

Optimized Self-Localization for SLAM in Dynamic Scenes Using Probability Hypothesis Density Filters

Christine Evers¹, Senior Member, IEEE, and Patrick A. Naylor¹, Senior Member, IEEE

Abstract—In many applications, sensors that map the positions of objects in unknown environments are installed on dynamic platforms. As measurements are relative to the observer's sensors, scene mapping requires accurate knowledge of the observer state. However, in practice, observer reports are subject to positioning errors. Simultaneous localization and mapping addresses the joint estimation problem of observer localization and scene mapping. State-of-the-art approaches typically use visual or optical sensors and therefore rely on static beacons in the environment to anchor the observer estimate. However, many applications involving sensors that are not conventionally used for Simultaneous Localization and Mapping (SLAM) are affected by highly dynamic scenes, such that the static world assumption is invalid. This paper proposes a novel approach for dynamic scenes, called Generalized Motion (GEM) SLAM. Based on probability hypothesis density filters, the proposed approach probabilistically anchors the observer state by fusing observer information inferred from the scene with reports of the observer motion. This paper derives the general, theoretical framework for GEM-SLAM, and shows that it generalizes existing Probability Hypothesis Density (PHD)-based SLAM algorithms. Simulations for a model-specific realization using range-bearing sensors and multiple moving objects highlight that GEM-SLAM achieves significant improvements over three benchmark algorithms.

Index Terms—Simultaneous localization and mapping, Bayes methods, nonlinear dynamical systems.

I. INTRODUCTION

SCENE mapping equips dynamic observers with the ability to explore and create three-dimensional representations of unknown environments, hence impacting on autonomous as well as guided systems, including robots, unmanned aerial vehicles, ad-hoc mobile networks, and planetary rovers. In practice, the detection of objects from the observer's sensor measurements is subject to uncertainty, whilst multipath effects in, e.g., communications, SONAR, and acoustic sensors, lead to false alarms. Moreover, dynamic objects result in spatio-temporal variations. Tracking algorithms [1]–[3] can be used to

exploit temporal models of the object dynamics and to estimate their smoothed trajectories from the distorted detections. Since object detections are relative to the observer's positional state, accurate knowledge of the observer trajectory in the environment is required in order to track objects across time and space. Reports of the observer position and motion can be obtained from, e.g., Global Positioning System (GPS) or gyroscopic sensors. However, in practice, the observer reports are subject to errors due to physical and mechanical limitations. For example, inertial sensor reports of robots decrease with hardware age due to the wear of the mechanical parts [4]. Dead reckoning [5], propagating the initial position using the reports, therefore leads to divergence between the estimated and ground truth observer trajectories.

Observer localization can be improved by identifying the observer state that optimally aligns the sensor measurements with the estimated objects' positions. Conditional on the estimated observer position, the objects are mapped using tracking algorithms. Observer localization and object mapping are therefore jointly dependent and present the estimation problem of SLAM.

SLAM developments to date led from the basic formulation of the simultaneous estimation problem [6], [7], through convergence and consistency studies [8], to the current focus on robust perception [9]. An overview of the open problems and future directions of SLAM research is provided in [10].

The aim and objective of this paper focus on two of the challenges highlighted in [10]. The aim of this paper is to provide a theoretical framework for SLAM that is suitable for sensors that facilitate perception beyond vision but are not yet conventionally used for SLAM, such as acoustic microphone arrays [11], [12]. Since many non-conventional sensors are deployed in environments where the positions of the observer and the objects are highly time-varying, e.g., underwater, the objective of this paper is the development of an approach that is specifically designed for dynamic environments.

SLAM was originally developed for the use of LIDAR range finders. Research rapidly focused on visual sensors as monocular, stereoscopic, and RGB-D cameras became available as commodity hardware [13]–[15]. Visual SLAM can be broadly classified into approaches that use either local or global descriptors. Local descriptors [16], [17] involve detection algorithms that abstract objects into salient features. Global descriptors [18] process the image without a detection phase, and are hence particularly suitable for place recognition.

However, global descriptors are often difficult to acquire for non-visual sensors. For example, acoustic signals correspond to the superposition of multiple interfering source signals, each of which is the result of a convolution with the acoustic transmission channel. Processing of audio content without a detection phase is therefore difficult. As such, feature-based SLAM

Manuscript received March 24, 2017; revised August 21, 2017 and October 30, 2017; accepted November 1, 2017. Date of publication November 21, 2017; date of current version January 16, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mats Bengtsson. This work was supported in part by the U.K. Engineering and Physical Sciences Research Council Fellowship Grant Reference EP/P001017/1 and in part by the European Union's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement 609465. (Corresponding author: Christine Evers.)

The authors are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: c.evers@imperial.ac.uk; p.naylor@imperial.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2017.2775590

is often a more appropriate choice over global descriptors for modalities other than visual sensors.

The majority of feature-based SLAM approaches are graph-based [19]–[21]. Graph-based SLAM typically relies on non-convex or non-linear optimization techniques. Hence, convergence to local minima may result in estimates unsuitable for navigation [10]. Moreover, pose graph optimization is susceptible to false feature detections. For improved robustness, loop closure [22] is typically used to recalibrate the observer trajectory when place recognition detects a previously visited scene. Fundamentally, loop closure relies on the assumption that each scene must contain immovable objects, of which at least some are visible each time the scene is revisited. However, this static-world assumption is invalid in dynamic environments. For example, underwater scenarios are subject to severe changes in both the positions of objects, such as vessels, as well as the seabed, i.e., the transmission channel for passive or active SONAR. In fully dynamic environments, loop closure is therefore not applicable, causing significant limitations for large-scale autonomy. Even on short scales, SLAM in dynamic scenes to date remains challenging [10].

For dynamic scenes, probabilistic filtering is a natural alternative to graph-based approaches. Perhaps the most widely used probabilistic feature-based approach is Factored Solution To Simultaneous Localization and Mapping (FastSLAM) [23]–[25]. A particle filter [26] samples multiple candidates of the observer state from the prior dynamical model of the observer motion. For each observer candidate, or particle, an ensemble of Extended Kalman Filters (EKFs) is used to estimate the feature positions, one at a time. Data association is used to establish a one-to-one mapping between the detections and features. However, in realistic conditions affected by false alarms and missing detections, robust data association is one of the fundamental challenges of SLAM algorithms [27].

For robust estimation against false alarms and missing detections, the Rao-Blackwellized PHD (RB-PHD) filter [28], [29], formulates the SLAM problem using Random Finite Sets (RFSs) to avoid the need for heuristic data association altogether. For each observer particle, one realization of a PHD filter [30], [31] is used to track the set of features. The most likely observer state is identified based on the Probability Density Function (pdf) of the set of detections conditional on the feature estimates. However, the expression of this pdf as proposed in [28], [29] is of combinatorial nature and hence computationally intractable. Therefore, the RB-PHD filter approximates the multi-detection pdf by considering the contribution of a single feature only. Alternatively, [32] approximates the intractable likelihood by a randomized computation of the equivalent matrix permanent. The work in [33] avoids approximating the multi-detection pdf by extending [28] to a Labeled Multi-Bernoulli (LMB) filter [34]. However, as reported in [34], the feature update for each observer particle and time step can require several seconds of computational time. Hence the LMB filter remains computationally prohibitive for many large-scale SLAM problems.

Rather than approximating the multi-detection pdf, the Single Cluster PHD (SC-PHD) filter [35]–[38] uses the closed-form solution. The SLAM problem is considered as the estimation of a cluster process of features, where the cluster centre corresponds to the observer state. Thus, the observer state is estimated as the centroid that aligns all feature detections with the feature estimates in the map. However, as the SC-PHD filter relies only

on feature information for observer localization, applications where features are highly dynamic and subject to uncertainty can lead to short-term ambiguities and hence long-term divergence of the observer estimates.

For robust SLAM performance in dynamic and uncertain environments, this paper proposes a novel approach, called GEM-SLAM, that fuses the knowledge inferred from feature mapping based on the SC-PHD filter, with reports of the observer motion, thereby generalizing existing PHD-based SLAM approaches. Such observer reports are typically obtained from inertial sensors, but may also include path planning instructions, motor outputs, or GPS coordinates. By considering the reports as measurements of the observer position, GEM-SLAM exploits uncertainties in both the observer dynamics and report accuracy to propagate the posterior observer pdf in time. Observer localization is optimized by intersecting the space of likely observer positions corresponding to the posterior pdf with the space of observer positions corresponding to the closed-form multi-detection likelihood of the features.

