Priority-Based Tracking of Extended Objects

KEVIN WYFFELS
MARK CAMPBELL

Inspired by human perception, a novel framework for dynamically allocating algorithmic and computational resources to achieve variable precision tracking of extended objects is presented. Probabilistic object relevancy metrics reflect the priority of each tracked object to the consumer of the tracking output, and are leveraged to trigger mode transitions in a hybrid system implementation of the proposed priority-based framework. In this way, the bulk of the algorithmic and computational resources are reserved for tracking objects of highest priority with high-precision methods, while low priority objects are tracked with inexpensive, qualitative methods. An example implementation of the proposed framework is provided for an autonomous driving application, in which the consumer of the tracking output is an anticipatory path planner. Simulation results demonstrate the ability of the framework to automatically trade computational complexity for tracking precision as a function of an object's priority to the tracking consumer.

Manuscript received October 31, 2015; revised February 18, 2016; released for publication May 10, 2016.

Refereeing of this contribution was handled by Karl Granstrom.

Authors' address: Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY, 14850 USA (E-mail: {klw94, mc288}@cornell.edu).

1557-6418/17/$17.00 © 2017 JAIF

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 [12], [13], [25], [27]. 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.

Given that computers/robots with finite computational resources are often employed to perform human tasks such as navigation or driving, many of the requirements of human perception discussed above apply equally well to computer/robotic perception. However, analogs to the complementary human perception characteristics are largely absent from robotic algorithms. Therefore, inspired by human cognition, this work proposes a priority-based framework for allocating algorithmic and computational resources as a function of *priority* in extended object tracking (EOT); the automatic allocation of computational resources as a function of priority is a novel contribution to the EOT field.

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. 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. While the proposed priority-based framework is general enough to

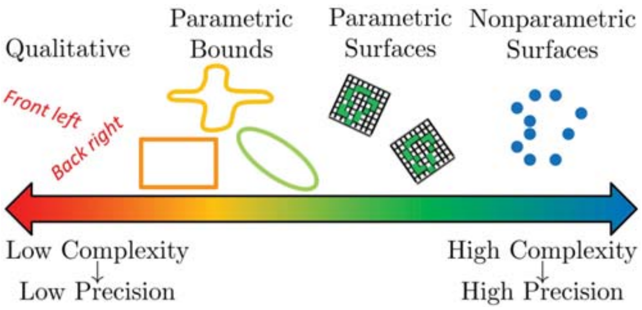

Fig. 1. Distilled spectrum of extent model complexity and tracking precision, which tend to be directly correlated.

consider any EOT methods, the authors are particularly interested in the general case in which the extended object size and shape is unknown a priori. Therefore, tracking approaches that rely on this information, such as spatial distribution models [14], are not included in the following discussion of prior work.

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. 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 that provided by monocular cameras or human input [28]. In a conceptually related approach, traditional static occupancy grid mapping concepts have been extended to characterize dynamic scenes [2], [32]. These approaches tend to be extremely efficient, but their qualitative/topological state representation is too imprecise for many applications, such as those requiring agents to safely 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 [6], ellipses [5], [7], [10], [23], [24], and rectangles [8], [26], [31]. 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, which is the goal of star convex random hypersurface models (RHM) [9]. 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) [3]. 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 [18]–[20], [29], [30], [34]–[37]. These models are extremely flexible in providing rich, 2 or 3-dimensional (3D) renderings 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.

Akin to human perception, within a given EOT application the *relevance* of each object to the consumer of the tracking output, henceforth referred to as *the consumer*, may vary from object-to-object or from instant-to-instant. For example, in a surveillance application, objects exhibiting anomalous behavior may be more relevant than those exhibiting benign behavior; in navigation, nearby objects may be more relevant than distant objects; or in a pursuit application, the lead object may be more relevant than other followers. In these cases, the EOT requirements may also vary with object relevance, deeming a single appropriate EOT method difficult to identify.

This work addresses this issue by proposing a *priority*-based tracking framework for extended objects, where *priority* refers to the object's relevance to the consumer. The proposed framework is implemented via a hybrid system model in which each discrete mode represents a different EOT method with unique characteristics on the complexity-precision spectrum of Fig. 1. Further, probabilistic object relevancy metrics are designed to reflect the time-varying priority of each object to the consumer, and leveraged to inform the hybrid system switching strategy. In this way, objects most relevant to the consumer are allocated more resources and tracked with higher precision, while objects of peripheral relevance are efficiently accounted for with minimal computational burden. To demonstrate the ability of the framework to prioritize objects by automatically trading computation for tracking precision, an example implementation of the hybrid framework is provided for an autonomous driving application in which the consumer is an anticipatory planner.

Section 2 formally defines the general object tracking problem, Section 3 introduces the proposed hybrid system implementation of the priority-based EOT framework, Sections 4 and 5 provide an example implementation of the hybrid framework for an autonomous driving application, Section 6 presents a discussion of simulation results, and finally 7 provides some concluding remarks.

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,

$$p(X_{1:K}^{1:N_O} \mid Z_{1:K}) = \prod_{n=1}^{N_O} p(X_{1:K}^{n} \mid Z_{1:K}, X_{1:K}^{1:n-1}) \tag{1}$$

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, $p(X_{1:K}^n \mid X_{1:K}^m) = p(X_{1:K}^n)\ \forall n \neq m$, 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:

$$p(X_{1:K}^{1:N_O} \mid Z_{1:K}) = \prod_{n=1}^{N_O} p(X_{1:K}^{n} \mid Z_{1:K}) \tag{2}$$

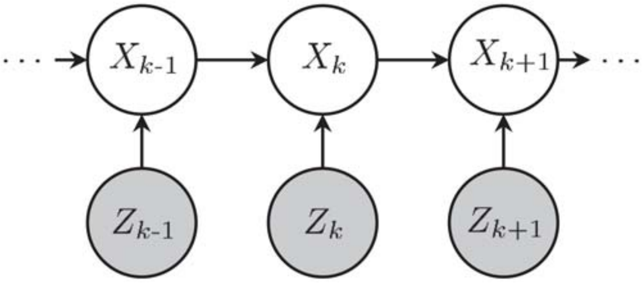

Fig. 2. Graphical representation of the Hidden Markov model (HMM) used to represent the single-object tracking problem. Shaded nodes denote observed variables, and unshaded nodes denote the hidden, i.e. latent, variables to be estimated.

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

$$p(X_{1:K} \mid Z_{1:K}) \tag{3}$$

Further, computation of the posterior in (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. 2:

$$\begin{aligned} p(Z_k \mid X_k, Z_\ell) &= p(Z_k \mid X_k) \quad \forall \ell \neq k \\ p(X_k \mid X_{k-1}, X_{k-\ell}) &= p(X_k \mid X_{k-1}) \quad \forall \ell > 1 \end{aligned} \tag{4}$$

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:

$$p(X_k \mid Z_{1:k}) \quad \forall k \in \{1, \ldots, K\} \tag{5}$$

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. 2, (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, $p(X_{k-1} \mid Z_{1:k-1})$, is propagated forward in time via the prescribed stochastic object dynamics model represented by the transition density, $p(X_{k-1}, X_k)$. The result is the prior state belief at the current time step: $p(X_k \mid Z_{1:k-1})$.
2) *Update step*: the prior state belief at the current time step, $p(X_k \mid Z_{1:k-1})$, 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: $p(X_k \mid Z_{1:k})$.

Estimating (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).

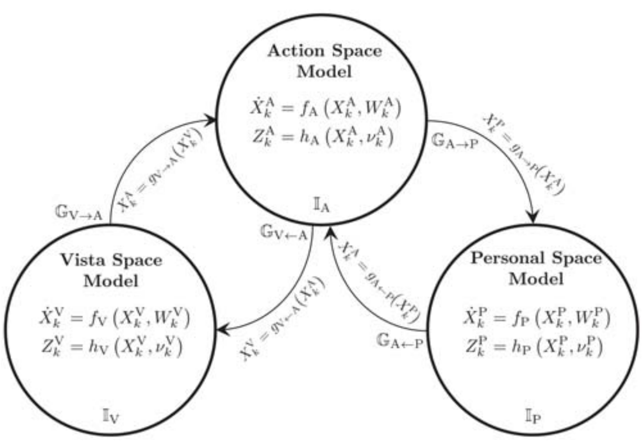
Action Space Model
$\dot{X}_k^{\mathrm{A}} = f_{\mathrm{A}}\left(X_k^{\mathrm{A}}, W_k^{\mathrm{A}}\right)$
$Z_k^{\mathrm{A}} = h_{\mathrm{A}}\left(X_k^{\mathrm{A}}, \nu_k^{\mathrm{A}}\right)$
$\mathbb{I}_{\mathrm{A}}$
Vista Space Model
$\dot{X}_k^{\mathrm{V}} = f_{\mathrm{V}}\left(X_k^{\mathrm{V}}, W_k^{\mathrm{V}}\right)$
$Z_k^{\mathrm{V}} = h_{\mathrm{V}}\left(X_k^{\mathrm{V}}, \nu_k^{\mathrm{V}}\right)$
$\mathbb{I}_{\mathrm{V}}$
Personal Space Model
$\dot{X}_k^{\mathrm{P}} = f_{\mathrm{P}}\left(X_k^{\mathrm{P}}, W_k^{\mathrm{P}}\right)$
$Z_k^{\mathrm{P}} = h_{\mathrm{P}}\left(X_k^{\mathrm{P}}, \nu_k^{\mathrm{P}}\right)$
$\mathbb{I}_{\mathrm{P}}$
$\mathbb{G}_{\mathrm{V}\to\mathrm{A}}$
$\dot{X}_k^{\mathrm{A}} = g_{\mathrm{V}\to\mathrm{A}}(X_k^{\mathrm{V}})$
$\mathbb{G}_{\mathrm{A}\to\mathrm{P}}$
$\dot{X}_k^{\mathrm{P}} = g_{\mathrm{A}\to\mathrm{P}}(X_k^{\mathrm{A}})$
$\mathbb{G}_{\mathrm{V}\leftarrow\mathrm{A}}$
$\dot{X}_k^{\mathrm{V}} = g_{\mathrm{V}\leftarrow\mathrm{A}}(X_k^{\mathrm{A}})$
$\mathbb{G}_{\mathrm{A}\leftarrow\mathrm{P}}$
$\dot{X}_k^{\mathrm{A}} = g_{\mathrm{A}\leftarrow\mathrm{P}}(X_k^{\mathrm{P}})$

Fig. 3. Hybrid model of the proposed tracking framework, where $\mathbb{G}$ and $\mathbb{I}$ denote the guards and invariants governing the discrete mode transitions.

