}_k &= \boldsymbol{P}_{X^{\mathrm{c}}_k|Z_{k-1},Z_k} \boldsymbol{P}^{-1}_{Z_k} \end{aligned} \tag{4.25}$$

For this work, the information in each lidar scan is fused via a two-step process; first, a rough alignment is performed by fusing a measurement of the object extremities (extreme bearings, and minimum range), dubbed the *metadata measurement*, followed by a detailed *surface measurement* to fuse the remaining information in the lidar scan.

The following subsections provide the models for each measurement type, as well as the corresponding measurement model dependent parameters of the joint distribution in (4.24) required to perform the update in (4.25); $\bar{Z}_k$, $\boldsymbol{P}_{Z_k}$, and $\boldsymbol{P}_{X^{\mathrm{c}}_k|Z_{k-1},Z_k}$.

Metadata Measurement

The metadata measurement, inspired by [38], is fused first to serve as a rough alignment of the *a priori* state belief with the lidar scan prior to fusing the detailed surface measurement. Specifically, the clockwise-most (cw) bearing, β^{cw}, counterclockwise-most (ccw) bearing, β^{ccw}, and minimum range, ρ, to the object

are extracted from the lidar scan and fused as a measurement:

$$Z_k = \begin{bmatrix} \beta_k^{\text{cw}} \\ \beta_k^{\text{ccw}} \\ \rho_k \end{bmatrix} + \nu_k^{\text{meta}} \tag{4.26}$$

where $\nu_k^{\text{meta}} \sim \mathcal{N}\left(\mathbf{0}_{3\times 1}, \boldsymbol{R}^{\text{meta}}\right)$ is the unbiased sensor noise, with covariance, $\boldsymbol{R}^{\text{meta}}$, derived from the range and bearing precision provided by the lidar sensor specification.

The required parameters of the joint distribution over the metadata measurement and the object state in (4.24) are computed from the *a priori* state belief and measurement model, (4.26), via the sigma point transform [28]. Note that the (two or three) individual lidar returns corresponding to the metadata in (4.26) are removed from the scan prior to fusing the surface measurement to avoid double counting information.

Surface Measurement

The surface measurement model makes use of the detailed shape estimate to fuse information from the remainder of the lidar scan. To compute the expected measurement distribution over the remaining lidar returns, a new vector of latent discrete variables, C_{E_k}, is introduced where $C_{E_k}(\ell) = j$ denotes the latent event that the ℓth scan point at time k is returned from edge j. Given C_{E_k}, the sensor model is written as:

$$Z_{k|C_{E_k}} = \left(\boldsymbol{H}_{k|C_{E_k}} \otimes \boldsymbol{I}_2\right) \cdot V_k + \nu_k^{\text{surf}} \tag{4.27}$$

where $\nu_k^{\text{surf}} \sim \mathcal{N}\left(\mathbf{0}_{N_{Z_k}\times 1}, \boldsymbol{R}_k^{\text{surf}}\right)$ is the unbiased sensor noise with covariance, $\boldsymbol{R}_k^{\text{surf}}$, computed by projecting the range and bearing precision provided by the lidar sensor specification into Cartesian space, linearizing about the observed scan points.

The measurement matrix, $\boldsymbol{H}_{k|C_{E_k}}$, is defined as:

$$\boldsymbol{H}_{k|C_{E_k}}(\ell, n) = \begin{cases} (1-\alpha_{\ell|C_{E_k}}) & \text{if } E\left(C_{E_k}(\ell), 1\right) = n \\ \alpha_{\ell|C_{E_k}} & \text{if } E\left(C_{E_k}(\ell), 2\right) = n \\ 0 & \text{otherwise} \end{cases} \tag{4.28}$$

where $\alpha_{\ell|C_{E_k}}$ is the linear interpolation gain along edge $C_{E_k}(\ell)$ for scan point ℓ, and $E(i, j)$ denotes the index of the $j^{\text{th}} \in \{1, 2\}$ shape point making up edge i. In this way, the measurement model in (4.27) represents each scan point as a linear interpolation along the shape edge from which it was returned.

The joint distribution over the scan and object state, (4.24), is then found by marginalizing over the discrete latent variable, C_{E_k}:

$$\begin{aligned} p\left(X_k^{\text{c}}, Z_k | E_k\right) &= \sum_{C_{E_k}} p\left(X_k^{\text{c}}, Z_k, C_{E_k} | E_k\right) \\ &= \sum_{C_{E_k}} p\left(X_k^{\text{c}}, Z_k | C_{E_k}, E_k\right) \cdot p\left(C_{E_k} | E_k\right) \end{aligned} \tag{4.29}$$

where $p\left(X_k^{\text{c}}, Z_k | C_{E_k}, E_k\right)$ is the joint distribution from (4.24), further conditioned on a candidate setting of the latent discrete variable, C_{E_k}. The conditional mean, covariance, and measurement-state correlation of $p\left(X_k^{\text{c}}, Z_k | C_{E_k}, E_k\right)$ are computed from the measurement model in (4.27), specifically:

$$
\begin{aligned}
\bar{Z}_{k|C_{E_k}} &= \mathbb{E}\left[Z_{k|C_{E_k}}\right] \\
&= \left(\boldsymbol{H}^{*}_{k|C_{E_k}} \otimes \boldsymbol{I}_2\right) \bar{V}_k \\
\boldsymbol{P}_{Z_k|C_{E_k}} &= \mathbb{E}\left[\left(Z_{k|C_{E_k}} - \bar{Z}_{k|C_{E_k}}\right)\left(Z_{k|C_{E_k}} - \bar{Z}_{k|C_{E_k}}\right)^T\right] \\
&= \left(\boldsymbol{H}^{*}_{k|C_{E_k}} \otimes \boldsymbol{I}_2\right) \boldsymbol{P}_{V_k} \left(\boldsymbol{H}^{*T}_{k|C_{E_k}} \otimes \boldsymbol{I}_2\right) + \boldsymbol{R}^{\text{surf}}_k \\
\boldsymbol{P}_{X^{\text{c}}_k, Z_k|C_{E_k}} &= \mathbb{E}\left[\left(X^{\text{c}}_k - \bar{X}^{\text{c}}_k\right)\left(Z_{k|C_{E_k}} - \bar{Z}_{k|C_{E_k}}\right)^T\right] \\
&= \boldsymbol{P}_{X^{\text{c}}_k}\left(\boldsymbol{H}^{*T}_{k|C_{E_k}} \otimes \boldsymbol{I}_2\right)
\end{aligned}
\tag{4.30}
$$

where $\mathbb{E}[\cdot]$ denotes the expectation, and the *optimal* measurement matrix, $\boldsymbol{H}^{*}_{k|C_{E_k}}$, is (4.28) evaluated at the maximum likelihood interpolation gains, $\alpha^{*}_{\ell|C_{E_k}} \forall\, \ell, C_{E_k}$, found to be:

$$
\alpha^{*}_{\ell|C_{E_k}} = \max\left(\min\left(\bar{\alpha}^{*}_{\ell|C_{E_k}}, 1\right), 0\right) \tag{4.31}
$$

where:

$$
\bar{\alpha}^{*}_{\ell|C_{E_k}} = \frac{\bar{\delta}^{T}_{E_k}\left(C_{E_k}(\ell)\right)\left[Z_k(\ell) - \bar{V}_k\left(E\left(C_{E_k}(\ell), 1\right)\right)\right]}{\bar{\delta}^{T}_{E_k}\left(C_{E_k}(\ell)\right)\bar{\delta}_{E_k}\left(C_{E_k}(\ell)\right)} \tag{4.32}
$$

and:

$$
\bar{\delta}_{E_k}\left(C_{E_k}(\ell)\right) = \bar{V}_k(E\left(C_{E_k}(\ell), 2\right)) - \bar{V}_k(E\left(C_{E_k}(\ell), 1\right)) \tag{4.33}
$$

The second term in (4.29), $p\left(C_{E_k}|E_k\right)$, is the discrete probability that the candidate C_{E_k} is correct, defined as:

$$
p\left(C_{E_k}|E_k\right) = \frac{\mathcal{N}\left(Z_k; \bar{Z}_{k|C_{E_k}}, \boldsymbol{P}_{Z_k|C_{E_k}}\right)}{\sum\limits_{C_{E_k}} \mathcal{N}\left(Z_k; \bar{Z}_{k|C_{E_k}}, \boldsymbol{P}_{Z_k|C_{E_k}}\right)} \tag{4.34}
$$