The novel contributions of this paper are therefore as follows. 1) We derive the theoretical framework and show that GEM-SLAM fuses observer information from features with the observer reports. Furthermore, illustrative examples will demonstrate that GEM-SLAM reduces ambiguities in observer localization arising from dynamic scenes, leading to improved accuracy in the observer and feature estimates. 2) The proposed framework is derived as a general, model-independent formulation. GEM-SLAM is therefore sensor-agnostic and can be realized for different sensor-specific models, including sensors that are not conventionally used for SLAM. For illustrative purposes and performance evaluation, this paper presents in addition to the general formulation a model-specific realization for range-bearing sensors in dynamic scenes. 3) A detailed discussion of the relationship between GEM-SLAM, the RB-PHD and SC-PHD filter as well as FastSLAM shows that GEM-SLAM generalizes existing PHD approaches, expressing the RB-PHD and SC-PHD filter as special cases. Using the model-specific realization, the estimation accuracy of GEM-SLAM is compared against the three benchmark approaches using simulated data. Simulations are used for the lack of available open-source datasets appropriate for SLAM in dynamic scenes. The results demonstrate that GEM-SLAM results in significant improvements in observer localization and feature mapping accuracy for increasing report errors, False Alarm Rates (FARs) and dynamic features.

This paper is structured as follows: Section II provides the problem formulation. The necessary background theory on RFSs is summarized in Section III. The proposed approach is derived and its relationships analyzed in Section IV. A model-specific realization is provided in Section V. The experimental setup is described in Section VI, results are discussed in Section VII, and conclusions are drawn in Section VIII.

II. PROBLEM FORMULATION

A general formulation describes the positional observer state, \mathbf{r}_t , at time step t , as a function of the control input, \mathbf{u}_t , and the process noise, \mathbf{v}_t , i.e.,

$$\mathbf{r}_t = f(\mathbf{r}_{t-1}, \mathbf{u}_t, \mathbf{v}_t), \quad (1)$$

where $f(\cdot)$ captures the observer dynamics (the reader is referred to the nomenclature table in Appendix C). Reports, \mathbf{y}_t , of the

observer motion are modelled as a function of the states and report error, \mathbf{w}_t , i.e.,

$$\mathbf{y}_t = h(\mathbf{r}_t, \mathbf{w}_t), \quad (2)$$

where $h(\cdot)$ is the report model.

Feature-based SLAM represents the state of each of N_t objects by a single, salient feature. The absolute state, $\mathbf{s}_{t,n}^a$, of feature $n \in 1, \dots, N_t$, in the global coordinate frame is given by

$$\mathbf{s}_{t,n}^a = d(\mathbf{s}_{t-1,n}^a, \mathbf{n}_{t,n}), \quad (3)$$

where $d(\cdot)$ captures the dynamics of the feature and the process noise, $\mathbf{n}_{t,n}$, allows for variations of the feature state from the dynamical model. Objects may appear in and leave the scene over time, therefore leading to feature birth and death. The number of features hence is time-varying and unknown *a priori*, such that the N_t feature states can be described as a RFS [39] with realizations, \mathcal{S}_t , modelled as

$$\mathcal{S}_t = \left[\bigcup_{n=1}^{N_{t-1}} P(\mathbf{s}_{t-1,n}) \right] \cup B_t, \quad (4)$$

where $\mathbf{s}_{t,n}$ is the state of feature n relative to the observer state, \mathbf{r}_t , the set cardinality is $|\mathcal{S}_t| = N_t$, the newborn feature process is B_t , and $P(\mathbf{s}_{t-1,n})$ is the survival process:

$$P(\mathbf{s}_{t-1,n}) = \begin{cases} \{\mathbf{s}_{t,n}\}, & \text{if feature } n \text{ is persistent} \\ \emptyset, & \text{if feature } n \text{ terminates} \end{cases} \quad (5)$$

where \emptyset denotes the empty set.

Given the sensor measurements, M_t detections, $\omega_{t,m}$, $m = 1, \dots, M_t$, of the N_t features are obtained relative to \mathbf{r}_t , i.e.,

$$\omega_{t,m} = \bar{g}(\mathbf{s}_{t,n}, \mathbf{e}_{t,m}), \quad (6)$$

where the mapping from the feature to the detection space is given by $\bar{g}(\cdot)$, and $\mathbf{e}_{t,m}$ describes the detection error.

In practice, detection algorithms often result in missing detections and false alarms. The feature detections thus are modelled by a RFS with realizations, Ω_t , such that

$$\Omega_t = \left[\bigcup_{n=1}^{N_t} D(\mathbf{s}_{t,n}) \right] \cup K_t, \quad (7)$$

where $D(\mathbf{s}_{t,n})$ models the detection process as

$$D(\mathbf{s}_{t,n}) = \begin{cases} \{\omega_{t,m}\} & \text{if feature } n \text{ is detected} \\ \emptyset & \text{if the feature is not detected} \end{cases} \quad (8)$$

for $m \in 1, \dots, M_t$. The term K_t in (7) is the RFS of false alarms, commonly modelled as a Poisson point process [30], [40], [41] of $N_{t,c}$ Independent and Identically Distributed (IID) false alarms distributed uniformly over the sensor Field of View (FoV) [29], [36]. The likelihood, $\kappa(\omega_{t,m} | \mathbf{r}_t)$, of a single-detection corresponding to a false alarms is hence

$$\kappa(\omega_{t,m} | \mathbf{r}_t) = \lambda_c \mathcal{U}(\text{FoV}), \quad (9)$$

where $\mathcal{U}(\cdot)$ denotes a uniform pdf, and λ_c is the maximum number of false alarms in the surveillance region.

III. FISST PREREQUISITES

This section provides a summary of Finite Set Statistics (FISST) required for the derivation of GEM-SLAM.

A. Multi-feature Probability Density Function

Assuming the observer state, \mathbf{r}_t , is known, the posterior feature pdf, $p(\mathcal{S}_t | \mathbf{r}_t, \Omega_{1:t})$, is propagated sequentially via

$$p(\mathcal{S}_t | \mathbf{r}_t, \Omega_{1:t}) = \frac{p(\Omega_t | \mathbf{r}_t, \mathcal{S}_t) p(\mathcal{S}_t | \mathbf{r}_t, \Omega_{1:t-1})}{\int p(\Omega_t | \mathbf{r}_t, \mathcal{S}_t) p(\mathcal{S}_t | \mathbf{r}_t, \Omega_{1:t-1}) \delta \mathcal{S}_t}, \quad (10)$$

where $p(\Omega_t | \mathbf{r}_t, \mathcal{S}_t)$ is the detection likelihood. The predicted pdf, $p(\mathcal{S}_t | \mathbf{r}_t, \Omega_{1:t-1})$, is obtained from the prior, $p(\mathcal{S}_t | \mathbf{r}_t, \mathcal{S}_{t-1})$, and the posterior, $p(\mathcal{S}_{t-1} | \mathbf{r}_t, \Omega_{1:t-1})$, at $t-1$ and relative to \mathbf{r}_t as

$$p(\mathcal{S}_t | \mathbf{r}_t, \Omega_{1:t-1}) = \int p(\mathcal{S}_t | \mathbf{r}_t, \mathcal{S}_{t-1}) p(\mathcal{S}_{t-1} | \mathbf{r}_t, \Omega_{1:t-1}) \delta \mathcal{S}_{t-1}. \quad (11)$$

The set integral, $\int \cdot \delta \mathbf{W}$, for any, \mathbf{W} , is defined as [30]

$$\begin{aligned} \int p(\mathbf{W}) \delta \mathbf{W} &= p(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int \dots \int p(\{\mathbf{w}_1, \dots, \mathbf{w}_n\}) d\mathbf{w}_1 \dots d\mathbf{w}_n. \end{aligned} \quad (12)$$

The summation in (12) enumerates the hypotheses that any number of objects can be contained in \mathbf{W} . The set integral is hence equivalent to marginalizing over all subsets $\{\mathbf{w}_1, \dots, \mathbf{w}_n\} \subset \mathbf{W}$, $\forall n = 0, \dots, \infty$. However, as a consequence, (10) is combinatorially intractable.