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 the relative importance of close objects compared to distant ones. In characterizing this phenomenon, cognitive scientists 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 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 [13] [12].

Many natural analogies exist between human and robot/computer perception; both operate under resource constraints (cognition vs. computation), and both utilize sensor information that often degrades with distance from the observer, to name only two. Given these analogies, and the fact that computers/robots are often designed to perform human tasks, such as surveillance or navigation, a priority-based tracking framework inspired by the human perception concepts of *attention* and *focus* is proposed here, which automatically trades computational and algorithmic resources for tracking precision as a function of object relevance to the consumer of the EOT output.

Fig. 3 depicts the hybrid system model designed to implement the proposed priority-based EOT framework. Each discrete mode, Vista, Action, and Personal, represents a unique EOT approach chosen from the left, center, and right, respectively, of the complexity-precision spectrum depicted in Fig. 1. Further, probabilistic object relevancy metrics inform the mode switching strategy such that, as an object becomes increasingly *relevant* to the consumer, the tracker transitions along the path: Vista → Action → Personal, causing the overall tracking framework to transition from left to right on the spectrum depicted in Fig. 1. In this way, objects most relevant to the consumer are allocated more resources and tracked with higher precision, while objects of peripheral relevance are efficiently accounted for with inexpensive EOT methods.

The parameters of the hybrid system model of Fig. 3, i.e. the modal EOT methods and object relevancy metrics, should be chosen to reflect the specific EOT application and goals motivating the use of the proposed priority-based framework. In this way, the *optimal* parameterization of the proposed priority-based framework in Fig. 3 is highly application and consumer dependent, and therefore beyond the scope of this work. However, for demonstration purposes, an example parameterization for an autonomous driving application is provided in the coming sections, coupled with some discussion of equally valid alternatives. For this example, the *consumer* of the tracker output is an anticipatory planning routine tasked with planning control inputs to *safely progress* the vehicle toward its destination. In this vein, the consumer defines object *priority* in terms of its potential contribution to the current plan. Specifically, while all objects in the local environment are considered when planning a future path, those that have potential to violate the planner's *safety* requirement are of the highest priority, i.e. those that pose an immediate risk of collision, followed by those with potential to violate the planner's *liveness* requirement, i.e. those that inhibit the ego-vehicle's progress toward the goal location.

While beyond the scope of this paper, the hybrid system framework depicted in Fig. 3 can also be outfit to address alternative tracking goals. For instance, consider the goal of achieving tracking robustness. An EOT approach robust to occlusion could be selected when driving in a cluttered environment, e.g. [1], [15], [34], [35], [37], 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. [24], 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 (2). Further, the model can be outfit to transition according to *ability*-based (or other) metrics, rather than object relevance. For instance, general object 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.

TABLE I
Suggested modal tracking methods

Vista Space	Action Space	Personal Space
Qualitative	*Parametric Bound*	*Nonparametric Surface*
• Temporal occupancy grid [2] • Markov chain occupancy grid [32] • Topological	• Circular Disk [6] • Ellipse [5], [7], [10], [23], [24] • Rectangle [8], [26], [31] • Star Convex RHM [9]	• 2D Point Cloud [29], [30], [34]–[37] • 3D Point Cloud [18]–[20] • 3D Surface Reconstruction e.g. KinectFusion [21]

[18]–[20], 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.

4. MODAL TRACKING APPROACHES

The focus of this work is to invoke high precision EOT methods for objects that are *relevant* to the consumer, 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. 1, respectively. A partial list of existing EOT approaches appropriate for each mode is provided in Table I, and those chosen for the autonomous driving example are presented in detail in the coming sections.

Note that, while not a requirement of the priority-based EOT framework, all measurement models chosen for the autonomous driving example 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.

4.1. Vista Space Model

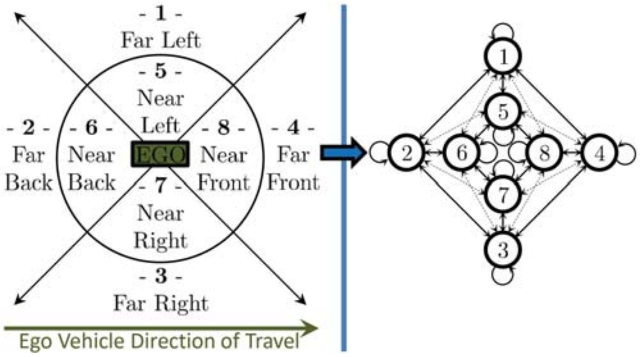

Fig. 4. Qualitative abstraction of the perceptual space of the ego robot. **Left:** Eight qualitative discrete states comprised of two range sets and four bearing quadrants. **Right:** Graphical model representation of the available transitions between the qualitative states of the discrete abstraction. Dashed arrows denote transitions enabled by the discrete time nature of the filter driven by the finite temporal resolution of the sensor.

Vista mode is reserved for objects with the most peripheral significance to the consumer; 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. 4 (left), the topology of which is encoded in the state transition diagram in Fig. 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:

$$\mathbb{B}_{X_k^{\mathrm{V}}} = \frac{\pi}{2}\cdot\left\{X_k^{\mathrm{V}} + \left[-\frac{1}{2},\quad \frac{1}{2}\right)\right\}$$

$$\mathbb{R}_{X_k^{\mathrm{V}}} = \begin{cases} [0,\quad \bar{\rho}) & \text{if } X_k^{\mathrm{V}} \in \text{Near} \\ [\bar{\rho},\quad \infty) & \text{if } X_k^{\mathrm{V}} \in \text{Far} \end{cases} \tag{6}$$

where $\bar{\rho}$ is the user-defined range boundary between ‘Near’ and ‘Far’; i.e. the circle in Fig. 4 (left).

Qualitative state representations have gained interest in robotic/computer applications, such as relational mapping [28], due to their efficiency, scalability, and natural synergy with inexpensive ordinal sensor information, such as that provided by monocular cameras or human input. The qualitative states depicted in Fig. 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, $p(X_k^{\mathrm{V}} \mid Z_{1:k})$, is distributed according to a categorical distribution.

4.1.1. Belief Prediction:

Minding the conditional independence rules defined in (4), the posterior categorical distribution over the object state at time $k-1$ is predicted forward to time

k as follows:

$$\begin{aligned} p(X_k^{\mathrm{V}} \mid Z_{1:k-1}) &= \sum_{X_{k-1}^{\mathrm{V}}} p(X_{k-1}^{\mathrm{V}}, X_k^{\mathrm{V}} \mid Z_{1:k-1}) \\ &= \sum_{X_{k-1}^{\mathrm{V}}} p(X_k^{\mathrm{V}} \mid X_{k-1}^{\mathrm{V}}) p(X_{k-1}^{\mathrm{V}} \mid Z_{1:k-1}) \end{aligned} \tag{7}$$

where $p(X_{k-1}^{\mathrm{V}} \mid Z_{1:k-1})$ is the posterior state distribution at time $k-1$, and $p(X_k^{\mathrm{V}} \mid X_{k-1}^{\mathrm{V}})$ is the symmetric discrete state transition density defined over the transition graph on the right of Fig. 4:

$$p(X_k^{\mathrm{V}} \mid X_{k-1}^{\mathrm{V}}) = \frac{\mathcal{L}(X_k^{\mathrm{V}} \mid X_{k-1}^{\mathrm{V}})}{\sum_{X_k^{\mathrm{V}}} \mathcal{L}(X_k^{\mathrm{V}} \mid X_{k-1}^{\mathrm{V}})} \tag{8}$$

where the conditional likelihood function is defined as:

$$\mathcal{L}(X_k^{\mathrm{V}} = \imath) \mid X_{k-1}^{\mathrm{V}} = \jmath) = \begin{cases} \mathcal{L}_{\mathbf{I}} & \text{if } \imath = \jmath \\ \mathcal{L}_{\bowtie} & \text{if } \imath \bowtie \jmath \\ \mathcal{L}_{\tilde{\bowtie}} & \text{if } \imath \tilde{\bowtie} \jmath \\ 0 & \text{if } \imath \not\bowtie \jmath \end{cases} \tag{9}$$

and $\bowtie$, $\not\bowtie$, and $\tilde{\bowtie}$ denote adjacency, non-adjacency, and diagonal adjacency of qualitative states (Fig. 4 left), and appear as solid, missing, and dashed edges between graph nodes in Fig. 4 (right), respectively. When applied to the graph in Fig. 4, the conditional distribution in (8) is depicted as the following symmetric, positive definite matrix:

$$p(X_k^{\mathrm{V}} \mid X_{k-1}^{\mathrm{V}}) =$$

$$X_{k-1}^{\mathrm{V}} \left\{ \overbrace{\begin{bmatrix} p_{\mathbf{I}} & p_{\bowtie} & 0 & p_{\bowtie} & p_{\bowtie} & p_{\tilde{\bowtie}} & 0 & p_{\tilde{\bowtie}} \\ p_{\bowtie} & p_{\mathbf{I}} & p_{\bowtie} & 0 & p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\tilde{\bowtie}} & 0 \\ 0 & p_{\bowtie} & p_{\mathbf{I}} & p_{\bowtie} & 0 & p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\tilde{\bowtie}} \\ p_{\bowtie} & 0 & p_{\bowtie} & p_{\mathbf{I}} & p_{\tilde{\bowtie}} & 0 & p_{\tilde{\bowtie}} & p_{\bowtie} \\ p_{\bowtie} & p_{\tilde{\bowtie}} & 0 & p_{\tilde{\bowtie}} & p_{\mathbf{I}} & p_{\bowtie} & 0 & p_{\bowtie} \\ p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\tilde{\bowtie}} & 0 & p_{\bowtie} & p_{\mathbf{I}} & p_{\bowtie} & 0 \\ 0 & p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\tilde{\bowtie}} & 0 & p_{\bowtie} & p_{\mathbf{I}} & p_{\bowtie} \\ p_{\tilde{\bowtie}} & 0 & p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\bowtie} & 0 & p_{\bowtie} & p_{\mathbf{I}} \end{bmatrix}}^{X_k^{\mathrm{V}}} \right. \tag{10}$$

where:

$$p_{(\cdot)} = \frac{\mathcal{L}_{(\cdot)}}{\mathcal{L}_{\mathbf{I}} + 3\mathcal{L}_{\bowtie} + 2\mathcal{L}_{\tilde{\bowtie}}} \tag{11}$$

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

4.1.2. Belief Update:

Minding the conditional independence rules defined in (4), the prior categorical distribution at time k, (7), is updated to reflect the observation at time k via the following equation:

$$\begin{aligned} p(X_k^{\mathrm{V}} \mid Z_{1:k}) &= \frac{p(X_k^{\mathrm{V}}, Z_k \mid Z_{1:k-1})}{p(Z_k \mid Z_{1:k-1})} \\ &= \frac{p(Z_k \mid X_k^{\mathrm{V}}) \cdot p(X_k^{\mathrm{V}} \mid Z_{1:k-1})}{\sum_{X_k^{\mathrm{V}}=1}^{8} p(Z_k \mid X_k^{\mathrm{V}}) \cdot p(X_k^{\mathrm{V}} \mid Z_{1:k-1})} \end{aligned} \tag{12}$$

where $p(X_k^{\mathrm{V}} \mid Z_{1:k-1})$ is the prior computed in (7). The conditional measurement likelihood, $p(Z_k \mid X_k^{\mathrm{V}})$, 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 (6):

$$p(Z_k \mid X_k^{\mathrm{V}}) = \sum_{\ell=1}^{n_k^z} (\beta_{z_\ell} \in \mathbb{B}_{X_k^{\mathrm{V}}}) \cap (\rho_{z_\ell} \in \mathbb{R}_{X_k^{\mathrm{V}}}) \tag{13}$$

where β_{z_ℓ} and ρ_{z_ℓ} denote the bearing and range to sensor return $z_\ell \in Z_k \; \forall \ell \in \{1, \ldots, n_k^z\}$. Note that the argument to the sum in (13) evaluates to 1, for points that lie within the discrete region of space corresponding to X_k^{V}, and 0 for those that do not; in this way, (13) counts the observations supporting qualitative state X_k^{V}.

4.2. Action Space Model

Action mode is reserved for objects of increasing significance to the consumer. For the autonomous driving example, these objects have a significant impact on the planning routine (i.e. the consumer), but are not at an immediate risk of collision [16], [17]. Therefore, a reasonable estimate of the object's position, velocity, and approximate size are desired to effectively anticipate their future behavior, and effectively plan around them. To this end, the extended object tracking approach chosen for the action space in the autonomous driving example is the random matrix method introduced in [23] and studied further in [24] and [5]. 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:

$$X_k^{\mathrm{A}} = \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{bmatrix}_k \tag{14}$$

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

$$\mathbf{E}_k = \begin{bmatrix} e_x & e_{x,y} \\ e_{y,x} & e_y \end{bmatrix}_k \tag{15}$$

The tracking problem in (5) is then to estimate the joint distribution over the object state and elliptical extent given the history of measurements, $Z_{1:k}$:

$$p(X_k^{\mathrm{A}}, \mathbf{E}_k \mid Z_{1:k}) \tag{16}$$

The joint distribution in (16) can be factored exactly into the product of vector and matrix valued distributions:

$$p(X_k^{\mathrm{A}}, \mathbf{E}_k \mid Z_{1:k}) = p(X_k^{\mathrm{A}} \mid \mathbf{E}_k, Z_{1:k}) p(\mathbf{E}_k \mid Z_{1:k}) \tag{17}$$

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

$$\begin{aligned} p(X_k^{\mathrm{A}} \mid \mathbf{E}_k, Z_{1:k}) &= \mathcal{N}(\bar{X}_{k|k}^{\mathrm{A}}, \mathbf{P}_{k|k}^{\mathrm{A}}) \\ p(\mathbf{E}_k \mid Z_{1:k}) &= \mathcal{W}^{-1}(\mathbf{\Psi}_{k|k}, \alpha_{k|k}) \end{aligned} \tag{18}$$

Thus, the posterior distribution in (16) is fully specified by the Gaussian mean, $\bar{X}_k^{\mathrm{A}}$, and covariance, $\mathbf{P}_k^{\mathrm{A}}$, coupled with the inverse Wishart scale matrix, $\mathbf{\Psi}_k$, 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, $\bar{\mathbf{E}}_k$, variance of the $(\imath, \jmath)$th element, $(\sigma_{k|k}^{\imath,\jmath})^2$, and covariance between the $(\imath, \jmath)$th and (ℓ, m)th elements, $\sigma_{k|k}^{(\imath,\jmath),(\ell,m)}$, of the extent matrix belief, $p(\mathbf{E}_k \mid Z_{1:k})$, are computed from the inverse Wishart parameters as:

$$\bar{\mathbf{E}}_{k|k} = \frac{\mathbf{\Psi}_{k|k}}{\alpha_{k|k} + d - 1} \tag{19}$$

$$\begin{aligned} (\sigma_{k|k}^{\imath,\jmath})^2 &= \frac{(\alpha_{k|k} - d + 1)(\psi_{k|k}^{\imath,\jmath})^2 + (\alpha_{k|k} - d - 1)\psi_{k|k}^{\imath,\imath}\psi_{k|k}^{\jmath,\jmath}}{(\alpha_{k|k} - d)(\alpha_{k|k} - d - 1)^2(\alpha_{k|k} - d - 3)} \\ \sigma_{k|k}^{(\imath,\jmath),(\ell,m)} &= \frac{2\psi_{k|k}^{\imath,\jmath}\psi_{k|k}^{\ell,m} + (\alpha_{k|k} - d - 1)(\psi_{k|k}^{\imath,\ell}\psi_{k|k}^{\jmath,m} + \psi_{k|k}^{\imath,m}\psi_{k|k}^{\jmath,\ell})}{(\alpha_{k|k} - d)(\alpha_{k|k} - d - 1)^2(\alpha_{k|k} - d - 3)} \end{aligned} \tag{20}$$

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

4.2.1 Belief Prediction:

The kinematic states evolve according to a stochastic continuous time differential equation model of the form in (66) provided in Appendix A. Therefore, the Kalman filter prediction equations in (69) are used to compute the mean and covariance of the prior state belief at time k:

$$p(X_k^{\mathrm{A}} \mid Z_{1:k-1}) = \mathcal{N}(\bar{X}_{k|k-1}^{\mathrm{A}}, \mathbf{P}_{k|k-1}^{\mathrm{A}}) \tag{21}$$

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

$$\begin{aligned} \mathbf{F}_k^{\mathrm{A}} &= \begin{bmatrix} \mathbf{0}_{2\times2} & \mathbf{I}_{2\times2} \\ (\mathbf{M}_{\mathrm{g},k}^{\mathrm{ego}})^T \ddot{\mathbf{M}}_{\mathrm{g},k}^{\mathrm{ego}} & -2(\mathbf{M}_{\mathrm{g},k}^{\mathrm{ego}})^T \dot{\mathbf{M}}_{\mathrm{g},k}^{\mathrm{ego}} \end{bmatrix} \\ \mathbf{G}_k^{\mathrm{A}} &= \begin{bmatrix} \mathbf{0}_{2\times2} \\ -(\mathbf{M}_{\mathrm{g},k}^{\mathrm{ego}})^T \end{bmatrix} \end{aligned} \tag{22}$$