The joint distribution in (4.29) is a Gaussian mixture, which can significantly increase the complexity of the measurement fusion. To avoid this complexity, rather than resorting to Gaussian mixture filtering methods, (4.29) is approximated by a single Gaussian of the form in (4.24). This is a reasonable approximation due to the fact that the characteristic high precision of lidar sensors

makes the scan point to edge correspondences, C_{E_k}, obvious under most conditions, resulting in a low entropy $p(C_{E_k}|Z_k)$; therefore, the mixture in (4.29) is dominated by a small number of mixands. Further, due to the nature of (4.34), if uncertainty in C_{E_k} does exist, it will be among edges that predict a similar conditional measurement distribution, $p\left(X_k^c, Z_k|C_{E_k}, E_k\right)$, (typically adjacent edges); therefore, there will be significant overlap among the constituent mixands with non-negligible weight, and the mixture in (4.29) will be well approximated by a single mixand. The single Gaussian approximation to the joint distribution in (4.29) is achieved by matching first and second moments:

$$\begin{aligned} \bar{Z}_k &= \sum_{C_{E_k}} p(C_{E_k}|Z_k) \cdot \bar{Z}_{k|C_{E_k}} \\ \boldsymbol{P}_{Z_k} &= \sum_{C_{E_k}} p(C_{E_k}|Z_k) \cdot \left[\boldsymbol{P}_{Z_k|C_{E_k}} + \boldsymbol{P}_{\bar{Z}_k|C_{E_k}}\right] \\ \boldsymbol{P}_{X_k^c,Z_k} &= \sum_{C_{E_k}} p(C_{E_k}|Z_k) \cdot \boldsymbol{P}_{X_k^c,Z_k|C_{E_k}} \end{aligned} \tag{4.35}$$

where the covariance of the mixand means is defined as:

$$\boldsymbol{P}_{\bar{Z}_k|C_{E_k}} = \left(\bar{Z}_{k|C_{E_k}} - \bar{Z}_k\right)\left(\bar{Z}_{k|C_{E_k}} - \bar{Z}_k\right)^T \tag{4.36}$$

The paramters in (4.35) are used in (4.25) to compute the posterior distribution over the object state given the observed surface scan.

4.7 Discrete Problem Partition

The discrete problem partition requires forming, evaluating, and pruning hypotheses of the discrete state variable, E, which encodes the number and connectivity of the shape points in forming object surfaces. The full solution to this problem involves forming hypotheses for all possible combinations of connections at each time step; this approach has complexity combinatorial in the

number of vertices, and exponential in the number of time steps (duration of the track), deeming it infeasible for even mildly complex shapes over short time horizons. However, the rigid body assumption employed suggests that the true connectivity does not change with time, $E_k = E \; \forall k$, reducing the algorithm complexity to constant in the number of time steps. Further, certain characteristics of the lidar sensor preclude a vast majority of connectivity combinations, greatly reducing the computational complexity with the number of vertices.

4.7.1 Connectivity of salient scan points

The sensor generates points over the horizontal field-of-view (fov) via a continuous rotation, providing an inherent ordering of the points in the direction of rotation; this, coupled with the assumption that lidar sensors cannot penetrate object surfaces of interest (otherwise lidar would be a poor choice of sensing modality for the application), gives rise to a single valid hypothesis for the connectivity of the salient lidar scan points, $Z_k^* \subseteq Z_k$. Specifically, each point forms a surface with its immediate neighbor(s) in the inherently ordered sequence; one cw and one ccw from the perspective of the sensor. Note that the two end points of each scan have one open connection to account for occluded surfaces on either side of the object. Leveraging these sensor characteristics dramatically reduces the connectivity complexity from $n!$ possible combinations among the n salient scan points, to 1 feasible connectivity hypothesis among the salient scan points.

4.7.2 Hypothesis Formation

Given that there is only one valid connectivity, E, among the salient lidar scan points, the algorithm is initialized from the first lidar scan of the object with a single hypothesis. However, the shape model must adapt as viewpoint changes reveal previously occluded object surfaces not yet represented in the shape model. These events are identified by thresholding the mean normalized innovation squared, $\bar{\epsilon}_k$, of the surface measurements in (4.27):

$$\bar{\epsilon}_k = \frac{\epsilon_k}{2 \cdot N_{Z_k}} \tag{4.37}$$

where the normalized innovation squared, ϵ_k, is defined as:

$$\epsilon_k = \left(Z_k - \bar{Z}_k\right)^T \boldsymbol{P}_{Z_k}^{-1} \left(Z_k - \bar{Z}_k\right) \tag{4.38}$$

Specifically, if $\bar{\epsilon}_k > \tau_\epsilon$, where τ_ϵ is a user-defined threshold, then it is taken as evidence that the current belief of the shape model is insufficient to predict the current lidar scan, and the model growth procedure is instantiated in place of the surface measurement update.

The updated set of shape points, $V_{k|Z_k}$, is obtained by appending the salient scan points, $Z_k^* \subseteq Z_k$, that do not correspond to an existing shape point to the existing set of shape points, $V_{k|Z_{k\text{-}1}}$. Specifically,

$$V_{k|Z_k} = \left(V_{k|Z_{k\text{-}1}}, V_{k|C_{V_k}}\right) \tag{4.39}$$

where:

$$V_{k|C_{V_k}} = \{z_\ell^* | z_\ell^* \in Z_k^*, C_{V_k}(\ell) = \emptyset\} \tag{4.40}$$

and $C_{V_k}(i) = j$ denotes that the ith salient scan point corresponds to the jth existing shape point, and $\emptyset$ denotes the null correspondence. The parameters of

the distribution over the continuous partition of the object state are updated as follows:

$$\begin{aligned} \bar{X}^{\mathrm{c}}_{k|\{Z_k, C_{V_k}\}} &= \begin{bmatrix} \bar{X}^{\mathrm{c}}_{k|Z_{k\text{-}1}} \\ \bar{V}^{\mathrm{new}}_{k|C_{V_k}} \end{bmatrix} \\ \boldsymbol{P}_{X^{\mathrm{c}}_k|\{Z_k, C_{V_k}\}} &= \begin{bmatrix} \boldsymbol{P}_{X^{\mathrm{c}}_{k|Z_{k\text{-}1}}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{R}^{\mathrm{surf}}_{V^{\mathrm{new}}_{k|C_{V_k}}} \end{bmatrix} \end{aligned} \tag{4.41}$$

where $\boldsymbol{R}^{\mathrm{surf}}_{V^{\mathrm{new}}_{k|C_{V_k}}}$ is the partition of lidar scan covariance, $\boldsymbol{R}^{\mathrm{surf}}_k$, corresponding the added shape points, $V_{k|C_{V_k}}$.

If the correspondences between existing shape points and the current salient scan points, C_{V_k}, are known, the new shape model connectivity, $E_{k|Z_k}$, is unique, and immediately apparent from additional knowledge of the connectivity within each of the individual contributing point sets, $V_{k|Z_{k\text{-}1}}$ and Z^*_k, as demonstrated in Fig. 4.4; therefore, multiple hypotheses would not be necessary. Note that the connectivity of $V_{k|Z_{k\text{-}1}}$, i.e. $E_{k|Z_{k\text{-}1}}$, is encoded within the parent hypothesis, and the connectivity of Z^*_k is straight forward as discussed in Section 4.7.1: *Connectivity of salient scan points*. However, the correspondences, C_{V_k}, are latent and must be estimated. Accurate estimates of these correspondences are critical in that errors have the potential to cause irrecoverable filter divergence. Therefore, all *reasonable* correspondences are considered as separate, parallel, hypotheses replacing the parent hypothesis, each giving rise to a different set of updated shape points, $V_{k|Z_k, \hat{C}_{V_k}}$, and connectivities, $E_{k|Z_k, \hat{C}_{V_k}}$.