B. Probability Hypothesis Density

The first-order multi-object moment, referred to as the PHD, $\lambda(\mathbf{w})$, can be used to approximate (10). The PHD expresses the probability that one of the multiple objects in \mathbf{W} has the state \mathbf{w} . If $p(\mathbf{W})$ is a Poisson point process [40], i.e., the number of objects, L , in \mathbf{W} is Poisson distributed and the feature states are IID, the pdf, $p(\mathbf{W})$, can be expressed in terms of its PHD, $\lambda(\mathbf{w})$, and *vice versa* as [30], [39]

$$p(\mathbf{W}) = e^{-L} \prod_{\mathbf{w} \in \mathbf{W}} \lambda(\mathbf{w}), \quad (13)$$

$$\lambda(\mathbf{w}) = \int \delta_{\mathbf{W}}(\mathbf{w}) p(\mathbf{W}) \delta \mathbf{W}, \quad (14)$$

where $\delta_{\mathbf{W}}(\mathbf{w}) = \sum_{\mathbf{w}' \in \mathbf{W}} \delta_{\mathbf{w}'}(\mathbf{w})$ is the sum of Dirac-Delta functions concentrated at $\mathbf{w}' \in \mathbf{W}$. By definition, the PHD is not a pdf but rather an intensity, since [30]:

$$\int_{\mathcal{S}} \lambda(\mathbf{w}) d\mathbf{w} = \mathbb{E}[L], \quad (15)$$

where L is the number of objects in the region, \mathcal{S} , of the state space, and $\mathbb{E}[\cdot]$ is the expectation operator. The expression in (15) is a crucial property as estimation of $\lambda(\mathbf{w})$ simultaneously results in an estimate L .

C. Multi-Feature PHD for Known Observer Positions

The PHD filter was originally proposed in [30] for a static observer with known position. The results can be

straightforwardly extended to a dynamic observer with known states as summarized in (25) set at bottom of the page.

The posterior PHD, $\lambda(\mathbf{s}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t})$ for any $\mathbf{s}_t \in \mathcal{S}_t$, conditional on the *a priori* known state of a moving observer, models the hypotheses that 1) any detection may be due to a newborn source, 2) any detection may originate from any of the existing features, or 3) any existing feature may not be detected, i.e., [30]

$$\lambda(\mathbf{s}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t}) = \lambda_b(\mathbf{s}_t | \mathbf{r}_t) + \lambda_m(\mathbf{s}_t | \mathbf{r}_t) + \lambda_d(\mathbf{s}_t | \mathbf{r}_t) \quad (16)$$

where the PHDs of newborn, missed, and detected features, $\lambda_b(\mathbf{s}_t | \mathbf{r}_t)$, $\lambda_m(\mathbf{s}_t | \mathbf{r}_t)$, and $\lambda_d(\mathbf{s}_t | \mathbf{r}_t)$ respectively, are

$$\lambda_b(\mathbf{s}_t | \mathbf{r}_t) = p_b(\mathbf{s}_t) \lambda(\mathbf{s}_t | \mathbf{r}_t, \mathbf{\Omega}_t), \quad (17a)$$

$$\lambda_m(\mathbf{s}_t | \mathbf{r}_t) = (1 - p_d(\mathbf{s}_t)) \lambda(\mathbf{s}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t-1}), \quad (17b)$$

$$\lambda_d(\mathbf{s}_t | \mathbf{r}_t) = p_d(\mathbf{s}_t) \sum_{m=1}^{M_t} \frac{p(\boldsymbol{\omega}_{t,m} | \mathbf{r}_t, \mathbf{s}_t) \lambda(\mathbf{s}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t-1})}{\ell(\boldsymbol{\omega}_{t,m} | \mathbf{r}_t)}. \quad (17c)$$

where $\lambda(\mathbf{s}_t | \mathbf{r}_t, \mathbf{\Omega}_t)$ is the initial feature PHD, $p_b(\mathbf{s}_t)$ is the probability of newborn features, $p_d(\mathbf{s}_t)$ is the probability of detection, and $p(\boldsymbol{\omega}_{t,m} | \mathbf{r}_t, \mathbf{s}_t)$ and $\ell(\boldsymbol{\omega}_{t,m} | \mathbf{r}_t)$ are the single-detection likelihood and evidence respectively, i.e.,

$$\ell(\boldsymbol{\omega}_{t,m} | \mathbf{r}_t) = \kappa(\boldsymbol{\omega}_{t,m} | \mathbf{r}_t) + p(\boldsymbol{\omega}_{t,m} | \mathbf{r}_t), \quad (18)$$

$$\begin{aligned} p(\boldsymbol{\omega}_{t,m} | \mathbf{r}_t) \\ = \int p_d(\mathbf{s}_t) \lambda(\mathbf{s}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t-1}) p(\boldsymbol{\omega}_{t,m} | \mathbf{r}_t, \mathbf{s}_t) d\mathbf{s}_t. \end{aligned} \quad (19)$$

The predicted PHD, $\lambda(\mathbf{s}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t-1})$, is

$$\begin{aligned} \lambda(\mathbf{s}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t-1}) \\ = \int p_s(\mathbf{s}_{t-1}) p(\mathbf{s}_t | \mathbf{r}_t, \mathbf{s}_{t-1}) \lambda(\mathbf{s}_{t-1} | \mathbf{r}_t, \mathbf{\Omega}_{1:t-1}) d\mathbf{s}_{t-1}, \end{aligned} \quad (20)$$

where $p_s(\mathbf{s}_{t-1})$ is the survival probability, $p(\mathbf{s}_t | \mathbf{r}_t, \mathbf{s}_{t-1})$ is the single-feature transition density, and the PHD, $\lambda(\mathbf{s}_{t-1} | \mathbf{r}_t, \mathbf{\Omega}_{1:t-1})$ is obtained by transforming $\lambda(\mathbf{s}_{t-1} | \mathbf{r}_{t-1}, \mathbf{\Omega}_{1:t-1})$ to the current \mathbf{r}_t .

IV. PROPOSED GEM-SLAM ALGORITHM

SLAM sequentially estimates the joint marginal posterior pdf, $p(\mathbf{r}_t, \mathcal{S}_t | \mathbf{\Omega}_{1:t}, \mathbf{y}_{1:t})$. In this section, the proposed GEM-SLAM algorithm is derived.

A. Posterior pdf

Defining $\mathbf{X}_t \triangleq (\mathbf{r}_t, \mathcal{S}_t)$ and $\mathbf{Z}_t \triangleq (\mathbf{y}_t, \mathbf{\Omega}_t)$ as realizations of marked point processes [40] corresponding to the joint states and measurements respectively, the posterior SLAM pdf, $p(\mathbf{X}_t | \mathbf{Z}_{1:t})$, can be written using Bayes's theorem as

$$p(\mathbf{X}_t | \mathbf{Z}_{1:t}) = \frac{p(\mathbf{Z}_t | \mathbf{X}_t) p(\mathbf{X}_t | \mathbf{Z}_{1:t-1})}{\int p(\mathbf{Z}_t | \mathbf{X}_t) p(\mathbf{X}_t | \mathbf{Z}_{1:t-1}) \delta \mathbf{X}_t}, \quad (21)$$

where $p(\mathbf{Z}_t | \mathbf{X}_t)$ is the likelihood function of the observations, and the predicted density, $p(\mathbf{X}_t | \mathbf{Z}_{1:t-1})$, is obtained by marginalizing, \mathbf{X}_{t-1} , from the transition density, $p(\mathbf{X}_t | \mathbf{X}_{t-1})$, and the posterior at $t-1$, i.e.,

$$\begin{aligned} p(\mathbf{X}_t | \mathbf{Z}_{1:t-1}) \\ = \int p(\mathbf{X}_t | \mathbf{X}_{t-1}) p(\mathbf{X}_{t-1} | \mathbf{Z}_{1:t-1}) \delta \mathbf{X}_{t-1}. \end{aligned} \quad (22)$$

Similar to (10), the posterior SLAM pdf is combinatorially intractable, but can be approximated by its PHD as detailed in the following subsections. The derivations can be found in Appendix A.