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:

$$\begin{aligned} \mathbf{M}_{\mathrm{g},k}^{\mathrm{ego}} &= \begin{bmatrix} \cos\phi_{\mathrm{g},k}^{\mathrm{ego}} & -\sin\phi_{\mathrm{g},k}^{\mathrm{ego}} \\ \sin\phi_{\mathrm{g},k}^{\mathrm{ego}} & \cos\phi_{\mathrm{g},k}^{\mathrm{ego}} \end{bmatrix} \\ \dot{\mathbf{M}}_{\mathrm{g},k}^{\mathrm{ego}} &= \frac{\partial \mathbf{M}_{\mathrm{g},k}^{\mathrm{ego}}}{\partial \phi_{\mathrm{g},k}^{\mathrm{ego}}} \dot{\phi}_{\mathrm{g},k}^{\mathrm{ego}} \\ \ddot{\mathbf{M}}_{\mathrm{g},k}^{\mathrm{ego}} &= \frac{\partial \mathbf{M}_{\mathrm{g},k}^{\mathrm{ego}}}{\partial \phi_{\mathrm{g},k}^{\mathrm{ego}}} \ddot{\phi}_{\mathrm{g},k}^{\mathrm{ego}} - \mathbf{M}_{\mathrm{g},k}^{\mathrm{ego}} (\dot{\phi}_{\mathrm{g},k}^{\mathrm{ego}})^2 \end{aligned} \tag{23}$$

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:

$$W_k^{\mathrm{A}} \sim \mathcal{N}\left(\begin{bmatrix} \ddot{x}_{\mathrm{g},k}^{\mathrm{ego}} \\ \ddot{y}_{\mathrm{g},k}^{\mathrm{ego}} \end{bmatrix}, \mathbf{Q}_k^{\mathrm{A}} \right) \tag{24}$$

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

$$\mathbf{E}_k = \mathbf{M}_{\mathrm{g},\Delta k}^{\mathrm{ego}} \mathbf{E}_{k-1} (\mathbf{M}_{\mathrm{g},\Delta k}^{\mathrm{ego}})^T \tag{25}$$

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

$$\mathbf{M}_{\mathrm{g},\Delta k}^{\mathrm{ego}} = \begin{bmatrix} \cos(\phi_{\mathrm{g},k}^{\mathrm{ego}} - \phi_{\mathrm{g},k-1}^{\mathrm{ego}}) & -\sin(\phi_{\mathrm{g},k}^{\mathrm{ego}} - \phi_{\mathrm{g},k-1}^{\mathrm{ego}}) \\ \sin(\phi_{\mathrm{g},k}^{\mathrm{ego}} - \phi_{\mathrm{g},k-1}^{\mathrm{ego}}) & \cos(\phi_{\mathrm{g},k}^{\mathrm{ego}} - \phi_{\mathrm{g},k-1}^{\mathrm{ego}}) \end{bmatrix} \tag{26}$$

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

$$\begin{aligned} \mathbf{\Psi}_{k|k-1} &= \mathbf{M}_{\mathrm{g},\Delta k}^{\mathrm{ego}} \mathbf{\Psi}_{k-1|k-1} (\mathbf{M}_{\mathrm{g},\Delta k}^{\mathrm{ego}})^T \\ \alpha_{k|k-1} &= \exp\left(\frac{-\delta t}{\tau}\right)(\alpha_{k-1|k-1} - 2) + 2 \end{aligned} \tag{27}$$

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

4.2.2. Belief Update:

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

$$Z_k^{\mathrm{A}} = \frac{1}{n_k^z} \sum_{\ell=1}^{n_k^z} z_k^{\ell} \tag{28}$$

where $z_k^{\ell}\ \forall \ell \in \{1, \ldots, n_k^z\}$ are the individual raw sensor returns at time k. The measurement in (28) is modeled as:

$$\begin{aligned} Z_k^{\mathrm{A}} &= \mathbf{H} X_k^{\mathrm{A}} + \nu_k \\ \mathbf{H} &= [\mathbf{I}_2, \mathbf{0}_2] \end{aligned} \tag{29}$$

where the measurement noise is defined as:

$$\begin{aligned} \nu_k &\sim \mathcal{N}(0_2, \bar{\mathbf{R}}_k) \\ \bar{\mathbf{R}}_{k|k-1} &= \boldsymbol{\Psi}_{k|k-1} + \mathbf{R}_k^{\mathrm{A}} \end{aligned} \tag{30}$$

where $\mathbf{R}_k^{\mathrm{A}}$ 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 (30) is bloated by the object extent scale matrix, $\boldsymbol{\Psi}_{k|k-1}$, thus tracking precision degrades for large objects.

Given the sensor model in (29), the measurement related parameters of the joint Gaussian distribution over the object state and measurement in (71) are given as:

$$\begin{aligned} \bar{Z}_k^{\mathrm{A}} &= \mathbf{H}\bar{X}_{k|k-1}^{\mathrm{A}} \\ \mathbf{P}_{Z_k}^{\mathrm{A}} &= \mathbf{H}\mathbf{P}_{k|k-1}^{\mathrm{A}}\mathbf{H}^T + \frac{1}{n_k^z}\bar{\mathbf{R}}_{k|k-1} \\ \mathbf{P}_{X_{k|k-1}Z_k}^{\mathrm{A}} &= \mathbf{P}_{k|k-1}^{\mathrm{A}}\mathbf{H}^T \end{aligned} \tag{31}$$

and the Kalman filter equations in (73), provided in Appendix A, are used to compute the posterior distribution over the object state in (17) and (18):

$$p(X_k^{\mathrm{A}} \mid \mathbf{E}_k, Z_{1:k}) = \mathcal{N}(\bar{X}_{k|k}^{\mathrm{A}}, \mathbf{P}_{k|k}^{\mathrm{A}}) \tag{32}$$

The parameters of the inverse Wishart distribution are updated to get the posterior distribution over the object extent in (17) and (18),

$$p(\mathbf{E}_k \mid Z_{1:k}) = \mathcal{W}^{-1}(\boldsymbol{\Psi}_{k|k}, \alpha_{k|k}) \tag{33}$$

using the following equations:

$$\begin{aligned} \boldsymbol{\Psi}_{k|k} &= \frac{1}{\alpha_{k|k}}(\alpha_{k|k-1}\boldsymbol{\Psi}_{k|k-1} + \hat{\mathbf{N}}_{k|k-1} + \hat{\boldsymbol{\Sigma}}_{k|k-1}) \\ \alpha_{k|k} &= \alpha_{k|k-1} + n_k^z \end{aligned} \tag{34}$$

where:

$$\begin{aligned} \hat{\mathbf{N}}_{k|k-1} &= \Upsilon_{k|k-1}\mathbf{N}_{k|k-1}\Upsilon_{k|k-1}^T \\ \hat{\boldsymbol{\Sigma}}_{k|k-1} &= \Xi_{k|k-1}\boldsymbol{\Sigma}_k\Xi_{k|k-1}^T \end{aligned} \tag{35}$$

and:

$$\begin{aligned} \mathbf{N}_{k|k-1} &= (Z_k^{\mathrm{A}} - \bar{Z}_{k|k-1}^{\mathrm{A}})(Z_k^{\mathrm{A}} - \bar{Z}_{k|k-1}^{\mathrm{A}})^T \\ \boldsymbol{\Sigma}_k &= \sum_{\ell=1}^{n_k^z}(z_k^\ell - Z_k^{\mathrm{A}})(z_k^\ell - Z_k^{\mathrm{A}})^T \\ \Upsilon_{k|k-1} &= \boldsymbol{\Psi}_{k|k-1}^{1/2}(\mathbf{P}_{Z_k}^{\mathrm{A}})^{-1/2} \\ \Xi_{k|k-1} &= \boldsymbol{\Psi}_{k|k-1}^{1/2}\bar{\mathbf{R}}_{k|k-1}^{-1/2} \end{aligned} \tag{36}$$

Matrix square roots when computing $\Upsilon_{k|k-1}$ and $\Xi_{k|k-1}$ in (36) are computed via the Cholesky factorization.

4.3. Personal Space Model

Personal space is reserved for objects with paramount relevance to the consumer; for the autonomous driving example, these are objects deemed to be at immediate risk of collision. 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.

For the autonomous driving example, 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:

$$X_k^{\mathrm{P}} = \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \\ \phi \end{bmatrix}_k \tag{37}$$

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,

$$p(\chi_k \mid Z_{1:k}) = \mathcal{N}(Z_k, \mathbf{R}_k) \tag{38}$$

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

4.3.1. Belief Prediction:

Given that the first four object states in (37) are identical to X_k^{A} given in (14), they evolve according to the same model. The additional state, the orientation of the object, evolves according to the following scalar differential equation:

$$\dot{\phi}_k = \dot{\phi}_{\mathrm{g},k} - \dot{\phi}_{\mathrm{g},k}^{\mathrm{ego}} \tag{39}$$

Therefore, the parameters of the prior distribution at k,

$$p(X_k^{\mathrm{P}} \mid Z_{1:k-1}) = \mathcal{N}(\bar{X}_{k|k-1}^{\mathrm{P}}, \mathbf{P}_{k|k-1}^{\mathrm{P}}) \tag{40}$$

can be computed using the Kalman filter equations in (69), provided in Appendix A, using the following model:

$$\begin{aligned} \mathbf{F}_k^{\mathrm{P}} &= \begin{bmatrix} \mathbf{F}_k^{\mathrm{A}} & \mathbf{0}_{4\times 1} \\ \mathbf{0}_{1\times 4} & 0 \end{bmatrix} \\ \mathbf{G}_k^{\mathrm{P}} &= \begin{bmatrix} \mathbf{G}_k^{\mathrm{A}} & \mathbf{0}_{2\times 1} \\ \mathbf{0}_{1\times 2} & -1 \end{bmatrix} \\ W_k^{\mathrm{P}} &\sim \mathcal{N}(\bar{W}_k^{\mathrm{P}}, \mathbf{Q}_k^{\mathrm{P}}) \end{aligned} \tag{41}$$

where:

$$\begin{aligned} \bar{W}_k^{\mathrm{P}} &= \begin{bmatrix} \bar{W}_k^{\mathrm{A}} \\ \dot{\phi}_{\mathrm{g},k}^{\mathrm{ego}} \end{bmatrix} \\ \mathbf{Q}_k^{\mathrm{P}} &= \begin{bmatrix} \mathbf{Q}_k^{\mathrm{A}} & \mathbf{0}_{2\times 1} \\ \mathbf{0}_{1\times 2} & q_{\dot{\phi}_{\mathrm{g},k}} \end{bmatrix} \end{aligned} \tag{42}$$