The proposed approach is indifferent to the many sensible strategies for finding the reasonable candidates for $\hat{C}_{V_k}$, generically referred to as *k-best assignment algorithms*, and commonly used within data association routines. In evaluating the proposed work, the Murty algorithm [39] was used in conjunction with the Munkres (Hungarian) optimal assignment algorithm [32] to provide a ranked

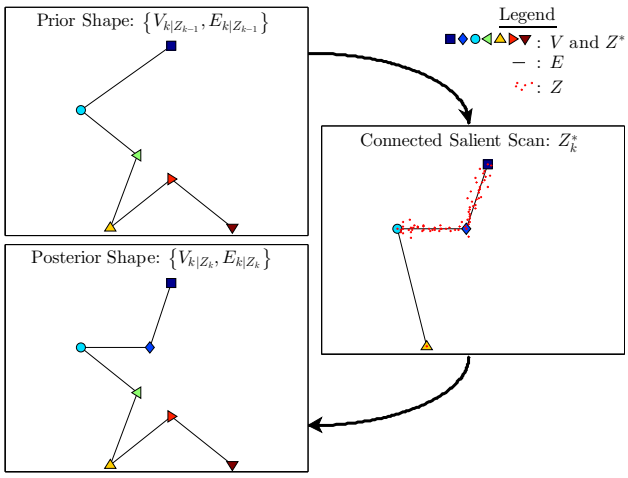

Figure 4.4: Demonstration of the shape model growth to include previously occluded object surfaces revealed with changing viewpoint. Point correspondences are identified between subplots via matching colors and markers. Notice that only the salient scan point that does not correspond to any existing shape points (blue diamond) is added to the shape model, and the posterior connectivity is unique and immediately apparent given the connectivity of the prior and salient scan points, as well as the point correspondences between the two sets.

list of the best candidates for $\hat{C}_{V_k}$; however, other algorithms, perhaps with better computational properties, could be used. The cost function used to evaluate each individual correspondence is:

$$J(i,j) = -\ln\left[p\left(C_{V_k}(i) = j\right)\right] \tag{4.42}$$

where the correspondence likelihood is defined as:

$$p\left(C_{V_k}(i) = j\right) = \mathcal{N}\left(\mathbf{0}_{2\times 1}; z_i^* - \bar{v}_j, \boldsymbol{R}_{z_i^*}^{\text{surf}} + \boldsymbol{P}_{v_j}\right) \tag{4.43}$$

where z_i^* is the ith point in Z_k^* with associated covariance $\boldsymbol{R}_{z_i^*}^{\text{surf}}$, and $v_j \sim \mathcal{N}\left(\bar{v}_j, \boldsymbol{P}_{v_j}\right)$ is the jth existing shape point in $V_k \sim \mathcal{N}\left(\bar{V}_k, \boldsymbol{P}_{V_k|Z_{k-1}}\right)$. The probability of each new hypothesis, $E_{\ell,k}$, one for each sufficiently likely correspondence candidate, $\hat{C}_{V_k,\ell}$, is initialized according to the probability that its unique correspondence

candidate is correct:

$$p\left(E_{\ell,k}|Z_{1:k}\right) = \frac{p\left(C_{V_k} = \hat{C}_{V_k,\ell}\right)}{\sum_{\ell} p\left(C_{V_k} = \hat{C}_{V_k,\ell}\right)} p\left(E_k|Z_{1:k\text{-}1}\right) \tag{4.44}$$

where $p\left(E_k|Z_{1:k\text{-}1}\right)$ is the prior probability that the parent hypothesis is correct, and $p\left(C_{V_k} = \hat{C}_{V_k,\ell}\right)$ is the joint probability of the individual, independent, point assignments making up the ℓth candidate set of correspondences, $\hat{C}_{V_k,\ell}$:

$$p\left(C_{V_k} = \hat{C}_{V_k,\ell}\right) = \prod_{i} p\left(C_{V_k}(i) = \hat{C}_{V_k,\ell}(i)\right) \tag{4.45}$$

where the marginal probabilities within the product are defined in (4.43).

As the shape point estimates, $\hat{V}_k$, within each hypothesis are refined over time given more data, their degree of saliency may diminish. That is, a point that appears salient in the raw, noisy, lidar scan may later prove to be well predicted by its neighboring shape points as the shape model estimate is refined. While these superfluous shape points will not degrade the tracking performance, they add unnecessary complexity to the shape model, and, therefore, should be identified and pruned. These points can be identified using the same feature extraction algorithm used to identify the salient points in the lidar scan; i.e. any shape points that are not identified as *features* of the shape model should be pruned.

4.7.3 Discrete Inference

The discrete inference problem is to recursively estimate the posterior categorical distribution over the hypotheses of E given the data, $p\left(E_k|Z_{1:k}\right)$. The prediction step of the recursion is the identity mapping due to the rigid body assump-

tion, $E_k = E \; \forall k$:

$$p\left(E_k|Z_{1:k\text{-}1}\right) = p\left(E_{k\text{-}1}|Z_{1:k\text{-}1}\right) \tag{4.46}$$

where $p\left(E_k|Z_{1:k\text{-}1}\right)$ is the prior edge belief at time step k, and $p\left(E_{k\text{-}1}|Z_{1:k\text{-}1}\right)$ is the posterior edge belief at time step $k-1$. Using Bayes' rule and the conditional independence assumptions encoded in the HMM of Fig. 4.3, the update step of the recursion can be rewritten as:

$$p\left(E_k|Z_{1:k}\right) = \frac{p\left(Z_k|E_k\right) \cdot p\left(E_k|Z_{1:k\text{-}1}\right)}{\sum\limits_{E_k} p\left(Z_k|E_k\right) \cdot p\left(E_k|Z_{1:k\text{-}1}\right)} \tag{4.47}$$

where $p\left(Z_k|E_k\right)$ is the measurement likelihood defined as:

$$p\left(Z_k|E_k\right) = \mathcal{N}\left(Z_k; \bar{Z}_k, \boldsymbol{P}_{Z_k}\right) \tag{4.48}$$

Note that the update step in (4.47) happens twice with each lidar scan, once for the metadata measurement in (4.26), and once for the surface measurement in (4.27).

To minimize computational effort, hypotheses are pruned when their probability, $p\left(E_k|Z_{1:k}\right)$, drops below a user defined threshold, deeming them sufficiently unlikely to represent the true object shape.

4.8 Baseline Tracker for Simulation and Experiment

To showcase the benefits that detailed shape estimates provide extended object tracking, a stripped down version of the proposed PH-SET filter is chosen as a baseline for comparison. Specifically, the object state is reduced to only the kinematic partition, Θ, and the filter prediction and measurement update procedures are reduced to only those described in (4.18) and (4.26), respectively.

Given that the baseline filter lacks a belief of the object shape, the most recent lidar scan points returned from the object are stored in the reference frame described by the first three kinematic states, (x, y, ϕ), and used to predict the distribution over the object metadata in (4.26) at the next time step to drive the filter metadata measurement update; a method inspired by [38].

Performance differences between the baseline and proposed filters in tracking the kinematic states can be attributed entirely to the more efficient use of sensor information by way of the detailed sensor model presented in (4.27); a feature enabled by the detailed shape estimates.

Note that the benefits of detailed shape estimates extend beyond improving the precision and accuracy of kinematic state estimates; for instance, they provide a rich source of information for higher level reasoning about object type/class, occupied space, etc. However, that higher level reasoning is beyond the scope of this paper, therefore, those benefits are not evaluated.

4.9 Simulation Results

To demonstrate the object tracking performance improvement enabled by the detailed shape estimates, the proposed work is evaluated against the baseline tracker in the simulated planar scenario depicted in Fig. 4.5. The scenario involves a dynamic ego vehicle (blue rectangle) tasked with tracking the dynamic object (green star) relative to its own (ego) coordinate frame. Data is simulated for a 360° field-of-view single-scan planar lidar firing at 12.5 Hz with 0.5° bearing resolution, fixed to centroid of the ego vehicle. Random sensor noise is sampled independently for each beam in each scan from $\mathcal{N}\left(0, 1\ \text{cm}^2\right)$ and added to

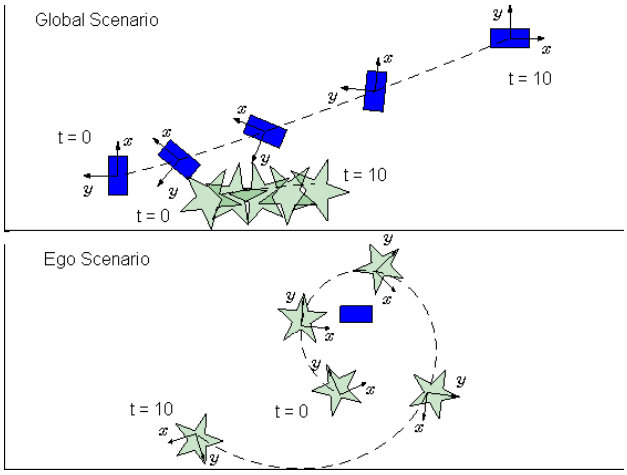

Figure 4.5: Simulated scenario in both global (top) and ego (bottom) coordinates. The blue rectangle denotes the ego (sensory) vehicle, while the green star denotes the tracked object.

the lidar range returns to simulate the accuracy of realistic lidar sensors.