B. Predicted GEM-SLAM PHD

Using the probability chain rule, the predicted pdf in (22) can be written as

$$\begin{aligned} p(\mathbf{X}_t | \mathbf{Z}_{1:t-1}) \\ = \int p(\mathbf{r}_t | \mathbf{r}_{t-1}) p(\mathbf{r}_{t-1} | \mathbf{y}_{1:t-1}) p(\mathcal{S}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t-1}) d\mathbf{r}_{t-1}, \end{aligned} \quad (23)$$

where $p(\mathcal{S}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t-1})$ is defined in (11), and $p(\mathbf{r}_t | \mathbf{r}_{t-1})$ and $p(\mathbf{r}_{t-1} | \mathbf{y}_{1:t-1})$ are the observer prior pdf and posterior pdf at $t-1$ respectively. As derived in Appendix A-A, the joint predicted PHD of (23) is given by

$$\begin{aligned} \lambda(\mathbf{x}_t | \mathbf{Z}_{1:t-1}) \\ = \int p(\mathbf{r}_t | \mathbf{r}_{t-1}) p(\mathbf{r}_{t-1} | \mathbf{y}_{1:t-1}) \lambda(\mathbf{s}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t-1}) d\mathbf{r}_{t-1}, \end{aligned} \quad (24)$$

for any $\mathbf{x}_t \in \mathbf{X}_t$, where $\lambda(\mathbf{s}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t-1})$ is given in (20). The joint predicted PHD therefore extrapolates $p(\mathbf{r}_{t-1} | \mathbf{y}_{1:t-1})$ using the observer dynamics specified by the prior, $p(\mathbf{r}_t | \mathbf{r}_{t-1})$, and marginalizes the previous observer state, \mathbf{r}_{t-1} , from the result as well as the predicted feature PHD.

C. Updated GEM-SLAM PHD

Similar to (23), the probability chain rule can be applied to (21), leading to (25) below. Note that Bayes's theorem for the posterior observer pdf, $p(\mathbf{r}_t | \mathbf{y}_{1:t})$, and similarly $p(\mathcal{S}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t})$, can be arranged to

$$p(\mathbf{y}_t | \mathbf{r}_t) p(\mathbf{r}_t | \mathbf{y}_{1:t-1}) = p(\mathbf{y}_t) p(\mathbf{r}_t | \mathbf{y}_{1:t}), \quad (26)$$

where $p(\mathbf{y}_t | \mathbf{r}_t)$ and $p(\mathbf{r}_t | \mathbf{y}_{1:t-1})$ are the likelihood and predicted observer pdf respectively. Inserting (26) into (25):

$$p(\mathbf{X}_t | \mathbf{Z}_{1:t}) = \frac{p(\mathbf{r}_t | \mathbf{y}_{1:t}) p(\mathbf{\Omega}_t | \mathbf{r}_t)}{\int p(\mathbf{r}_t | \mathbf{y}_{1:t}) p(\mathbf{\Omega}_t | \mathbf{r}_t) d\mathbf{r}_t} p(\mathcal{S}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t}) \quad (27)$$

$$p(\mathbf{X}_t | \mathbf{Z}_{1:t}) = \frac{p(\mathbf{y}_t | \mathbf{r}_t) p(\mathbf{r}_t | \mathbf{y}_{1:t-1}) p(\mathbf{\Omega}_t | \mathbf{r}_t, \mathcal{S}_t) p(\mathcal{S}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t-1})}{\iint p(\mathbf{y}_t | \mathbf{r}_t) p(\mathbf{r}_t | \mathbf{y}_{1:t-1}) p(\mathbf{\Omega}_t | \mathbf{r}_t, \mathcal{S}_t) p(\mathcal{S}_t | \mathbf{\Omega}_{1:t-1}) \delta \mathcal{S}_t d\mathbf{r}_t}, \quad (25)$$

where the multi-detection evidence, $p(\Omega_t | \mathbf{r}_t)$, is given by the Chapman-Kolmogorov equation [42] as

$$p(\Omega_t | \mathbf{r}_t) = \int p(\Omega_t | \mathbf{r}_t, \mathcal{S}_t) p(\mathcal{S}_t | \mathbf{r}_t, \Omega_{1:t-1}) \delta \mathcal{S}_t. \quad (28)$$

For tractability of (27), $p(\mathbf{X}_t | \mathbf{Z}_{1:t})$ is approximated by its PHD, $\lambda(\mathbf{x}_t | \mathbf{Z}_{1:t})$, which, from Appendix A-B, is

$$\lambda(\mathbf{x}_t | \mathbf{Z}_{1:t}) = \frac{\mathcal{L}(\Omega_t | \mathbf{r}_t) p(\mathbf{r}_t | \mathbf{y}_{1:t})}{\int \mathcal{L}(\Omega_t | \mathbf{r}_t) p(\mathbf{r}_t | \mathbf{y}_{1:t}) d\mathbf{r}_t} \lambda(\mathbf{s}_t | \mathbf{r}_t, \Omega_{1:t}), \quad (29)$$

where $\lambda(\mathbf{s}_t | \mathbf{r}_t, \Omega_{1:t})$ is provided in (16) and (17). The multi-detection evidence, $\mathcal{L}(\Omega_t | \mathbf{r}_t)$, is given by

$$\mathcal{L}(\Omega_t | \mathbf{r}_t) \triangleq e^{-N_{t,c} - N_{t|t-1}} \prod_{m=1}^{M_t} \ell(\omega_{t,m} | \mathbf{r}_t), \quad (30)$$

where $\ell(\omega_{t,m} | \mathbf{r}_t)$ is defined in (18), and $N_{t,c}$ and $N_{t|t-1}$ respectively are the number of false alarms and predicted features obtained by applying (15).

The result in (29) is crucial as it allows for full exploitation of the joint dependency between the features and the observer. The features are dependent on the observer through $\lambda(\mathbf{s}_t | \mathbf{r}_t, \Omega_{1:t})$. Simultaneously, the observer pdf is dependent on the features through the multi-detection evidence, $\mathcal{L}(\Omega_t | \mathbf{r}_t)$, facilitating probabilistic anchoring of the observer using the features. This point is further elaborated on in Section V-C and Section V-E for the model-specific GEM-SLAM realization. The necessary equations for the general formulation of GEM-SLAM are summarized in Algorithm 1.

To provide insight into the physical meaning of the multi-detection evidence in (30), recall from (18) that the single-detection evidence, $\ell(\omega_{t,m} | \mathbf{r}_t)$, models the probability of the detection being either due to a false alarm with probability $\kappa(\omega_{t,m} | \mathbf{r}_t)$, or due to a feature with probability $p(\omega_{t,m} | \mathbf{r}_t)$. For reasons of illustration, consider expanding the closed-form solution in (30) using (18), leading to the combinatorial expression:

$$\mathcal{L}(\Omega_t | \mathbf{r}_t) = e^{-N_{t,c} - N_{t|t-1}} \left[\prod_{m=1}^{M_t} \kappa(\omega_{t,m} | \mathbf{r}_t) + \sum_{\mathbf{W} \boxminus \Omega_t} \prod_{\mathbf{w} \in \mathbf{W}} p(\mathbf{w} | \mathbf{r}_t) \prod_{\mathbf{z} \in \{\Omega_t - \mathbf{W}\}} \kappa(\mathbf{z} | \mathbf{r}_t) \right], \quad (31)$$

where $\mathbf{W} \boxminus \Omega_t$ denotes the partitions of set Ω_t not including the empty set, and $\{\Omega_t - \mathbf{W}\}$ denotes the set difference. The expression in (31) highlights that GEM-SLAM enumerates explicitly all hypotheses of the origin of detections, i.e., 1) all M_t detections are false alarms; 2) all detections are due to features at $\mathbf{W} = \Omega_t$; and 3) each set partition $\mathbf{W} \boxminus \Omega_t$ are due to features, whilst the remaining detections, $\Omega_t - \mathbf{W}$, are false alarms. As the expression in (31) involves combinatorial terms, it is computationally intractable. This problem is avoided by the closed-form expressions in (30) and (18). In fact, the formulation of the multi-detection likelihood is one of the fundamental differences between the RB-PHD, SC-PHD filter and GEM-SLAM as discussed in the following subsection.