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:

$$\chi_k = \mathbf{M}_{\Delta k}^{\text{Block}}(\chi_{k-1} - T_{k-1}^{\text{Block}}) + T_k^{\text{Block}} \tag{43}$$

where:

$$\mathbf{M}_{\Delta k}^{\text{Block}} = \mathbf{I}_{n_{k-1}^z} \otimes \begin{bmatrix} \cos(\phi_k - \phi_{k-1}) & -\sin(\phi_k - \phi_{k-1}) \\ \sin(\phi_k - \phi_{k-1}) & \cos(\phi_k - \phi_{k-1}) \end{bmatrix}$$

$$T_k^{\text{Block}} = \mathbf{1}_{n_{k-1}^z \times 1} \otimes \begin{bmatrix} x_k \\ y_k \end{bmatrix} \tag{44}$$

$\otimes$ denotes the kronecker product, and (x_k, y_k, ϕ_k) refer to the position and orientation object states in (37). The parameters of the prior distribution over the object extent at k,

$$p(\chi_k \mid Z_{1:k-1}) = \mathcal{N}(\bar{\chi}_{k|k-1}, \mathbf{P}_{\chi_{k|k-1}}) \tag{45}$$

can be computed with the Sigma Point Transform [22].

4.3.2. Belief Update:

The measurement of objects in personal space is inspired by [29], [30], and defined as an observation of the extremities of the object:

$$Z_k^{\text{P}} = \begin{bmatrix} \beta_k^{\text{cw}} \\ \beta_k^{\text{ccw}} \\ \rho_k \end{bmatrix} \tag{46}$$

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 (46) is defined as:

$$Z_k^{\text{P}} = h(\chi_k) + \nu_k \tag{47}$$

where χ_k is the extent model, and $h(\cdot)$ is a function extracting the measurement metadata in (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 (71), specifically, $\bar{Z}_k^{\text{P}}$, $\mathbf{P}_{Z_k}^{\text{P}}$, and $\mathbf{P}_{X_{k|k-1}Z_k}^{\text{P}}$, are computed from the prior distributions over the state, (40), and object extent, (45), using the Sigma Point Transform [22]. Finally, the parameters of the posterior state distribution,

$$p(X_k^{\text{P}} \mid Z_{1:k}) = \mathcal{N}(\bar{X}_{k|k}^{\text{P}}, \mathbf{P}_{k|k}^{\text{P}}) \tag{48}$$

are computed with the Kalman filter update equations in (73), and the distribution over the object extent is updated with (38); i.e. replacing the prior extent belief with the most recent lidar scan.

5. MODE TRANSITIONS

The mode transitions among, Vista, Action, and Personal spaces, depicted in Fig. 3, are fully defined by their guards, $\mathbb{G}$, and invariants, $\mathbb{I}$, informed by the object relevancy metrics, as well as the state transition functions of the form, $X_k^{\imath} = g_{\jmath\rightarrow\imath}(X_k^{\jmath})$, which transform the state belief from the source mode representation, $\jmath$, to that of the destination mode, $\imath$.

TABLE II
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

In hybrid system theory, the *invariants* are a set of conditions that must be satisfied for the system to operate within each discrete mode. This is in contrast to the *guards*, which are a set of conditions that must be satisfied to invoke each discrete mode transition. For the autonomous vehicle example, the hybrid system of Fig. 3 is deterministic; i.e. the invariants are chosen to be perfectly aligned with the guards, such that, at any given instant, there is a single valid mode of operation, and all 3 modes are reachable. The focus of this work is to invoke high precision EOT methods for objects that are *relevant* to the consumer, and inexpensive EOT methods for objects of peripheral relevance. Therefore, metrics informing the guards and invariants should be chosen to reflect a measure of the consumer's definition of object *relevance*. A list of example relevance definitions coupled with suggestions for relevancy metrics is provided in Table II. The definitions and metrics chosen for the autonomous vehicle example are presented in the following sections along with their associated guards, invariants, and mode transition functions.

5.1. Vista ↔ Action

5.1.1. Probabilistic object relevancy metric:

For the autonomous driving example, the relevancy metric informing transitions between Vista and Action modes is chosen as the probability that the object is 'Far' from the ego-vehicle, $p(\text{Far}_k \mid Z_{1:k})$; 'Far' is defined by the discrete abstraction in Fig. 4, as the object occupying any of the first four qualitative states. Therefore, the far probability metric is computed as:

$$\begin{aligned} p(\text{Far}_k \mid Z_{1:k}) &= p\left(\bigcup_{X_k^{\text{V}}=1}^{4} X_k^{\text{V}} \mid Z_{1:k}\right) \\ &= \sum_{X_k^{\text{V}}=1}^{4} p(X_k^{\text{V}} \mid Z_{1:k}) \end{aligned} \tag{49}$$

where the equality of the first and second lines of (49) is conditioned on the fact that the qualitative states are disjoint. For the transition from Vista to Action, V→A, $p(X_k^{\text{V}} \mid Z_{1:k})$ is the current state belief posterior computed during the measurement update step. For the reverse transition, V←A, $p(X_k^{\text{V}} \mid Z_{1:k})$ is computed from

(12) with a uniform prior over the qualitative states, X_k^{V}. Specifically,

$$p(X_k^{\mathrm{V}} \mid Z_{1:k}) = \frac{p(Z_k \mid X_k^{\mathrm{V}})}{\sum_{X_k^{\mathrm{V}}} p(Z_k \mid X_k^{\mathrm{V}})} \tag{50}$$

where $p(Z_k \mid X_k^{\mathrm{V}})$ is the vista mode measurement likelihood defined in (13).

5.1.2. Guards:

The guards are defined by thresholding the probabilistic object relevancy metric defined above. Specifically,

$$\begin{aligned} \mathbb{G}_{\mathrm{V}\rightarrow\mathrm{A}} &\triangleq p(\mathrm{Far}_k \mid Z_{1:k}) \leq p_{\mathrm{V,A}} \\ \mathbb{G}_{\mathrm{V}\leftarrow\mathrm{A}} &\triangleq p(\mathrm{Far}_k \mid Z_{1:k}) > p_{\mathrm{A,V}} \end{aligned} \tag{51}$$

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

5.1.3. 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:

$$\begin{aligned} \boldsymbol{\Psi}_k &= \frac{1}{n_k^z}\boldsymbol{\Sigma}_k \\ \alpha_k &= n_k^z \end{aligned} \tag{52}$$

where $\boldsymbol{\Sigma}_k$ is defined in (36), and n_k^z is the number of measurement returns at time k. The Gaussian parameters of the state distribution are computed as:

$$\begin{aligned} \bar{X}_k^{\mathrm{A}} &= \begin{bmatrix} Z_k^{\mathrm{A}} \\ -\dot{x}_{\mathrm{g},k}^{\mathrm{ego}} \\ -\dot{y}_{\mathrm{g},k}^{\mathrm{ego}} \end{bmatrix} \\ \mathbf{P}_k^{\mathrm{A}} &= \begin{bmatrix} \frac{1}{n_k^z}\bar{\mathbf{R}}_k & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{P}_{V_0}^{\mathrm{A}} \end{bmatrix} \end{aligned} \tag{53}$$

where Z_k^{A} and $\bar{\mathbf{R}}_k$ are the mean and covariance of the Action space centroid measurement, defined in (28) and (30), respectively, and $\mathbf{P}_{V_0}^{\mathrm{A}}$ 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 (22) and (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 (50), which was already computed to evaluate the guard, $\mathbb{G}_{\mathrm{V}\leftarrow\mathrm{A}}$, in (51).

5.2. Action ↔ Personal

5.2.1. Probabilistic object relevancy metric:

For the autonomous driving example, 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, $p(\mathcal{C}_{k:k+h} \mid Z_{1:k})$. The anticipated probability of collision, $p(\mathcal{C}_{k:k+h} \mid Z_{1:k})$, is taken as the maximum *instantaneous* collision probability over each time step in the horizon, h:

$$p(\mathcal{C}_{k:k+h} \mid Z_{1:k}) = \max_{\ell\in\{1,\ldots,h\}} [p(\mathcal{C}_{k+\ell} \mid Z_{1:k})] \tag{54}$$

Conceptually, the instantaneous collision probabilities, $p(\mathcal{C}_{k+\ell} \mid Z_{1:k})\ \forall \ell \in \{1,\ldots,h\}$, 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. 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:

$$p(\mathcal{C}_{k+\ell} \mid Z_{1:k}) = \exp(-\tfrac{1}{2}(\bar{D}^{\min}_{\mathcal{O}_{k+\ell|k}})^T \mathbf{P}^{-1}_{D^{\min}_{\mathcal{O}_{k+\ell|k}}} \bar{D}^{\min}_{\mathcal{O}_{k+\ell|k}}) \tag{55}$$

where:

$$D^{\min}_{\mathcal{O}_{k+\ell|k}} \sim \mathcal{N}(\bar{D}^{\min}_{\mathcal{O}_{k+\ell|k}}, \mathbf{P}_{D^{\min}_{\mathcal{O}_{k+\ell|k}}}) \tag{56}$$

is the vector from the ego vehicle to the closest point in the collision region, $\mathcal{O}_{k+\ell}$, the mean and covariance of which can be computed using the sigma point transform [22].