Fig. 4.6 shows the evolution of the maximum-a-posteriori (MAP) object shape estimate, (4.15), for a sampling of seven simulation time steps. Each shape estimate is projected into a common object-fixed coordinate frame for direct comparison, from which the object appears static; the corresponding global scenario, and the scenario as seen from the perspective of the ego vehicle (the coordinate frame from which the kinematic states are defined), are provided in the top and bottom panes of Fig. 4.5, respectively.

Notice the ability of the shape estimate to evolve to include previously occluded object surfaces revealed to the sensor via viewpoint changes. Further, the filter is able to accurately track previously observed shape points during occlusion events, despite a lack of direct observation. For example, the top vertex of the star is observed at the beginning of the simulation, becomes occluded at $t \approx 3.8$ s, and remains occluded until $t \approx 8.56$ s; despite the lack of direct observation for approximately half of the simulation, the tracker is able to maintain

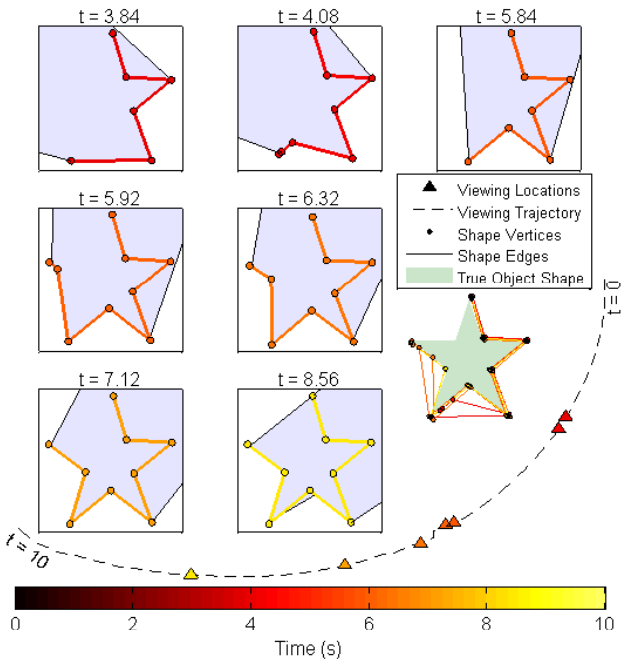

Figure 4.6: Evolution of the MAP star shape estimate projected into a common object-fixed coordinate frame (seven time steps are shown). Each snapshot shows the shape estimate at a unique time step superimposed upon the occlusion shadow cast by the visible surface of the object (light blue region). Beside the snapshots, the full sensor trajectory as seen from the object-fixed coordinate frame is shown, and the shape estimates from each of the seven snapshots are superimposed upon the ground truth shape (green star); the sensor positions corresponding to each of the seven snapshots are denoted with a triangle marker. Sensor positions and shape estimates are pseudo-colored by their time step.

an accurate and precise estimate of the vertex position in space through the unknown and complex dynamics depicted in Fig. 4.5. This characteristic is made possible through accurate modeling of the inherent correlations and dependencies among the visible and occluded shape points, as well as the kinematic states (i.e. inferring the *joint* pdf over the shape and kinematic states). Lastly, the filter does not make assumptions about surfaces that have yet to be observed, accurately reflecting the complete lack of information about portions of the object

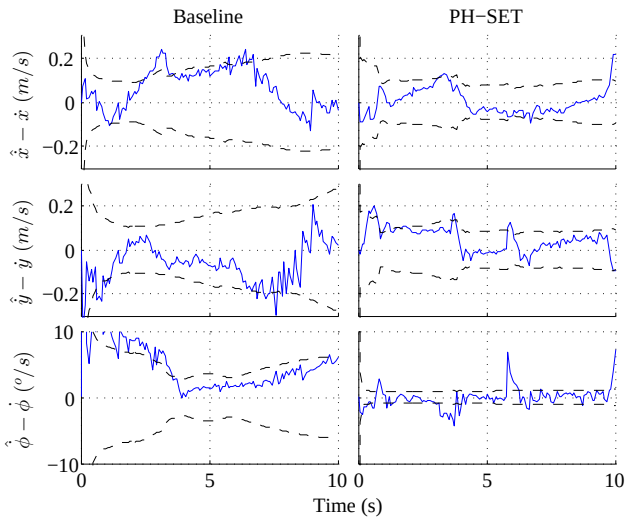

Figure 4.7: Performance in tracking the kinematic velocity states for both the baseline (left) and PH-SET (right) filters. The MAP estimate errors from ground truth is plotted along with the 1-σ covariance bounds estimated by each filter.

shape that have been occluded heretofore. This is apparent in Fig. 4.6, in that all vertices and edges represented in the estimated shape, at any given time, have been directly observed by the sensor at some time in the past; unobserved portions of the object remain absent from the shape estimate until evidence of their existence and appearance warrants their inclusion.

Fig. 4.7 depicts the error in tracking the kinematic velocity states in (4.17), along with their estimated 1-σ bounds, for both the baseline (left) and the PH-SET (right) filters. Note that the PH-SET filter estimate is both more accurate (smaller magnitude errors) and more precise (tighter bounds on the standard deviation of the error) than the baseline filter; a benefit credited to more efficient use of sensor information by way of explicit sensor models derived from detailed shape estimates.

4.10 Experimental Results

The PH-SET approach was also validated using experimental data collected from a Velodyne HDL-64 laser scanner instrumented to Cornell University's autonomous vehicle platform [37]. The Velodyne returns a 3D point cloud of approximately one million range samples per second from 64 lasers sweeping a full 360° horizontal FOV at 15 Hz. Ground returns are classified using a Gaussian Process (GP) trained on hand-labeled data from the first revolution of the Velodyne, and removed from the raw scan data prior to computing the 2D hull measurements defined in (4.2).

The experiment scenario involves a static ego-vehicle tasked with tracking a 1995 Kubota L2850 tractor maneuvering in a natural, non-planar environment. The tractor was chosen as a test case because it has a complex shape that varies greatly with height, deeming it a worthy challenge for the PH-SET approach. Fig. 4.8 depicts the scenario as an overhead view of a sampling of the hull measurements, defined in Section 4.3, strung together with an approximation of the full tractor trajectory (black dashed line); the full trajectory is approximated by tracing a denser set of measurements than shown in the figure. Scan points are pseudo-colored by their height above the ground model to emphasize that the 2D shape model is estimated from data sampled from varying heights on the object.

Fig. 4.9 shows the evolution of the MAP object shape estimate for a sampling of six experimental time steps. Each shape estimate is projected into a common object-fixed coordinate frame for direct comparison, from which the object appears static; the corresponding global scenario is provided in Fig. 4.8.

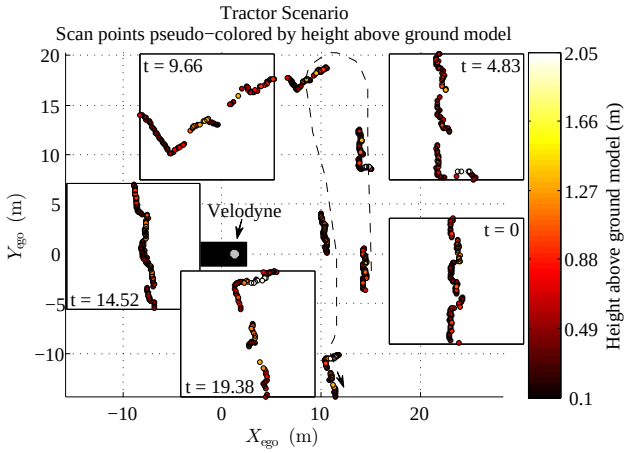

Figure 4.8: Five time snapshots of sensor data returned from the tractor over the 20 s scenario. Sensor frames are strung together with a dashed line denoting the approximate trajectory of the tractor. Lidar scan points are pseudo-colored by their height above the ground model.

As discussed for Fig. 4.6 in reference to the simulation, the shape estimate is able to evolve to include previously occluded object surfaces, accurately tracks previously observed object surfaces during occlusion events, and avoids making assumptions about object surfaces that have yet to be observed. Of particular interest here is that the filter is able to achieve all of this while maintaining an accurate and precise track of the object through a strong, unmodeled, 180° turn and acceleration maneuver from about 7s $< t <$ 12s. This indicates that the filter criterion for adapting the shape model to include new surfaces, i.e. thresholding the mean normalized innovation squared of the surface measurements, (4.37), is robust to other sources of uncertainty in the problem; in this case, strong deviations from the assumed dynamics model.