Algorithm 1: Summary of GEM-SLAM.

- 1: Feature prediction, (20);
 - 2: Feature update, (16)–(17);
 - 3: Single-detection evidence, (18);
 - 4: Multi-detection evidence, (30);
 - 5: Predicted SLAM PHD, (24);
 - 6: Updated SLAM PHD, (29).
-

D. Relationship to Existing Approaches

1) *SC-PHD SLAM*: The SC-PHD filter [36], [43] was the first approach based on the closed-form solution of the multi-detection evidence in (30). The SC-PHD filter formulates the SLAM PHD as:

$$\lambda^{\text{SC}}(\mathbf{x}_t | \mathbf{Z}_{1:t}) = p^{\text{SC}}(\mathbf{r}_t | \Omega_{1:t}) \lambda(\mathbf{s}_t | \mathbf{r}_t, \Omega_{1:t}), \quad (32)$$

where the observer posterior pdf in (32) is given by

$$p^{\text{SC}}(\mathbf{r}_t | \Omega_{1:t}) \triangleq \frac{\mathcal{L}(\Omega_t | \mathbf{r}_t) p(\mathbf{r}_t | \Omega_{1:t-1})}{\int \mathcal{L}(\Omega_t | \mathbf{r}_t) p(\mathbf{r}_t | \Omega_{1:t-1}) d\mathbf{r}_t}. \quad (33)$$

Therefore, the SC-PHD filter propagates the observer estimates using the predicted pdf, $p(\mathbf{r}_t | \Omega_{1:t-1})$, only. Recalling the predicted SLAM PHD in (24) and the dynamical model in (1), the SC-PHD filter therefore exploits the uncertainty in the dynamical model of the observer, but cannot incorporate the uncertainty in the reports.

In contrast, GEM-SLAM incorporates the uncertainty in the observer dynamics as well as the reports via the posterior pdf of the observer states. In (29), the observer information is optimally fused with information inferred from the features by intersecting the space of likely observer states of the posterior pdf with that of the multi-detection evidence. The optimal exploitation of additional information about the observer motion is crucial for dynamic scenes in order to disambiguate information inferred from the time-varying feature states.

2) *RB-PHD SLAM*: Similar to the SC-PHD filter, the RB-PHD filter [28], [29] relies on the predicted pdf of the observer state only. Moreover, the RB-PHD filter is based on the computationally intractable multi-detection evidence in (31). To address the combinatorial nature of the expression, the RB-PHD filter relies on an approximation, which considers the contribution of a single feature only, (see Appendix B):

$$\mathcal{L}^{\text{RB}}(\Omega_t | \mathbf{r}_t) \approx e^{-N_{t,c} - N_{t|t-1}} \left[\prod_{m=1}^{M_t} \kappa(\omega_{t,m} | \mathbf{r}_t) + \sum_{m=1}^{M_t} \hat{q}(\omega_{t,m} | \mathbf{r}_t) \prod_{\substack{\ell=1 \\ \ell \neq m}}^{M_t} \kappa(\omega_{t,\ell} | \mathbf{r}_t) \right], \quad (34)$$

where

$$\hat{q}(\omega_{t,m} | \mathbf{r}_t) \triangleq \int p_d(\mathbf{s}_t) p(\omega_{t,m} | \hat{\mathbf{s}}_t, \mathbf{r}_t) p(\hat{\mathbf{s}}_t | \mathbf{r}_t, \mathbf{Z}_{1:t-1}) d\hat{\mathbf{s}}_t$$

and $\hat{\mathbf{s}}_t$ is the single, selected feature. The multi-detection evidence of the RB-PHD filter in (34) therefore accounts for two hypotheses: either all detections are false alarms or all detections are false alarms with the exception of a single detection that is due to a feature. In contrast, GEM-SLAM in (30) accounts for

the hypotheses that any partition of detections may be due to features. GEM-SLAM therefore reduces to the RB-PHD filter if $N_t = 1$ and $p(\mathbf{y}_t | \mathbf{r}_t) = \text{const}$.

3) *FastSLAM*: FastSLAM [23], [24], as an approach based on Random Variables (RVs) rather than RFSs, assumes that all features are static and exist permanently in the scene. Furthermore, it is assumed that at most one feature detection, ω_t , is observed at each time step. The joint posterior pdf of the observer state and features is therefore factorized as [23]:

$$p(\mathbf{X}_t | \mathbf{r}_t, \boldsymbol{\omega}_{1:t}, \mathbf{y}_{1:t}) \approx p(\mathbf{r}_t | \boldsymbol{\omega}_t, \mathbf{y}_t) \prod_{n=1}^{N_t} p(\mathbf{s}_n | \mathbf{r}_t, \boldsymbol{\omega}_{1:t}, k_t),$$

where k_t is the detection-to-feature assignment, obtained by heuristic [44] or probabilistic [25], [45] data association, and $p(\mathbf{r}_t | \boldsymbol{\omega}_t, \mathbf{y}_t)$ is the observer posterior pdf, evaluated using the current, single detection ω_t . Using Bayes's theorem, the single-feature update is hence expressed by

$$p(\mathbf{s}_n | \mathbf{r}_t, \boldsymbol{\omega}_{1:t}, k_t) = \frac{p(\boldsymbol{\omega}_t | \mathbf{r}_t, \mathbf{s}_n, k_t) p(\mathbf{s}_n | \boldsymbol{\omega}_{1:t-1})}{\int p(\boldsymbol{\omega}_t | \mathbf{r}_t, \mathbf{s}_n, k_t) p(\mathbf{s}_n | \boldsymbol{\omega}_{1:t-1}) d\mathbf{s}_n}. \quad (35)$$

The feature PHD in (16) reduces to (35) for a detection probability of $p_d = 1$, survival probability $p_s = 1$, false alarm intensity $\kappa(\boldsymbol{\omega}_t | \mathbf{r}_t) = 0$, and detection cardinality of $M_t = 1$. Furthermore, assume that the multi-detection RFS, $\boldsymbol{\Omega}_t$, can be considered as a sequence of M_t single detections observed between $[t-1, t]$. The multi-detection evidence in (30) thus reduces to $\mathcal{L}(\boldsymbol{\Omega}_t | \mathbf{r}_t) = e^{-N_t |t-1} p(\boldsymbol{\omega}_t | \mathbf{r}_t)$. Therefore, under the above assumptions, GEM-SLAM reduces to FastSLAM by a scaling factor of $e^{-N_t |t-1}$.

V. MODEL-SPECIFIC REALIZATION

This section provides a model-specific realization of GEM-SLAM for range-bearing sensors and dynamic features.

A. Model

To provide a model-specific realization of GEM-SLAM, consider as an illustrative example a range-bearing sensor installed in the body of a humanoid robot. The observer state is defined as $\mathbf{r}_t \triangleq [\tilde{\mathbf{r}}_t^T, v_t, \gamma_t]^T$ where $\tilde{\mathbf{r}}_t = [x_t, y_t, z_t]^T$ is the 3D position, v_t denotes the velocity, and γ_t is the orientation. Assuming that the observer moves in the direction of its orientation, the position, $\tilde{\mathbf{r}}_t$, becomes a non-linear function of the orientation. The observer state can hence be separated into a linear subspace, $\mathbf{p}_t \triangleq [\tilde{\mathbf{r}}_t^T, v_t]^T$, and its non-linear dependency, γ_t , such that $\mathbf{r}_t \triangleq [\mathbf{p}_t^T, \gamma_t]^T$. Hence:

$$\mathbf{p}_t = \mathbf{F}_t \mathbf{p}_{t-1} + \mathbf{v}_{t,v}, \quad \mathbf{v}_{t,v} \sim \mathcal{N}(\mathbf{0}_{4 \times 1}, \boldsymbol{\Sigma}_{t,v,v}) \quad (36a)$$

$$\gamma_t = \vartheta(\gamma_{t-1} + v_{t,\gamma}), \quad v_{t,\gamma} \sim \mathcal{N}(0, \sigma_{t,\gamma}^2) \quad (36b)$$