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 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. [16], [17].

5.2.2. Guards:

The guards are defined by thresholding the probabilistic object relevancy metric defined above. Specifically,

$$\begin{aligned} \mathbb{G}_{\mathrm{A}\rightarrow\mathrm{P}} &\triangleq p(\mathcal{C}_{k:k+h} \mid Z_{1:k}) \geq p_{\mathrm{A,P}} \\ \mathbb{G}_{\mathrm{A}\leftarrow\mathrm{P}} &\triangleq p(\mathcal{C}_{k:k+h} \mid Z_{1:k}) < p_{\mathrm{P,A}} \end{aligned} \tag{57}$$

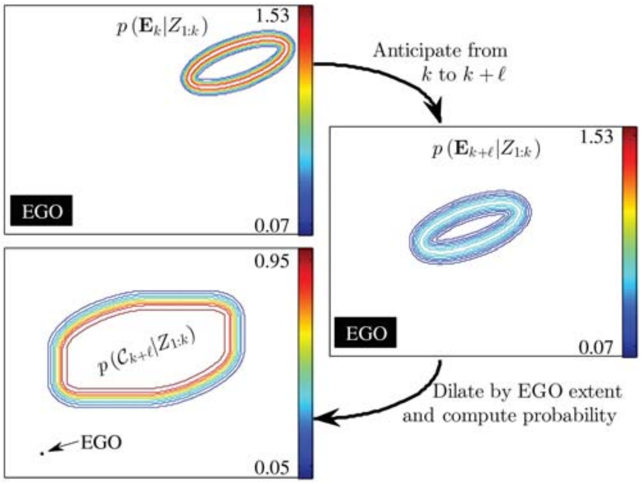
$p(\mathbf{E}_k|Z_{1:k})$
1.53
0.07
EGO
Anticipate from k to $k+\ell$
$p(\mathbf{E}_{k+\ell}|Z_{1:k})$
1.53
0.07
EGO
0.95
$p(\mathcal{C}_{k+\ell}|Z_{1:k})$
0.05
EGO
Dilate by EGO extent and compute probability

Fig. 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.

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

5.2.3. Transition functions:

Given the similarity of the state representations in (14) and (37), the state transition functions for A→P is:

$$\bar{X}_k^{\text{P}} = \begin{bmatrix} \bar{X}_k^{\text{A}} \\ 0 \end{bmatrix}$$

$$\mathbf{P}_k^{\text{P}} = \begin{bmatrix} \mathbf{P}_k^{\text{A}} & 0_{4\times 1} \\ 0_{1\times 4} & \epsilon \end{bmatrix} \tag{58}$$

where ϵ is a small positive number indicating perfect knowledge of the initial relative orientation, ϕ_k, while maintaining the positive definite requirement of $\mathbf{P}_k^{\text{P}}$; 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, $p(\chi_k \mid Z_{1:k})$, is initialized from the current lidar scan returned from the object via (38).

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

$$\bar{X}_k^{\text{A}} = \bar{X}_k^{\text{P}}(1:4)$$

$$\mathbf{P}_k^{\text{A}} = \mathbf{P}_k^{\text{P}}(1:4, 1:4) \tag{59}$$

and the inverse Wishart parameters of the distribution over the object extent are initialized from the lidar scan as in (52).

5.3. Invariants

For the autonomous driving example, the invariants are chosen such that as a guard enables a transition between two modes, the invariant for the source mode is violated, and the invariant for the destination mode is satisfied. Specifically:

$$\mathbb{I}_{\text{V}} \triangleq p(\text{Far}_k \mid Z_{1:k}) > p_{\text{A,V}}$$

$$\mathbb{I}_{\text{A}} \triangleq (p(\text{Far}_k \mid Z_{1:k}) \leq p_{\text{V,A}}) \cap \ldots$$
$$(p(\mathcal{C}_{k:k+h} \mid Z_{1:k}) < p_{\text{A,P}})$$

$$\mathbb{I}_{\text{P}} \triangleq p_{\text{P,A}} \leq p(\mathcal{C}_{k:k+h} \mid Z_{1:k}) \tag{60}$$

In this way, deterministic transitions are triggered as soon the guard is satisfied.

6. SIMULATION RESULTS

To demonstrate the ability of the proposed priority-based framework in Fig. 3 to automatically trade computation for tracking precision as a function of object relevance, the framework, as parameterized in Sections 4 and 5 for the autonomous driving example, is evaluated over the two simulated scenarios depicted in Figs. 6 and 7. The scenario depicted in Fig. 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. 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 (≈ 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 $\mathcal{N}(0, 1\ \text{cm}^2)$ 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 III. All simulations were coded in Matlab with all 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*, $\gamma_k^{n-\sigma}$, is defined as:

$$\gamma_k^{n-\sigma} = \bar{\mathcal{A}}^{-1}(p(X_k \mid Z_{1:k}), n) \tag{61}$$

where $\bar{\mathcal{A}}(p(X_k \mid Z_{1:k}), n)$ denotes the expected value of the area enclosed by the $n-\sigma$ confidence bound. To obtain a fair comparison with objects in Vista space,

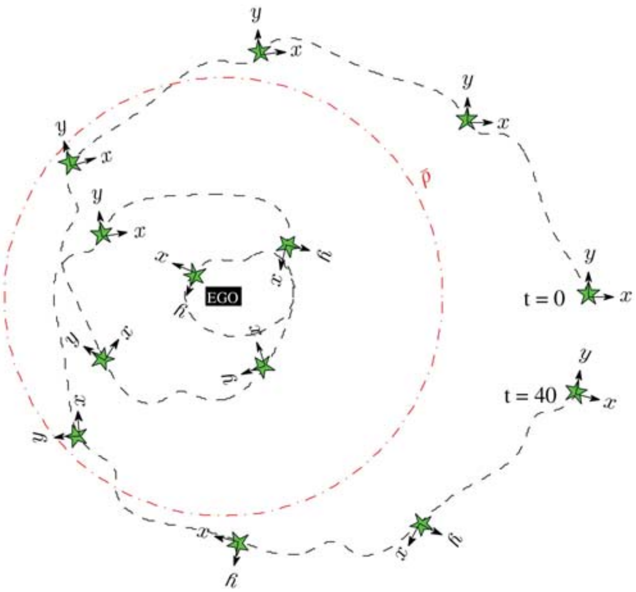

Fig. 6. Simulated scenario of a star shaped object spiraling in toward the ego-vehicle and then back out over a period of 40 s; 13 time steps are shown.

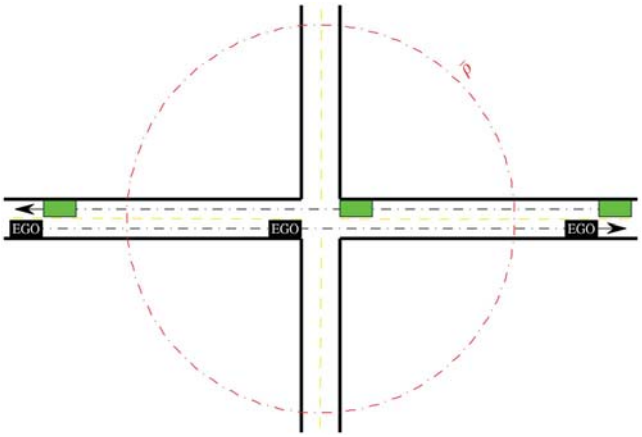

Fig. 7. Simulated scenario of a rectangular object passing the ego vehicle head-on in close proximity at an intersection; three time steps are shown.

TABLE III
Hybrid tracking parameters

Mode Transitions		Vista		Action		Personal	
$p_{\mathrm{V,A}}$	0.45	$\bar{\rho}$	30 m	τ	60 s		
$p_{\mathrm{A,V}}$	0.55	$p_{\mathbf{I}}$	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						

the object position belief is taken to be uniformly distributed over the qualitative region represented by the state, and the $n-\sigma$ confidence bound is interpreted as the area required to enclose the same probability, $p(n)$, as the $n-\sigma$ confidence bound of a Gaussian, where $n-\sigma$ refers to the Mahalanobis distance from the mean of the distribution:

$$p(n) = 1 - \exp\left(-\frac{n^2}{2}\right) \tag{62}$$

Specifically, for the discrete abstraction in Fig. 4:

$$\bar{\mathcal{A}}(p(X_k \mid Z_{1:k}), n) = \begin{cases} p(n)\cdot\pi\cdot\dfrac{\mathbb{E}[\rho_{\mathrm{o}}^2 - \rho_{\mathrm{i}}^2]}{4} & \text{if Vista} \\ n^2\cdot\pi\cdot\sqrt{|\mathbf{P}_{k|k}^{xy}|} & \text{otherwise} \end{cases} \tag{63}$$

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 (63) becomes:

$$\begin{aligned}\mathbb{E}[\rho_{\mathrm{o}}^2 - \rho_{\mathrm{i}}^2] = p(X_k \in \mathrm{Far} \mid Z_{1:k})\cdot\rho_{\max}^2 \cdots \\ - [2\cdot p(X_k \in \mathrm{Far} \mid Z_{1:k}) - 1]\cdot\bar{\rho}^2\end{aligned} \tag{64}$$

where $\rho_{\max}$ denotes the maximum range of the sensor, $p(\mathrm{Far} \mid Z_{1:k})$ is defined in (49). Note that the expression in (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. 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.