Fig. 4.10 depicts the estimates and 1-σ bounds of the kinematic velocity states for both the baseline (left) and PH-SET (right) filters. First, notice that the PH-SET filter estimate is more precise than the baseline filter (tighter 1-σ

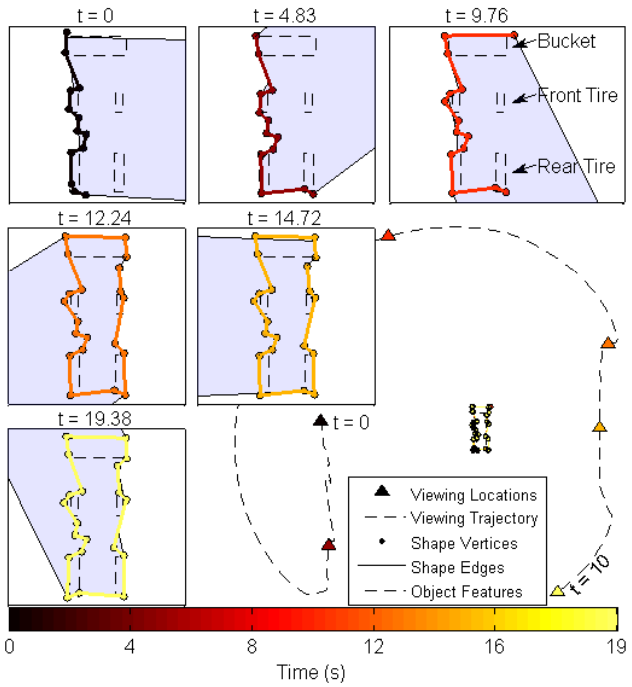

Figure 4.9: Evolution of the MAP tractor shape estimate projected into a common object-fixed coordinate frame (six time steps are shown). Each snapshot shows the shape estimate at a unique time step superimposed upon the occlusion shadow cast by the visible surface of the object (light blue region). Beside the snapshots, the full sensor trajectory as seen from the object-fixed coordinate frame is shown, and the shape estimates from each of the six snapshots are superimposed upon each other; the sensor positions corresponding to each of the six snapshots are denoted with a triangle marker. Sensor positions and shape estimates are pseudo-colored by their time step. Appropriately sized tractor features are placed by hand to orient the viewer, and are labeled in the top right frame.

bounds) due to a more efficient use of the information in each lidar scan. Specifically, via the refined sensor model in (4.27) enabled by the detailed shape estimate, the filter is able to extract information about the object state from *every* lidar point returned from the 2D hull of the object; alternatively, the baseline filter discards the information in each lidar beam that does not contribute to the metadata in (4.26). Next, notice that the final leg of the tractor trajectory in

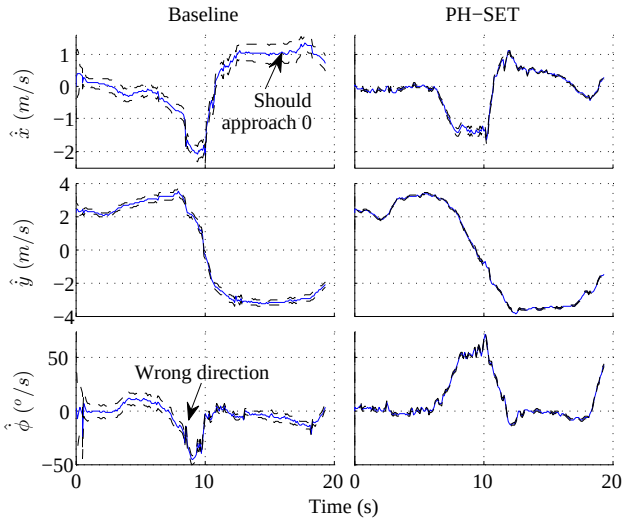

Figure 4.10: Mean and 1-σ bound of the velocity state estimate histories for both the baseline (left) and PH-SET (right) filters.

Fig. 4.8 asymptotically approaches $x \approx 12$ m coming out of the turn ($t \approx 12$ s), followed by a slight dip in x leading into the start of another left-hand turn at $t \approx 18$ s. The PH-SET filter estimate of the velocity, $\hat{\dot{x}}$, (top right of Fig. 4.10) matches this trajectory by gradually reducing after the turn ($t > 12$ s) to a small negative value over the duration of the final leg of the trajectory before climbing positive again; meanwhile, the baseline filter estimate of $\dot{x}$ (top left of Fig. 4.10) incorrectly reflects an approximately constant increase in x position coming out of the turn. Lastly, notice that the baseline filter estimate of $\dot{\phi}$ (bottom left of Fig. 4.10) incorrectly reflects a clockwise turn midway through the trajectory, while the PH-SET filter estimate (bottom right of Fig. 4.10) accurately reflects a counter-clockwise turn, consistent with the true trajectory depicted in Fig. 4.8. This final point is emphasized in Fig. 4.11, which shows the MAP estimate of the tractor orientation, $\hat{\phi}$, (bottom) and its rate of change, $\hat{\dot{\phi}}$, (top) for both filters superimposed. Notice that the baseline filter estimates of ϕ and $\dot{\phi}$ are contradictory in sign, and do not track the full 180° turn present in the true

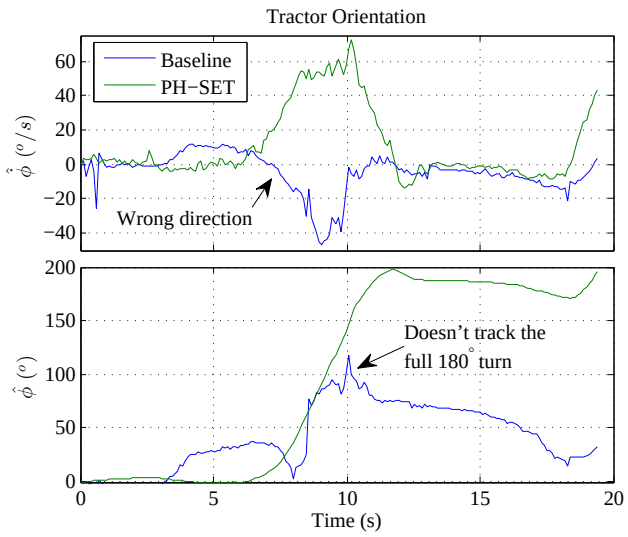

Figure 4.11: MAP estimate of object orientation, ϕ, (bottom) and its time derivative, $\dot{\phi}$, (top) for both the baseline and PH-SET filters, superimposed.

scenario; evidence that the object orientation, decoupled from object heading, is not sufficiently observable without a detailed shape estimate.

4.11 Conclusions

A novel approach to estimating the detailed shape of arbitrary extended objects jointly with their kinematics is proposed, dubbed the *Polygonal Hull Shape Estimation and Tracking* approach, or PH-SET for short. The PH-SET approach is evaluated via simulation and experimental data collected for objects with complex shapes maneuvering in natural, unstructured environments characteristic of autonomous driving applications. The simulation and experimental results demonstrate that detailed shape estimates not only provide a richer set of information for higher level reasoning about extended objects (e.g. about object type, or occupied space), they enable more efficient use of sensor information

by way of detailed sensor models, engendering improved observability of kinematic states, and improved precision and accuracy of kinematic state estimates.

CHAPTER 5

CONCLUSIONS

The research presented in this dissertation is inspired by human perception, and motivated by the recent effort to integrate fully autonomous vehicles alongside human driven vehicles within the existing transportation infrastructure. Specifically, the research presented here is aimed at improving the accuracy and precision in tracking arbitrary, uncooperative extended objects without *a priori* knowledge of the environment (e.g. the road network, traffic laws, etc.), or objects (e.g. type, shape, size, etc.). In general, the common underlying mechanism of the proposed work is the use of novel, detailed, non-traditional sensor models to interpret both the positive and negative information inherently present in the sensor data, and how that information relates to the extended object state. The use of these detailed sensor models mitigates the information loss inherent in interpreting the data returned from extended objects, thereby improving the accuracy and precision of object state estimates, and facilitating the overall safety of the robotic system. Specifically, Chapter 2 presents a rigorous probabilistic framework for modeling and fusing the inherently present, but traditionally ignored, *negative information* associated with occlusion events; an analog to human occlusion reasoning. The approach is complementary to existing Bayesian inference-based object trackers, and is based on a novel probabilistic occlusion model naturally derived from tracked random variables. Chapter 3 presents a novel hierarchical, priority-based tracking framework to allocate computational resources in proportion to the time-dependent relevancy of the object to the ego robot task; an analog to the human concepts of *attention* and *focus* in allocating finite cognitive resources. Probabilistic object relevancy metrics are developed to inform a hybrid system switching strategy, resulting in detailed,

high precision tracking of objects relevant to the ego robot task, and computationally efficient tracking of objects peripheral to the ego robot task. Lastly, Chapter 4 presents a rigorous multi-hypothesis framework for estimating an object's shape jointly with its kinematics. The detailed object shape estimates enable explicit and intricate sensor models that are precise in their interpretation of the sensor data, thereby engendering precise, low entropy, beliefs of the object's kinematic states. The work proposed here is evaluated in simulation and experiments involving Cornell University's autonomous Chevy Tahoe, *Skynet*, demonstrating that non-traditional sensor models are necessary for high precision tracking of arbitrary extended objects, and that humans are worth emulating en route to this goal. Specific conclusions pertaining to each contribution can be found at the end of each respective chapter.