where $\mathbf{0}_{P \times 1}$ is the $P \times 1$ zero vector, $\mathcal{N}(\cdot)$ denotes the Gaussian pdf, $\mathbf{v}_{t,v}$ is the process noise with covariance $\boldsymbol{\Sigma}_{t,v,v}$, and the orientation in (36b) obeys a random walk with process noise, $v_{t,\gamma}$, process noise variance, $\sigma_{t,\gamma}^2$, and $\vartheta(\alpha) = \text{mod}(\alpha, 2\pi)$. Furthermore,

$$\mathbf{F}_t \triangleq \begin{bmatrix} \mathbf{I}_3 & [\Delta_t \sin(\gamma_t), \Delta_t \cos(\gamma_t), 0]^T \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix}, \quad (37)$$

where \mathbf{I}_P is the $P \times P$ identity matrix and Δ_t is the time delay between $t-1$ and t . The observer reports, $\mathbf{y}_t \triangleq [y_{t,v}, y_{t,\gamma}]^T$, containing the reported speed, $y_{t,v}$, and orientation, $y_{t,\gamma}$, are noisy measurements of the state, \mathbf{r}_t , i.e.,

$$y_{t,v} = \mathbf{h} \mathbf{p}_t + w_{t,v}, \quad w_{t,v} \sim \mathcal{N}(0, \sigma_{t,w_v}^2) \quad (38a)$$

$$y_{t,\gamma} = \vartheta(\gamma_t + w_{t,\gamma}), \quad w_{t,\gamma} \sim \mathcal{N}(0, \sigma_{t,w_\gamma}^2) \quad (38b)$$

where $\mathbf{h} \triangleq [0, 0, 0, 1]^T$ and $w_{t,v}$ and $w_{t,\gamma}$ are the velocity and orientation report noise with variance σ_{t,w_v}^2 and σ_{t,w_γ}^2 .

The state of each feature, $n \in 1, \dots, N_t$, is defined as $\mathbf{s}_{t,n}^a \triangleq [x_{t,n}, y_{t,n}, z_{t,n}, \dot{x}_{t,n}, \dot{y}_{t,n}, \dot{z}_{t,n}]^T$, containing the source position and velocity, such that

$$\mathbf{s}_{t,n}^a = \mathbf{D}_t \mathbf{s}_{t-1,n}^a + \mathbf{n}_{t,n}, \quad \mathbf{n}_{t,n} \sim \mathcal{N}(\mathbf{0}_{6 \times 1}, \mathbf{Q}_t) \quad (39)$$

where $\mathbf{n}_{t,n}$ is the process noise with covariance \mathbf{Q}_t , and the dynamical model, \mathbf{D}_t , follows a constant-velocity model. Using a range-bearing sensor, each detection is defined as $\boldsymbol{\omega}_{t,m} \triangleq [r_{t,m}, \phi_{t,m}, \theta_{t,m}]^T$, with sensor-to-feature range, $r_{t,m}$, azimuth, $\phi_{t,m}$, and elevation, $\theta_{t,m}$, where

$$\boldsymbol{\omega}_{t,m} = g(\mathbf{s}_{t,n}) + \mathbf{e}_{t,m}, \quad \mathbf{e}_{t,m} \sim \mathcal{N}(\mathbf{0}_{3 \times 1}, \mathbf{R}_{t,m}) \quad (40)$$

where $g(\cdot)$ is the Cartesian-to-spherical transformation, $\mathbf{e}_{t,m}$ is the detection error with covariance, $\mathbf{R}_{t,m}$, and $\mathbf{s}_{t,n}$ is the sensor-relative feature state:

$$\mathbf{s}_{t,n} = \boldsymbol{\Gamma}(\gamma_t) \mathbf{s}_{t,n}^a + [\hat{\mathbf{r}}_t^T, \mathbf{0}_{1 \times 3}]^T, \quad (41a)$$

$$\boldsymbol{\Gamma}(\gamma_t) \triangleq \begin{bmatrix} \cos \gamma_t & -\sin \gamma_t & \mathbf{0}_{1 \times 4} \\ \sin \gamma_t & \cos \gamma_t & \mathbf{0}_{1 \times 4} \\ \mathbf{0}_{4 \times 1} & \mathbf{0}_{4 \times 1} & \mathbf{I}_4 \end{bmatrix}. \quad (41b)$$

B. GM-PHD Filter for Feature Mapping

For the Gaussian state-space in (39) and (40), the feature PHD in (16) reduces to a Gaussian Mixture Model (GMM), [31]:

$$\lambda(\mathbf{s}_t | \mathbf{r}_t, \boldsymbol{\Omega}_{1:t-1}) = \sum_{j=1}^{J_{t-1}} w_{t|t-1}^{(j)} \mathcal{N}(\mathbf{s}_t | \mathbf{m}_{t|t-1}^{(j)}, \boldsymbol{\Sigma}_{t|t-1}^{(j)}), \quad (42a)$$

$$\lambda_d(\mathbf{s}_t | \mathbf{r}_t) = \sum_{m=1}^{M_t} \sum_{j=1}^{J_{t-1}} w_t^{(j,m)} \mathcal{N}(\mathbf{s}_t | \mathbf{m}_t^{(j,m)}, \boldsymbol{\Sigma}_t^{(j,m)}). \quad (42b)$$

where J_{t-1} is the number of Gaussian Mixture (GM) components at $t-1$, $w_{t|t-1}^{(j)}$ and $w_t^{(j,m)}$ are the predicted and updated GM weights, and the predicted and updated mean, $\mathbf{m}_{t|t-1}^{(j)}$ and $\mathbf{m}_t^{(j,m)}$, and covariance, $\boldsymbol{\Sigma}_{t|t-1}^{(j)}$ and $\boldsymbol{\Sigma}_t^{(j,m)}$, are given by the

EKF [46], i.e.,

$$\mathbf{m}_{t|t-1}^{(j)} = \mathbf{D}_t \tilde{\mathbf{m}}_{t-1}^{(j)}, \quad (43a)$$

$$\Sigma_{t|t-1}^{(j)} = \mathbf{D}_t \tilde{\Sigma}_{t-1}^{(j)} \mathbf{D}_t^T + \mathbf{Q}_t, \quad (43b)$$

$$\mathbf{m}_t^{(j,m)} = \mathbf{m}_{t|t-1}^{(j)} + \mathbf{K}_t^{(j,m)} \left(\boldsymbol{\omega}_{t,m} - g(\mathbf{m}_{t|t-1}^{(j)}) \right), \quad (43c)$$

$$\Sigma_t^{(j,m)} = \left(\mathbf{I}_6 - \mathbf{K}_t^{(j,m)} \mathbf{G}_t^{(j)} \right) \Sigma_{t|t-1}^{(j)}, \quad (43d)$$

where the gain, $\mathbf{K}_t^{(j,m)}$, is given via the measurement Jacobian, $\mathbf{G}_t^{(j)}$, and innovation covariance, $\mathbf{S}_t^{(j,m)}$, as:

$$\mathbf{K}_t^{(j,m)} = \Sigma_{t|t-1}^{(j)} [\mathbf{G}_t^{(j)}]^T [\mathbf{S}_t^{(j,m)}]^{-1}, \quad (44a)$$

$$\mathbf{S}_t^{(j,m)} = \mathbf{G}_t^{(j)} \Sigma_{t|t-1}^{(j)} [\mathbf{G}_t^{(j)}]^T + \mathbf{R}_{t,m}. \quad (44b)$$

The previous mean and covariance, $\tilde{\mathbf{m}}_{t-1}^{(j)}$ and $\tilde{\Sigma}_{t-1}^{(j)}$, relative to the current \mathbf{r}_t , are obtained using (41) as

$$\begin{aligned} \tilde{\mathbf{m}}_{t-1}^{(j)} &= \Gamma(\gamma_t) \left[\left(\Gamma^{-1}(\gamma_{t-1}) \mathbf{m}_{t-1}^{(j)} + \mathbf{r}_{t-1} \right) - \mathbf{r}_t \right], \\ \tilde{\Sigma}_{t-1}^{(j)} &= \Gamma(\gamma_t) \Gamma^{-1}(\gamma_{t-1}) \Sigma_{t-1}^{(j)} \left[\Gamma^{-1}(\gamma_{t-1}) \right]^T \Gamma^T(\gamma_t). \end{aligned} \quad (45)$$