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

$$\epsilon_k = \delta t_{\mathrm{computation}}\cdot f_{\mathrm{sensor}} \tag{65}$$

Figs. 8 and 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 (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 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 (30), which scales directly with the size of the object extent, and is ultimately the source of the degraded tracking precision in Action space.

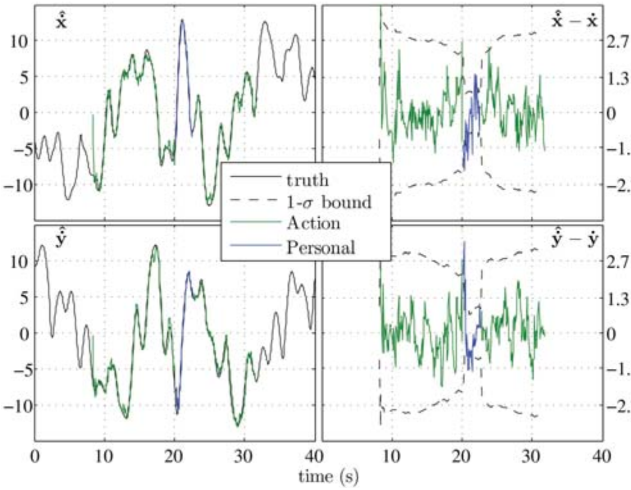

Fig. 8. Performance in tracking the velocity states for the star scenario: **Left:** MAP estimates overlaid on ground truth, **Right:** tracking error and $1-\sigma$ bounds. Note that there is not a concept of velocity in Vista space.

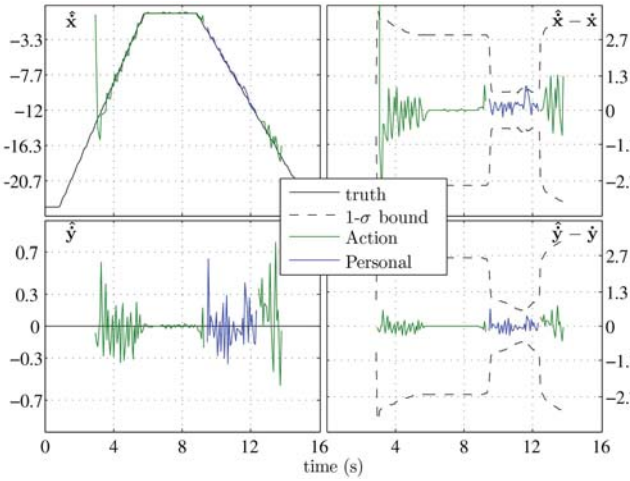

Fig. 9. Performance in tracking the velocity states for the intersection scenario: **Left:** MAP estimates overlaid on ground truth, **Right:** tracking error and $1-\sigma$ bounds. Note that there is not a concept of velocity in Vista space.

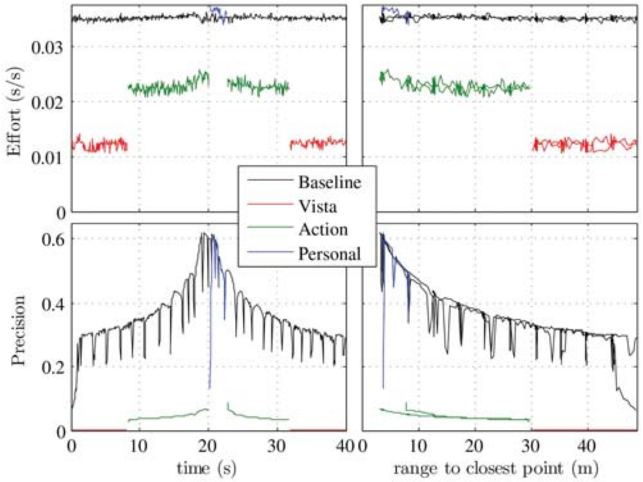

Fig. 10. Computational effort, ϵ, and precision, γ, as a function of time (left) and range to the closest point on the object (right) for the star scenario.

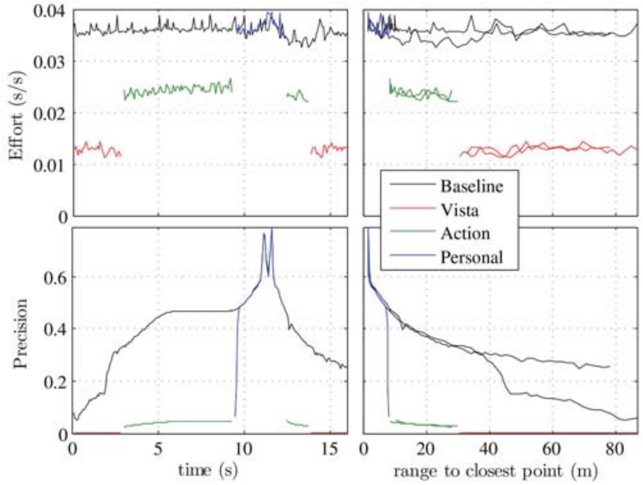

Fig. 11. Computational effort, ϵ, and precision, γ, as a function of time (left) and range to the closest point on the object (right) for the intersection scenario.

TABLE IV
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%

Also apparent in Figs. 8 and 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 (53) within two time steps (0.16 s) of the filter transition 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. 10 and 11 plot the precision, ϵ defined in (61), and computational effort, γ defined in (65), over time, for both the star and intersection scenarios, respectively; the combined summary of these metrics is provided in Table IV. *Baseline* refers to a tracker that operates solely in Personal mode to emphasize the contribution of the hybrid framework depicted in Fig. 3. The Personal 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

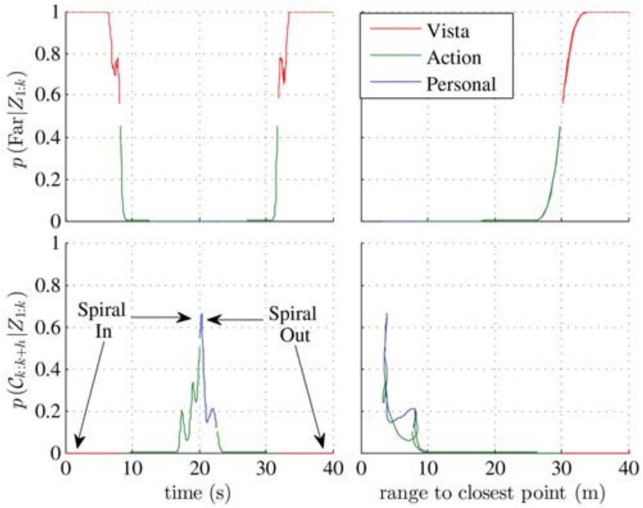

Fig. 12. Probabilistic object relevancy metric trajectories for the star scenario. Annotations referring to events in the ground truth scenario are provided for perspective. **Top Row:** Probability that the object is 'Far' away, governing the transitions between Vista and Action modes. **Bottom Row:** 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.

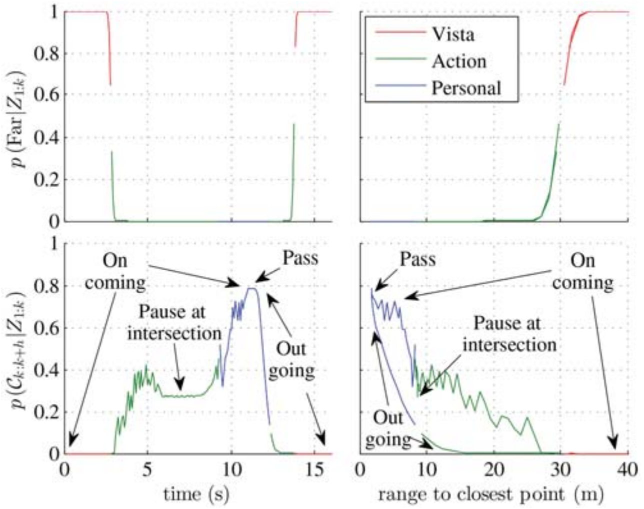

Fig. 13. Probabilistic object relevancy metric trajectories for the intersection scenario. Annotations referring to events in the ground truth scenario are provided for perspective. **Top Row:** Probability that the object is 'Far' away, governing the transitions between Vista and Action modes. **Bottom 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.

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. 10 is a direct consequence of the measurement model in (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. 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 viewpoint is changing most rapidly.

Lastly, Figs. 12 and 13 demonstrate the trajectories of the probabilistic object relevancy metrics for both scenarios. The flat region in the bottom left of Fig. 13 in the approximate range, $5.5\text{ s} < t < 8.5$ 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. 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. 13) than it is after they depart the intersection in opposing directions (portion of the curve labeled 'Out going' in Fig. 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. 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. 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. 12)—a direct consequence of the hybrid mode transition to Personal space.

7. CONCLUSION

Inspired by human perception, this paper introduces a novel method to dynamically allocate algorithmic and computational resources to achieve variable precision tracking of extended objects. 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 consumer of the tracker output, a priority-based tracking framework is proposed, enabling objects of critical importance to the consumer to be tracked with relatively expensive, high precision methods, and objects of peripheral importance to be tracked with relatively efficient, low precision methods. The proposed priority-based framework is a direct analog to the human perception concepts of *attention* and *focus*.