In summary, the following is a list of key novel contributions to the field of extended object tracking presented in this dissertation:

- perspective on the analogy between human and robot perception as it pertains to EOT
- a rigorous probabilistic occlusion model derived naturally from tracked random variables
- a rigorous negative information fusion framework that is complementary to traditional probabilistic inference strategies for estimating:
  - object kinematic state (Gaussian or mixture of Gaussians)
  - object existence (Bernoulli)
  - measurement-object correspondences (Categorical)

- a hierarchical, priority-based EOT framework to emulate the human perception attributes of focus and attention in allocating finite cognitive resources
- object relevancy metrics to drive mode transitions in the priority-based hybrid tracking framework
- an efficient, qualitative tracking approach for objects of peripheral importance to the ego robot task
- perspective on the fundamental distinction between *extended* and *group* objects, and how this distinction can be leveraged to improve the performance in tracking extended objects
- a non-parametric object shape model for arbitrary extended objects
- a method for extracting a 2D measurement of the hull of an extended object from a full 3D lidar scan
- Polygonal Hull Shape Estimation and Tracking (PH-SET): a rigorous probabilistic filtering framework for inferring the joint distribution over object kinematic state and non-parametric shape model for arbitrarily shaped extended objects, without requiring *a priori* information about object size, shape, or type

APPENDIX A

RELATIONSHIP BETWEEN TWO GAUSSIAN RANDOM VARIABLES

Given,

$$\Delta x = x_1 - x_2$$

$$x_1 \sim \mathcal{N}\left(\bar{x}_1, \sigma_{x_1}^2\right)$$

$$x_2 \sim \mathcal{N}\left(\bar{x}_2, \sigma_{x_2}^2\right)$$

then,

$$p\left(x_2 \geq x_1\right) = p\left(\Delta x \leq 0\right)$$

$$\Delta x \sim \mathcal{N}\left(\bar{x}_1 - \bar{x}_2, \sigma_{x_1}^2 + \sigma_{x_2}^2\right)$$

Therefore, it follows that:

$$p\left(x_2 \geq x_1\right) = \mathbf{\Phi}\left(\frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2}}\right) \tag{A.1}$$

where $\mathbf{\Phi}\left(\cdot\right)$ is the standard normal cumulative distribution function (CDF):

$$\begin{aligned} \mathbf{\Phi}\left(x\right) &= \frac{1}{\sqrt{2\pi}} \int_{-\inf}^{x} \exp\left(\frac{-t^2}{2}\right) dt \\ &= \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right] \end{aligned} \tag{A.2}$$

APPENDIX B

PROBABILITY OF THE UNION OF TWO EVENTS

The probability of the union of two events, x and y, is defined as:

$$p(x \cup y) = p(x) + p(y) - p(x \cap y) \tag{B.1}$$

where $p(x \cap y)$ is the joint probability:

$$p(x \cap y) = p(x) \cdot p(y|x)$$

$$p(y|x) = \begin{cases} p(y) & \text{if independent} \\ 0 & \text{if mutually exclusive} \end{cases}$$

APPENDIX C

RATE FUNCTION PARAMETERS

The rate function parameters introduced in Section 2.3.2 in Fig. 2.7, and assumed in this work, are provided in Table C.1. A summary of each parameter is provided below, where f_s denotes sensor frequency:

- $\underline{\gamma}$: the maximum rate parameter for undetected tracks.

$$\lim_{z_n^{\text{exist}} \to 0} \mathbf{\Gamma}\left(z_n^{\text{exist}}\right) = \underline{\gamma}$$

  Setting $\underline{\gamma} = 2\frac{f_s}{N_{\text{scan}}}$ results in undetected tracks perennially believed to be visible ($z_n^{\text{exist}} \leq \underline{z}_{0.99}^{\text{exist}}$) to be removed after N_{scan} sensor scans.

- $\overline{\gamma}$: the minimum rate parameter for undetected tracks.

$$\lim_{z_n^{\text{exist}} \to 1} \mathbf{\Gamma}\left(z_n^{\text{exist}}\right) = \overline{\gamma}$$

  Setting $\overline{\gamma} = 0$ allows undetected tracks perennially believed to be occluded ($z_n^{\text{exist}} \geq \overline{z}_{0.99}^{\text{exist}}$) to maintain a constant existence probability.

- $\tilde{z}^{\text{exist}}$: location of the inflection point of $\mathbf{\Gamma}(\cdot)$ for undetected tracks; a soft threshold on the interpretation of the existence measurement, such that $z^{\text{exist}} < \tilde{z}^{\text{exist}}$ corresponds to a measurement of track visibility, and $z^{\text{exist}} > \tilde{z}^{\text{exist}}$ corresponds to a measurement of track occlusion.

$$\mathbf{\Gamma}\left(\tilde{z}^{\text{exist}}\right) = \frac{1}{2}\left(\underline{\gamma} + \overline{\gamma}\right)$$

  As $\tilde{z}^{\text{exist}} \to 1$ the existence model becomes increasingly assertive, applying a higher standard for undetected tracks to persist.

- α: determines the shape of $\mathbf{\Gamma}(\cdot)$. As $\alpha \to \infty$, the existence model becomes more authoritative, applying a more definitive threshold on the interpretation of z^{exist}. As $\alpha \to 0$, the existence model approaches the baseline

Table C.1: Existence Model Parameters

Parameter	Valid Range	Suggested Value
$\underline{\gamma}$	$(\overline{\gamma}, \infty)$	$\frac{2}{3} f_s$
$\overline{\gamma}$	$\left[0, \underline{\gamma}\right)$	0
$\tilde{z}^{\text{exist}}$	$(0, 1)$	0.5
α	$[0, \infty)$	7
γ_g	$[0, \infty)$	$3 f_s$

tracker [38], applying a common rate parameter for all undetected tracks, regardless of their measured existence.

- γ_g: growth constant of the existence model for detected tracks. For $\gamma_g = 0$, track existence probability is preserved for tracks detected in sensor data, despite the positive evidence. For $\gamma_g \approx \infty$, the existence probability updates to 1 for all detected tracks, regardless of their existence probability prior, or measured existence.

To provide some insight into the influence of α and $\tilde{z}^{\text{exist}}$ on the rate parameter function, (2.29), the interval, $\left[\underline{z}_\xi^{\text{exist}}, \overline{z}_\xi^{\text{exist}}\right]$, over which $\xi \cdot 100\%$ of the rate parameter transient from $-\underline{\gamma}$ at $z_n^{\text{exist}} = 0$ to $-\overline{\gamma}$ at $z_n^{\text{exist}} = 1$, demonstrated in Fig. 2.7, can be computed by considering the following two conditions:

$$\mathbf{\Gamma}\left(\underline{z}_\xi^{\text{exist}}\right) - \mathbf{\Gamma}\left(\overline{z}_\xi^{\text{exist}}\right) = \xi\left(\underline{\gamma} - \overline{\gamma}\right)$$
$$\mathbf{\Gamma}\left(\underline{z}_\xi^{\text{exist}}\right) + \mathbf{\Gamma}\left(\overline{z}_\xi^{\text{exist}}\right) = \underline{\gamma} + \overline{\gamma}$$

Combining these conditions forms the following uncoupled expressions for the rate parameter at the left and right interval boundaries, respectively:

$$\mathbf{\Gamma}\left(\underline{z}_\xi^{\text{exist}}\right) = \frac{1}{2}\left[(1+\xi)\,\underline{\gamma} + (1-\xi)\,\overline{\gamma}\right]$$
$$\mathbf{\Gamma}\left(\overline{z}_\xi^{\text{exist}}\right) = \frac{1}{2}\left[(1-\xi)\,\underline{\gamma} + (1+\xi)\,\overline{\gamma}\right]$$

which can be solved using the $D_n = 0$ case of (2.29), to get:

$$\begin{aligned} \underline{z}_\xi^{\text{exist}} &= \frac{1}{1 + \left(\frac{1}{\tilde{z}^{\text{exist}}} - 1\right) \exp\left[-\frac{1}{\alpha} \ln\left(\frac{1-\xi}{1+\xi}\right)\right]} \\ \overline{z}_\xi^{\text{exist}} &= \frac{1}{1 + \left(\frac{1}{\tilde{z}^{\text{exist}}} - 1\right) \exp\left[-\frac{1}{\alpha} \ln\left(\frac{1+\xi}{1-\xi}\right)\right]} \end{aligned} \tag{C.1}$$

Note that as $\xi \to 1$, $\underline{z}_\xi^{\text{exist}} \to 0$ and $\overline{z}_\xi^{\text{exist}} \to 1$, as expected due to (2.30).