The GM weights are given by $w_{t|t-1}^{(j)} = p_s(\mathbf{m}_{t-1}^{(j)}) w_{t-1}^{(j)}$ and:

$$w_t^{(j,m)} = \frac{p_d(\mathbf{m}_{t|t-1}^{(j)}) w_{t|t-1}^{(j)}}{\ell(\boldsymbol{\omega}_{t,m} | \mathbf{r}_t)} \mathcal{N}\left(\boldsymbol{\omega}_{t,m} | g(\mathbf{m}_{t|t-1}^{(j)}), \mathbf{S}_t^{(j,m)}\right), \quad (46)$$

where $\ell(\boldsymbol{\omega}_{t,m} | \mathbf{r}_t)$ is defined in (18) with

$$p(\boldsymbol{\omega}_{t,m} | \mathbf{r}_t) = \sum_{j=1}^{J_{t-1}} w_{t|t-1}^{(j)} \mathcal{N}\left(\boldsymbol{\omega}_{t,m} | g(\mathbf{m}_{t|t-1}^{(j)}), \mathbf{S}_t^{(j,m)}\right). \quad (47)$$

Furthermore, the newborn PHD can be modelled as a detection-driven process [47], such that

$$\lambda_b(\mathbf{s}_t | \mathbf{r}_t) = \sum_{m=1}^{M_t} \sum_{j=1}^{J_b} w_{b,t}^{(j,m)} \mathcal{N}\left(\mathbf{s}_t | \mathbf{m}_{b,t}^{(j,m)}, \Sigma_{b,t}^{(j,m)}\right), \quad (48)$$

where J_b is the number of newborn components sampled from each of the M_t detections. The newborn GM mean is $\mathbf{m}_{b,t}^{(j,m)} = [g^{-1}(\boldsymbol{\omega}_{t,m})^T, \dot{x}_0, \dot{y}_0, \dot{z}_0]^T$, with covariance, $\Sigma_{b,t}^{(j,m)} = \mathbf{G}_t^{(j)} \mathbf{R}_{t,m} [\mathbf{G}_t^{(j)}]^T$, and weight, $w_{b,t}^{(j,m)} = p_b(\mathbf{m}_{b,t}^{(j,m)}) / J_b$, where $g^{-1}(\cdot)$ is the spherical-to-Cartesian transformation, and $(\dot{x}_0, \dot{y}_0, \dot{z}_0)$ are constant *a priori* initialization values of the feature velocity.

C. Rao-Blackwellized Particle Filter for Observer Localization

Recalling (36), the orientation, γ_t , is described by a wrapped Gaussian pdf [48], whilst \mathbf{p}_t is modelled by a Gaussian state-space that is non-linearly dependent on γ_t . To separate the linear Gaussian subspace from the non-Gaussian orientation, $p(\mathbf{r}_t | \mathbf{y}_{1:t})$ is factorized as

$$p(\mathbf{r}_t | \mathbf{y}_{1:t}) = p(\gamma_t | \mathbf{y}_{1:t,\gamma}) p(\mathbf{p}_t | \gamma_t, \mathbf{y}_{1:t,v}). \quad (49)$$

where $p(\gamma_t | \mathbf{y}_{1:t,\gamma})$ and $p(\mathbf{p}_t | \gamma_t, \mathbf{y}_{1:t,v})$ are the posterior pdfs of the orientation and linear subspace respectively. The formulation of GEM-SLAM facilitates estimation of \mathbf{r}_t using a Rao-Blackwellized particle filter [49], where γ_t is estimated by optimal Importance Sampling (IS) and \mathbf{p}_t is obtained from its optimal estimator formulated below. Using optimal IS, I particles, $\hat{\gamma}_t^{(i)}$, $i = 1, \dots, I$, of the orientation are drawn from

$$\hat{\gamma}_t^{(i)} \sim \mathcal{N}^w\left(\gamma_t | \mu_{t,\gamma}^{(i)}, \varsigma_{t,\gamma}^{(i)}\right), \quad (50)$$

where $\mathcal{N}^w(\cdot)$ denotes a wrapped Gaussian pdf, and the mean, $\mu_{t,\gamma}^{(i)}$, and variance, $\varsigma_{t,\gamma}^{(i)}$, are obtained by the wrapped Kalman Filter (KF) [50].

As \mathbf{p}_t is described by a linear Gaussian state-space in (36a) and (38a), the optimal estimator of \mathbf{p}_t is the KF. Therefore, for each $\hat{\gamma}_t^{(i)}$:

$$p(\mathbf{p}_t | \hat{\gamma}_t^{(i)}, \mathbf{y}_{1:t,v}) = \mathcal{N}\left(\mathbf{p}_t | \mu_{t,\mathbf{p}}^{(i)}, \varsigma_{t,\mathbf{p}}^{(i)}\right), \quad (51)$$

where the mean, $\mu_{t,\mathbf{p}}^{(i)}$, and covariance, $\varsigma_{t,\mathbf{p}}^{(i)}$, are provided by the KF equations based on the dynamical model in (37) evaluated at $\hat{\gamma}_t^{(i)}$. Applying (49)–(51) to (29):

$$\begin{aligned} \lambda(\mathbf{x}_t | \mathbf{Z}_{1:t}) &= \sum_{i=1}^I \alpha_t^{(i)} \delta_{\hat{\gamma}_t^{(i)}}(\gamma_t) p(\mathbf{p}_t | \hat{\gamma}_t^{(i)}, \mathbf{y}_{1:t,v}) \lambda(\mathbf{s}_t | \hat{\mathbf{r}}_t^{(i)}, \Omega_{1:t}) \\ &= \sum_{i=1}^I \alpha_t^{(i)} \delta_{\hat{\gamma}_t^{(i)}}(\gamma_t) p(\mathbf{p}_t | \hat{\gamma}_t^{(i)}, \mathbf{y}_{1:t,v}) \lambda(\mathbf{s}_t | \hat{\mathbf{r}}_t^{(i)}, \Omega_{1:t}) \end{aligned} \quad (52)$$

where $\hat{\mathbf{r}}_t^{(i)} \triangleq \left[[\mu_{t,\mathbf{p}}^{(i)}]^T, \hat{\gamma}_t^{(i)} \right]^T$, $\lambda(\mathbf{s}_t | \hat{\mathbf{r}}_t^{(i)}, \Omega_{1:t})$ is given by (16), and $\alpha_t^{(i)} = \hat{\alpha}_t^{(i)} / \sum_{j=1}^I \hat{\alpha}_t^{(j)}$ are the normalized IS weights. Using (29):

$$\hat{\alpha}_t^{(i)} = \alpha_t^{(i)} \mathcal{L}(\Omega_t | \hat{\mathbf{r}}_t^{(i)}) p(\mathbf{y}_t), \quad (53)$$

where $p(\mathbf{y}_t)$ is the observer report evidence, given as

$$\begin{aligned} p(\mathbf{y}_t) &= \mathcal{N}^w\left(\mathbf{y}_{t,\gamma} | \hat{\gamma}_t^{(i)}, \sigma_{t,\gamma}^2 + \sigma_{t,w_\gamma}^2\right) \\ &\quad \times \mathcal{N}\left(\mathbf{y}_{t,v} | \mathbf{h} \mu_{t,\mathbf{p}}^{(i)}, \mathbf{h} \varsigma_{t,\mathbf{p}}^{(i)} + \sigma_{t,v}^2 + \sigma_{t,w_v}^2\right), \end{aligned} \quad (54)$$

where $\sigma_{t,v}^2$ is the velocity component of Σ_{t,\mathbf{v}_v} . Furthermore, using (30), (18), (47):

$$\begin{aligned} \mathcal{L}(\Omega_t | \hat{\mathbf{r}}_t^{(i)}) &= e^{-N_{t,c}^{(i)} - N_{t|t-1}^{(i)}} \prod_{m=1}^{M_t} \left[\kappa\left(\boldsymbol{\omega}_{t,m} | \hat{\mathbf{r}}_t^{(i)}\right) \right. \\ &\quad \left. + \sum_{j=1}^{J_{t|t-1}} p_d(\mathbf{m}_{t|t-1}^{(i,j)}) w_{t|t-1}^{(i,j)} \mathcal{N}\left(\boldsymbol{\omega}_{t,m} | g(\mathbf{m}_{t|t-1}^{(i,j)}), \mathbf{S}_t^{(i,j,m)}\right) \right]. \end{aligned} \quad (55)$$