The priority-based EOT framework is parameterized for an example autonomous vehicle application in which the consumer of the tracking output is an anticipatory planner. Probabilistic object relevancy metrics are derived to convey the priority of an object to the consumer, and inform mode transitions in the hybrid model implementation of the priority-based EOT framework. Simulation results for two different scenarios are presented and 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 consumer (i.e. the anticipatory planner).

APPENDIX A KALMAN FILTER

This section provides the Kalman filter prediction and update equations [4], [11], [33].

A.1. Prediction

Given a stochastic linear vector differential equation model the form:

$$\begin{aligned}\dot{X}_k &= \mathbf{F}_k X_k + \mathbf{G}_k W_k\\ W_k &\sim \mathcal{N}(\bar{W}_k, \mathbf{Q}_k)\end{aligned} \tag{66}$$

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

$$p(X_{k-1} \mid Z_{1:k-1}) = \mathcal{N}(\bar{X}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}) \tag{67}$$

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

$$p(X_k \mid Z_{1:k-1}) = \mathcal{N}(\bar{X}_{k|k-1}, \mathbf{P}_{k|k-1}) \tag{68}$$

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

$$\begin{aligned}\bar{X}_{k|k-1} &= \int_{k-1}^{k} \dot{X}_t dt\\ \mathbf{P}_{k|k-1} &= \mathbf{\Phi}_{k-1}\mathbf{P}_{k-1|k-1}\mathbf{\Phi}_{k-1}^T + \mathbf{\Gamma}_{k-1}(\mathbf{Q}_k \cdot \delta t)\mathbf{\Gamma}_{k-1}^T\end{aligned} \tag{69}$$

where:

$$\mathbf{\Phi}_{k-1} = \int_{k-1}^{k} \mathbf{F}_t dt, \quad \text{and} \quad \mathbf{\Gamma}_{k-1} = \int_{k-1}^{k} \mathbf{G}_t dt \tag{70}$$

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

A.2. Update

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

$$p(X_k, Z_k \mid Z_{1:k-1}) = \mathcal{N}\left(\begin{bmatrix}\bar{X}_{k|k-1}\\ \bar{Z}_k\end{bmatrix}, \begin{bmatrix}\mathbf{P}_{k|k-1}, & \mathbf{P}_{X_{k|k-1}Z_k}\\ \mathbf{P}_{X_{k|k-1}Z_k}^T, & \mathbf{P}_{Z_k}\end{bmatrix}\right) \tag{71}$$

where $\bar{X}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$ are the mean and covariance of the prior state distribution computed in (69), and $\bar{Z}_k$, $\mathbf{P}_{Z_k}$, and $\mathbf{P}_{X_{k|k-1}Z_k}$ 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:

$$p(X_k \mid Z_{1:k}) = \mathcal{N}(\bar{X}_{k|k}, \mathbf{P}_{k|k}) \tag{72}$$

with parameters computed as:

$$\begin{aligned}\bar{X}_{k|k} &= \bar{X}_{k|k-1} + \mathbf{K}_k(Z_k - \bar{Z}_k)\\ \mathbf{P}_{k|k} &= \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{P}_{Z_k}\mathbf{K}_k^T\end{aligned} \tag{73}$$

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

$$\mathbf{K}_k = \mathbf{P}_{X_{k|k-1}Z_k}\mathbf{P}_{Z_k}^{-1} \tag{74}$$

ACKNOWLEDGMENT

This work was supported by the NSF Robust Intelligence grant IIS-1320490 and ARO grant W911NF-09-1-0466.

REFERENCES

[1] A. Andriynko, S. Roth, and K. 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.

[2] D. Arbuckle, A. Howard, and M. 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)*, October 2004, pp. 409–414.

[3] J. Aue, M. R. Schmid, T. Graf, J. Effertz, and P. 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.

[4] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan
Estimation with Applications to Tracking and Navigation.
John Wiley & Sons, Inc., 2001.

[5] M. Baum, M. Feldmann, D. Franken, U. D. Hanebeck, and W. 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.

[6] M. Baum and U. D. Hanebeck
Extended object tracking based on combined set-theoretic and stochastic fusion,
in *Proceedings of the 12th International Conference on Information Fusion (FUSION)*, July 2009, pp. 1288–1295.

[7] M. Baum and U. 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.

[8] M. Baum and U. D. Hanebeck
Tracking an extended object modeled as an axis-aligned rectangle,
in *Proceedings of Informatik 2009*, 2009.

[9] M. Baum and U. 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.

[10] M. Baum, B. Noack, and U. 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.

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

[12] J. E. Cutting
Reconceiving perceptual space,
in *Looking Into Pictures: An Interdisciplinary Approach to Pictorial Space*. MIT Press, June 2003, ch. 11.

[13] J. E. Cutting and P. 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*. Academic Press, 1995, pp. 69–117.

[14] K. Gilholm and D. Salmond
Spatial distribution model for tracking extended objects,
in *IEE Proceedings—Radar, Sonar, and Navigation*, vol. 152, no. 5, October 2005, pp. 364–371.

[15] K. Granstrom and O. Orguner
A phd filter for tracking multiple extended targets using random matrices,
IEEE Transactions on Signal Processing, vol. 60, no. 11, pp. 5657–5671, 2012.

[16] J. Hardy and M. Campbell
Contingency planning over probabilistic obstacle predictions for autonomous road vehicles,
IEEE Transactions on Robotics, vol. 29, no. 4, pp. 913–929, 2013.

[17] F. Havlak and M. Campbell
Discrete and conitnuous, probabilistic anticipation for autonomous robots in urban environments,
Transactions on Robotics, vol. 30, no. 2, pp. 461–474, Decemberl 2013.

[18] D. Held, J. Levinson, and S. 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.

[19] D. Held, J. Levinson, S. Thrun, and S. 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.

[20] E. Ilg, R. Kummerle, W. Burgard, and T. 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.

[21] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon
Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera,
in *ACM Symposium on User Interface Software and Technology*, October 2011. [Online]. Available: http://research.microsoft.com/apps/pubs/default.aspx?id=155416.

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

[23] W. Koch
Bayesian approach to extended object and cluster tracking using random matrices,
IEEE Transactions on Aerospace and Electronic Systems, vol. 44, no. 3, pp. 1042–1059, 2008.

[24] W. Koch and M. Feldmann
Cluster tracking under kinematical constraints using random matrices,
Robotics and Autonomous Systems, vol. 57, no. 3, pp. 296–309, 2009.

[25] M. S. Landy, L. T. Maloney, E. Johnston, and M. Young
Measurement and modeling of depth cue combination: in defense of weak fusion,
Vision Research, vol. 35, no. 3, pp. 389–412, February 1995.

[26] C. Lundquist, K. Granstrom, and O. Orguner
Estimating object shape of targets with a phd filter,
in *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, July 2011, pp. 1–8.

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

[28] M. McClelland, T. Estlin, and M. Campbell
Qualitative relational mapping for planetary rover exploration,
in *Proceedings of the AIAA Guidance, Navigation and Control Conference*, 2013.

[29] I. Miller, M. Campbell, et al.
Team cornell's skynet: Robust perception and planning in an urban environment,
Journal of Field Robotics, vol. 25, no. 8, pp. 493–527, 2008.

[30] I. Miller, M. Campbell, and D. Huttenlocher
Efficient, unbiased tracking of multiple dynamic obstacles under large viewpoint changes,
IEEE Transactions on Robotics, vol. 27, no. 1, pp. 29–46, 2011.

[31] A. Petrovskaya and S. Thrun
Model based vehicle detection and tracking for autonomous urban driving,
Autonomous Robots Journal, vol. 26, no. 2–3, pp. 123–139, 2009.

[32] J. Saarinen, H. Andreasson, and A. 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)*, October 2012, pp. 3489–3495.

[33] S. Thrun, W. Burgard, and D. Fox
Probabilistic Robotics.
MIT Press, 2006.

[34] K. Wyffels and M. 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.

[35] K. Wyffels and M. Campbell
Negative observations for multiple hypothesis tracking of dynamic extended objects,
in *Proceedings of the American Controls Conference (ACC 2014)*, June 2014.

[36] K. Wyffels and M. Campbell
Joint tracking and non-parametric shape estimation of arbitrary extended objects,
in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 1–8.

[37] K. Wyffels and M. Campbell
Negative information for occlusion reasoning in dynamic extended multiobject tracking,
IEEE Transactions on Robotics, vol. 31, no. 2, pp. 425–442, April 2015.

Kevin Wyffels received his B.S. degree from the University at Buffalo in 2005, M.S. degree from Rochester Institute of Technology in 2007, and M.S./Ph.D. degrees from Cornell University in 2014/2016. He is currently a research scientist at Ford Motor Company specializing in estimation, probabilistic inference, and robotic perception for fully autonomous vehicles. He is a member of IEEE.

Mark Campbell received his B.S. degree from CMU in Mechanical Engineering, and his M.S./Ph.D. degrees in Control and Estimation from MIT in 1993/1996. He is currently a Professor and the S. C. Thomas Sze Director of the Sibley School of Mechanical and Aerospace Engineering at Cornell University. In 2005–06, he was Visiting Scientist at the Insitu group, and an ARC International Fellow at the Australian Centre of Field Robotics. He received best paper awards from AIAA Propulsion and GNC conferences, and Frontiers in Education conference, and teaching awards from Cornell, University of Washington, and ASEE. He has been an invited speaker for the NAE Frontiers in Engineering Symposium and NAS Kavli Frontiers of Science Symposium. His research interests are in the areas of autonomous systems.