APPENDIX D

LOG-OF-SUMS TRICK

Consider an equation of the form:

$$y = \sum_{\ell=1}^{N} f(\ell) \tag{D.1}$$

where $f(\ell) \in (0, \infty)$ may be too small to compute accurately due to numerical underflow. If $\alpha_\ell = \ln(f(\ell))$ can be computed accurately $\forall\, \ell$, then the numerical stability of (D.1) can be improved by using the *log-of-sums* trick. Specifically,

$$y = \exp(\ln(y)) \tag{D.2}$$

where:

$$\begin{aligned} \ln(y) &= \ln\left\{\sum_{\ell=1}^{N} f(\ell)\right\} \\ &= \ln\left\{\sum_{\ell=1}^{N} \exp(\alpha_\ell)\right\} \\ &= \ln\left\{\exp\left(\max_{\ell}(\alpha_\ell)\right) \cdot \sum_{\ell=1}^{N} \frac{\exp(\alpha_\ell)}{\exp\left(\max_{\ell}(\alpha_\ell)\right)}\right\} \\ &= \max_{\ell}(\alpha_\ell) + \ln\left\{\sum_{\ell=1}^{N} \exp\left(\alpha_\ell - \max_{\ell}(\alpha_\ell)\right)\right\} \end{aligned} \tag{D.3}$$

Notice that the largest argument to the exponential function on the last line of (D.3) will be 0, eliminating the possibility of overflowing the exponential. Further, the use of α_ℓ rather than $f(\ell)$ in (D.3) improves the numerical stability compared to (D.1).

A common application is when $f(\ell)$ is a $d-$dimensional Gaussian distribution:

$$f(\ell) = \mathcal{N}(X_\ell; \mu_\ell, \mathbf{\Sigma}_\ell) \tag{D.4}$$

in which case α_ℓ can be computed directly as:

$$\alpha_\ell = -\ln\left(\sqrt{(2\pi)^d \cdot |\boldsymbol{\Sigma}_\ell|}\right) - \frac{1}{2}(X_\ell - \mu_\ell)^T \boldsymbol{\Sigma}_\ell^{-1}(X_\ell - \mu_\ell) \tag{D.5}$$

D.1 Kalman Filter

This section provides the Kalman filter prediction and update equations [6, 15, 46].

D.1.1 Prediction

Given a stochastic linear vector differential equation model the form:

$$\begin{aligned} \dot{X}_k &= \boldsymbol{F}_k X_k + \boldsymbol{G}_k W_k \\ W_k &\sim \mathcal{N}\left(\bar{W}_k, \boldsymbol{Q}_k\right) \end{aligned} \tag{D.6}$$

describing the object dynamics, and a Gaussian belief of the posterior object state at time $k-1$,

$$p\left(X_{k\text{-}1} | Z_{1:k\text{-}1}\right) = \mathcal{N}\left(\bar{X}_{k\text{-}1|k\text{-}1}, \boldsymbol{P}_{k\text{-}1|k\text{-}1}\right) \tag{D.7}$$

the Gaussian prior distribution at time k is obtained by predicting the posterior at $k-1$ over the time interval δt:

$$p\left(X_k | Z_{1:k\text{-}1}\right) = \mathcal{N}\left(\bar{X}_{k|k\text{-}1}, \boldsymbol{P}_{k|k\text{-}1}\right) \tag{D.8}$$

where the mean and covariance are computed using the following equations:

$$\begin{aligned} \bar{X}_{k|k\text{-}1} &= \int_{k-1}^{k} \dot{X}_t \, \mathrm{d}t \\ \boldsymbol{P}_{k|k\text{-}1} &= \boldsymbol{\Phi}_{k-1} \boldsymbol{P}_{k\text{-}1|k\text{-}1} \boldsymbol{\Phi}_{k-1}^T + \boldsymbol{\Gamma}_{k-1} \left(\boldsymbol{Q}_k \cdot \delta t\right) \boldsymbol{\Gamma}_{k-1}^T \end{aligned} \tag{D.9}$$

where:

$$\boldsymbol{\Phi}_{k\text{-}1} = \int_{k-1}^{k} \boldsymbol{F}_t \, \mathrm{d}t \text{ , and } \quad \boldsymbol{\Gamma}_{k\text{-}1} = \int_{k-1}^{k} \boldsymbol{G}_t \, \mathrm{d}t \tag{D.10}$$

$\bar{X}_{k|k\text{-}1}$, $\boldsymbol{\Phi}_{k\text{-}1}$, and $\boldsymbol{\Gamma}_{k\text{-}1}$ are computed using numerical integration techniques, such as Runge-Kutta.

D.1.2 Update

Given that the object state, X_k, and measurement, Z_k, are jointly Gaussian:

$$p\left(X_k, Z_k | Z_{1:k\text{-}1}\right) = \mathcal{N}\left(\begin{bmatrix} \bar{X}_{k|k\text{-}1} \\ \bar{Z}_k \end{bmatrix}, \begin{bmatrix} \boldsymbol{P}_{k|k\text{-}1}, \boldsymbol{P}_{X_{k|k\text{-}1}Z_k} \\ \boldsymbol{P}^T_{X_{k|k\text{-}1}Z_k}, \ \boldsymbol{P}_{Z_k} \end{bmatrix}\right) \tag{D.11}$$

where $\bar{X}_{k|k\text{-}1}$ and $\boldsymbol{P}_{k|k\text{-}1}$ are the mean and covariance of the prior state distribution computed in (D.9), and $\bar{Z}_k$, $\boldsymbol{P}_{Z_k}$, and $\boldsymbol{P}_{X_{k|k\text{-}1}Z_k}$ are the measurement mean, covariance, and state-measurement covariance derived from the particular sensor model. Then, the posterior distribution over the object state conditioned on the measurement is also Gaussian:

$$p\left(X_k | Z_{1:k}\right) = \mathcal{N}\left(\bar{X}_{k|k}, \boldsymbol{P}_{k|k}\right) \tag{D.12}$$

with parameters computed as:

$$\begin{aligned} \bar{X}_{k|k} &= \bar{X}_{k|k-1} + \boldsymbol{K}_k\left(Z_k - \bar{Z}_k\right) \\ \boldsymbol{P}_{k|k} &= \boldsymbol{P}_{k|k\text{-}1} - \boldsymbol{K}_k \boldsymbol{P}_{Z_k} \boldsymbol{K}_k^T \end{aligned} \tag{D.13}$$

where $\boldsymbol{R}_k$ is the measurement noise covariance provided by the sensor specification, and the Kalman gain, $\boldsymbol{K}_k$, is defined as:

$$\boldsymbol{K}_k = \boldsymbol{P}_{X_{k|k\text{-}1}Z_k} \boldsymbol{P}^{-1}_{Z_k} \tag{D.14}$$

BIBLIOGRAPHY

[1] Nisar R. Ahmed, Eric M. Sample, and Mark Campbell. Bayesian multicategorical soft data fusion for human-robot collaboration. *IEEE Transactions on Robotics*, 29(1):189–206, February 2013.

[2] D. Alspach and H. Sorenson. Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Transactions on Automatic Control*, 17(4):439 – 448, 1972.

[3] Anton Andriynko, Stefan Roth, and Konrad Schindler. An analytical formulation of global occlusion reasoning for multi-object tracking. In *Proceedings of the 13th IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 2011.

[4] Daniel Arbuckle, Andrew Howard, and Maja Mataric. Temporal occupancy grids: a method for classifying the spatio-temporal properties of the environment. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*, pages 409–414, October 2004.

[5] Jan Aue, Matthias R. Schmid, Thorsten Graf, Jan Effertz, and Peter Muehlfellner. Object tracking from medium level stereo camera data providing detailed shape estimation using local grid maps. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, June 2013.

[6] Yaakov Bar-Shalom, X. Rong Li, and Thiagalingam Kirubarajan. *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Inc., 2001.

[7] Marcus Baum, Michael Feldmann, Dietrich Franken, Uwe D. Hanebeck, and Wolfgang Koch. Extended object and group tracking: A comparison of random matrices and random hypersurface models. In *Proceedings of the IEEE ISIF Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF 2010)*, October 2010.

[8] Marcus Baum and Uwe D. Hanebeck. Extended object tracking based on combined set-theoretic and stochastic fusion. In *Proceedings of the 12th International Conference on Information Fusion (FUSION)*, pages 1288 – 1295, July 2009.

[9] Marcus Baum and Uwe D. Hanebeck. Random hypersurface models for extended object tracking. In *Proceedings of the 9th IEEE International Symposium on Signal Processing and Information Technology (ISSPIT 2009)*, December 2009.

[10] Marcus Baum and Uwe D. Hanebeck. Tracking an extended object modeled as an axis-aligned rectangle. In *Proceedings of Informatik 2009*, 2009.

[11] Marcus Baum and Uwe D. Hanebeck. Shape tracking of extended objects and group targets with star-convex rhms. In *Proceedings of the International Conference on Information Fusion (FUSION 2011)*, July 2011.

[12] Marcus Baum, Benjamin Noack, and Uwe D. Hanebeck. Extended object and group tracking with elliptic random hypersurface models. In *Proceedings of the International Conference on Information Fusion (FUSION 2010)*, July 2010.

[13] Adrian N. Bishop and Branko Ristic. Fusion of natural language propositions: Bayesian random set framework. In *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, pages 1–8, 2011.

[14] Wayne R. Blanding, Wolfgang Koch, and Ulrich Nickel. Adaptive phased-array tracking in ecm using negative information. In *IEEE Transactions on Aerospace and Electronic Systems*, volume 45, 2009.

[15] John L. Crassidis and John L. Junkins. *Optimal Estimation of Dynamic Systems*. Chapman & Hall/CRC Press, 2004.

[16] James E. Cutting. Reconceiving perceptual space. In *Looking Into Pictures: An Interdisciplinary Approach to Pictorial Space*, chapter 11. MIT Press, June 2003.

[17] James E. Cutting and Peter M. Vishton. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In *Handbook of perception and cognition, Vol. 5; Perception of space and motion*, pages 69–117. Academic Press, 1995.

[18] M. Feldmann and D. Franken. Tracking of extended objects and group targets using random matrices - a new approach. In *Proceedings of the 11th International Conference on Information Fusion (FUSION 2008)*, pages 1–9, July 2008.

[19] K. Gilholm and D. Salmond. Spatial distribution model for tracking extended objects. In *IEE Proceedings - Radar, Sonar, and Navigation*, volume 152, pages 364–371, October 2005.

[20] K. Granstrom, C. Lundquist, and O. Orguner. Extended target tracking

using a gaussian-mixture phd filter. *Aerospace and Electronic Systems, IEEE Transactions on*, 48(4):3268–3286, October 2012.

[21] Karl Granstrom and Omut Orguner. A phd filter for tracking multiple extended targets using random matrices. *IEEE Transactions on Signal Processing*, 60(11):5657 – 5671, 2012.

[22] Jason Hardy and Mark Campbell. Contingency planning over probabilistic obstacle predictions for autonomous road vehicles. *Transactions on Robotics*, 29(4):913–929, 2013.

[23] Frank Havlak and Mark Campbell. Discrete and conitnuous, probabilistic anticipation for autonomous robots in urban environments. *Transactions on Robotics*, 30(2):461–474, Decemberl 2013.

[24] David Held, Jesse Levinson, and Sebastian Thrun. Precision tracking with sparse 3d and and dense color 2d data. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2013)*, May 2013.

[25] David Held, Jesse Levinson, Sebastian Thrun, and Silvio Savarese. Combining 3d shape, color, and motion for robust anytime tracking. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2014)*, June 2014.

[26] Jan Hoffmann, Michael Spranger, Daniel Gohring, and Matthias Jungel. Making use of what you don't see: Negative information in markov localization. In *In Proceedings of the 2005 IEEE/RSJ International Conference of Intelligent Robots and Systems (IROS). IEEE*, 2006.

[27] Eddy Ilg, Rainer Kummerle, Wolfram Burgard, and Thomas Brox. Reconstruction of rigid body models from motion distorted laser range data using optical flow. In *Proceedings of the Robotics Science and Systems Conference (RSS 2014)*, July 2014.

[28] S.J. Julier and J.K. Uhlmann. A new extension of the kalman filter to nonlinear systems. In *Proceedings of the 11th International Symposium on Aerospace/Defence*, pages 401–422, 1997.

[29] Wolfgang Koch. On 'negative' information in tracking and sensor data fusion: Discussion of selected examples. In *Proceedings of the Seventh International Conference on Information Fusion*, pages 91–98, 2004.

[30] Wolfgang Koch. Bayesian approach to extended object and cluster tracking using random matrices. *IEEE Transactions on Aerospace and Electronic Systems*, 44(3):1042–1059, 2008.

[31] Wolfgang Koch and Michael Feldmann. Cluster tracking under kinematical constraints using random matrices. *Robotics and Autonomous Systems*, 57(3):296 – 309, 2009.

[32] H.W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistic Quarterly*, 2:83 – 97, 1955.

[33] Michael S. Landy, L. T. Maloney, E.B. Johnston, and M.J. Young. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, 35(3):389–412, February 1995.

[34] Christian Lundquist, Karl Granstrom, and Omut Orguner. Estimating object shape of targets with a phd filter. In *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, pages 1–8, July 2011.

[35] R Luo and M Kay. Data fusion and sensor intergration: state-of-the-art. In *1990s, Data Fusion in Robotics and Machine Intelligence*, pages 7–136, 1992.

[36] Mark McClelland, Tara Estlin, and Mark Campbell. Qualitative relational mapping for planetary rover exploration. In *Proceedings of the AIAA Guidance, Navigation and Control Conference*, 2013.

[37] Isaac Miller, Mark Campbell, et al. Team cornell's skynet: Robust perception and planning in an urban environment. *Journal of Field Robotics*, 25(8):493–527, 2008.

[38] Isaac Miller, Mark Campbell, and Daniel Huttenlocher. Efficient, unbiased tracking of multiple dynamic obstacles under large viewpoint changes. *IEEE Transactions on Robotics*, 27(1):29–46, 2011.

[39] K. G. Murty. An algorithm for ranking all the assignment in order of increasing cost. *Operations Research*, 16:682–687, 1968.

[40] Anna Petrovskaya and Sebastian Thrun. Model based vehicle detection and tracking for autonomous urban driving. *Autonomous Robots Journal*, 26(2-3):123–139, 2009.

[41] Matthew Powers, Ramprasad Ravichandran, Frank Dellaert, and Tucker

Balch. Improving multirobot multitarget tracking by communicating negative information. In Lynne Parker, Frank Schneider, and Alan Schultz, editors, *Multi-Robot Systems. From Swarms to Intelligent Automata Volume III*, pages 107–117. Springer Netherlands, 2005.

[42] Mark L. Psiaki, Jonathan R. Schoenberg, and Isaac T. Miller. Gaussian mixture approximation by another Gaussian mixture for "blob" filter resampling. In *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, pages 1–27, 2010.

[43] S. Reuter and K. Dietmayer. Pedestrian tracking using random finite sets. In *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, pages 1–8, July 2011.

[44] Jari Saarinen, Henrik Andreasson, and Achim Lilienthal. Independent markov chain occupancy grid maps for representation of dynamic environment. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*, pages 3489 – 3495, October 2012.

[45] John Schoenberg, Mark Campbell, and Isaac Miller. Posterior representation with a multi-modal likelihood using the Gaussian sum filter for localization in a known map. *Journal of Field Robotics*, 29:240–257, 2012.

[46] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics*. MIT Press, 2006.

[47] Karin Tischler and Heike S. Vogt. A sensor data fusion approach for the integration of negative information. In *2007 10th International Conference on Information Fusion*, July 2007.

[48] Kevin Wyffels and Mark Campbell. Modeling and fusing negative information for dynamic extended multi-object tracking. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2013)*, May 2013.

[49] Kevin Wyffels and Mark Campbell. Negative observations for multiple hypothesis tracking of dynamic extended objects. In *Proceedings of the American Controls Conference (ACC 2014)*, June 2014.

[50] Kevin Wyffels and Mark Campbell. Joint tracking and non-parametric shape estimation of arbitrary extended objects. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–8, May 2015.

[51] Kevin Wyffels and Mark Campbell. Negative information for occlusion reasoning in dynamic extended multiobject tracking. *IEEE Transactions on Robotics*, 31(2):425–442, April 2015.