Therefore, the observer particles are weighted by the multi-detection and observer report evidence terms. The detection uncertainty in the observer orientation and speed, as well as the rotational and translational agility are explicitly accounted for via the observer evidence in (54). The multi-detection evidence in (55) facilitates exploitation of the existence of multiple features for observer localization. The product in (55) approaches zero for an increasing number of detections corresponding to

TABLE I
MODEL-SPECIFIC DIFFERENCES BETWEEN GEM-SLAM AND THE BENCHMARK ALGORITHMS

Algorithm	Observer estimation	IS weight	Feature mapping
GEM-SLAM	Rao-Blackwellized IS, (49) and (50)	Report pdf & pdf of detection set given feature set, (52)-(54)	GM-PHD
SC-PHD	Prior IS, (56)	pdf of detection set given feature set, (54)	GM-PHD
RB-PHD	Prior IS, (56)	pdf of detection set given single feature, (33)	GM-PHD
FastSLAM	Prior IS, (56)	pdf of single detection given single feature, (19)	EKF

Algorithm 2: GEM-SLAM.

```

1: Input  $\left\{ \hat{\mathbf{r}}_{t-1}^{(i)}, \alpha_{t-1}^{(i)}, \left\{ \mathbf{m}_{t-1}^{(i,j)}, \boldsymbol{\Sigma}_{t-1}^{(i,j)}, w_{t-1}^{(i,j)} \right\}_{j=1}^{J_{t-1}} \right\}_{i=1}^I$ 
2: for  $i = 1, \dots, I$  do
3:   Sample  $\hat{\mathbf{r}}_t^{(i)}$  (50), (51);
4:   for  $j = 1, \dots, J_{t-1}$  do
5:     Evaluate  $\tilde{\mathbf{m}}_{t-1}^{(i,j)}, \tilde{\boldsymbol{\Sigma}}_{t-1}^{(i,j)}$  relative to  $\hat{\mathbf{r}}_t^{(i)}$  (45)
6:     Predict  $\mathbf{m}_{t|t-1}^{(i,j)}, \boldsymbol{\Sigma}_{t|t-1}^{(i,j)}, w_{t|t-1}^{(i,j)}$  (43);
7:   end for
8:   for  $m = 1, \dots, M_t$  do
9:     for  $j = 1, \dots, J_b$  do
10:      Evaluate  $\mathbf{m}_{b,t}^{(i,j,m)}, \boldsymbol{\Sigma}_{b,t}^{(i,j,m)}, w_{b,t}^{(i,j,m)}$ ;
11:    end for
12:    for  $j = 1, \dots, J_{t-1}$  do
13:      Evaluate  $\mathbf{S}_t^{(i,j,m)}, \mathbf{K}_t^{(i,j,m)}$  (44);
14:      Update  $\mathbf{m}_t^{(i,j,m)}, \boldsymbol{\Sigma}_t^{(i,j,m)}, w_t^{(i,j,m)}$  (43);
15:    end for
16:    Evaluate  $\ell(\omega_{t,m} | \hat{\mathbf{r}}_t^{(i)})$  (18), (47);
17:  end for
18:  Evaluate  $\mathcal{L}(\boldsymbol{\Omega}_t | \hat{\mathbf{r}}_t^{(i)})$  (55)
19:  GM reduction [51]
20:  Evaluate  $\hat{\alpha}_t^{(i)}$  (53);
21:  Normalize weights,  $\alpha_t^{(i)}$ ;
22: end for
23: Resampling [49];

```

low likelihood values. The optimal feature-to-detection alignment thus corresponds to the observer particle that maximizes (55). Therefore, the observer is probabilistically anchored by identifying the location and orientation that corresponds to the optimal alignment between the features and detections. The GEM-SLAM pseudo-code is provided in Algorithm 2.

D. Relationship to Existing Approaches

As PHD-based approaches, the RB-PHD and SC-PHD filter use the same feature model as GEM-SLAM, estimated using the Gaussian Mixture PHD (GM-PHD) filter in Section V-B. In contrast, FastSLAM, based on RVs rather than RFSs, assumes that feature detections are processed one-by-one with known data association, and uses a single EKF for each feature.

The fundamental difference between the benchmark algorithms and GEM-SLAM is the approach to observer localization. The RB-PHD and SC-PHD filter as well as FastSLAM model the observer state, \mathbf{r}_t , as the Cartesian feature coordinates, $\hat{\mathbf{r}}_t$, and use the observer reports as parameters of the control input, $\mathbf{u}_t = \Delta_t y_{t,v} [\sin(y_{t,\gamma}), \cos(y_{t,\gamma}), 0]^T$ [25]. Hence,

$$\hat{\mathbf{r}}_t = \hat{\mathbf{r}}_{t-1} + \mathbf{u}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}_{3 \times 1}, \boldsymbol{\Sigma}_{t,\hat{\mathbf{r}}}) \quad (56)$$

where $\boldsymbol{\Sigma}_{t,\hat{\mathbf{r}}}$ is the process noise covariance. As (33) incorporates the predicted pdf, but not the updated posterior, observer localization for the RB-PHD, SC-PHD filter, and FastSLAM is realized by prior IS, i.e., [36]

$$v_t^{(i)} \sim \mathcal{N}(v_t | y_{t,v}, \sigma_{t,v}^2) \quad (57a)$$

$$\hat{\gamma}_t^{(i)} \sim \mathcal{N}(\gamma_t | y_{t,\gamma}, \sigma_{t,v_\gamma}^2), \quad (57b)$$

such that $\hat{\alpha}_t^{(i)} = \alpha_t^{(i)} \mathcal{L}(\boldsymbol{\Omega}_t | \mathbf{r}_t^{(i)})$. The particle state, $\hat{\mathbf{r}}_t^{(i)}$, is evaluated from (57) using (36). Based on (57), prior IS hence incorporates the uncertainty in the dynamical model of the observer, but cannot model uncertainty in the reports. Therefore, prior IS is expected to lead to increased estimation variance and hence decreased performance compared to optimal IS used in GEM-SLAM [52]. Furthermore, the GEM-SLAM particle weights in (53) compensate for the report evidence in addition to the multi-feature evidence. As discussed in Section IV-D1, this fusion is equivalent to evaluating the intersection between the areas of likely observer positions corresponding to the report and multi-detection evidence respectively. GEM-SLAM is therefore expected to result in a significantly reduced area of likely observer positions compared to the benchmark approaches.

Table I summarizes the model-specific differences between GEM-SLAM, the RB-PHD and SC-PHD filter as well as FastSLAM. The following subsection highlights these differences using an illustrative example.

E. Illustrative Comparison

Consider the initial observer position at (5, 2, 1.8) m in a $6 \times 6 \times 2.5$ m³ room with $\gamma_0 = 0$ rad and $v_0 = 0.5$ m/s. The observer trajectory is simulated using (36) and (38) with $\sigma_{t,v_\gamma} = 0.08$ rad, $\sigma_{t,v_v} = 0.1$ m/s, $\sigma_{t,w_\gamma} = 0.18$ rad and $\sigma_{t,w_v} = 0.5$ m/s. Two static features are initially located at (4, 3, 1.8) m and (2, 2, 1.8) m, both moving with a velocity of (0.1, 0.1, 0) m/s in the x-y-z directions. The feature states and detections are simulated using (39) and (40) for $\mathbf{Q}_t = 10^{-9} \mathbf{I}_6$ and $\mathbf{R}_t = 0.1 \mathbf{I}_3$. The number of uniformly distributed false alarms was simulated from a Poisson distribution with $\lambda_c = 1$. The room is divided into a grid with 5 cm² resolution in x - and y -coordinates. The z -positions of the co-located observer sensor and sources are assumed known *a priori*.

For each observer position in the grid, the EKF equations are used to update the two feature states with each of the two detections, resulting in four new feature states. The resulting importance weights of GEM-SLAM, the SC-PHD filter, the RB-PHD filter, and FastSLAM displayed as contour plots in Fig. 1. The ‘‘confidence area’’ corresponding to the outermost contour level is evaluated numerically based on the distances between vertices along the contour lines. The resulting area for FastSLAM assuming known data association is shown in Fig. 1(a) and

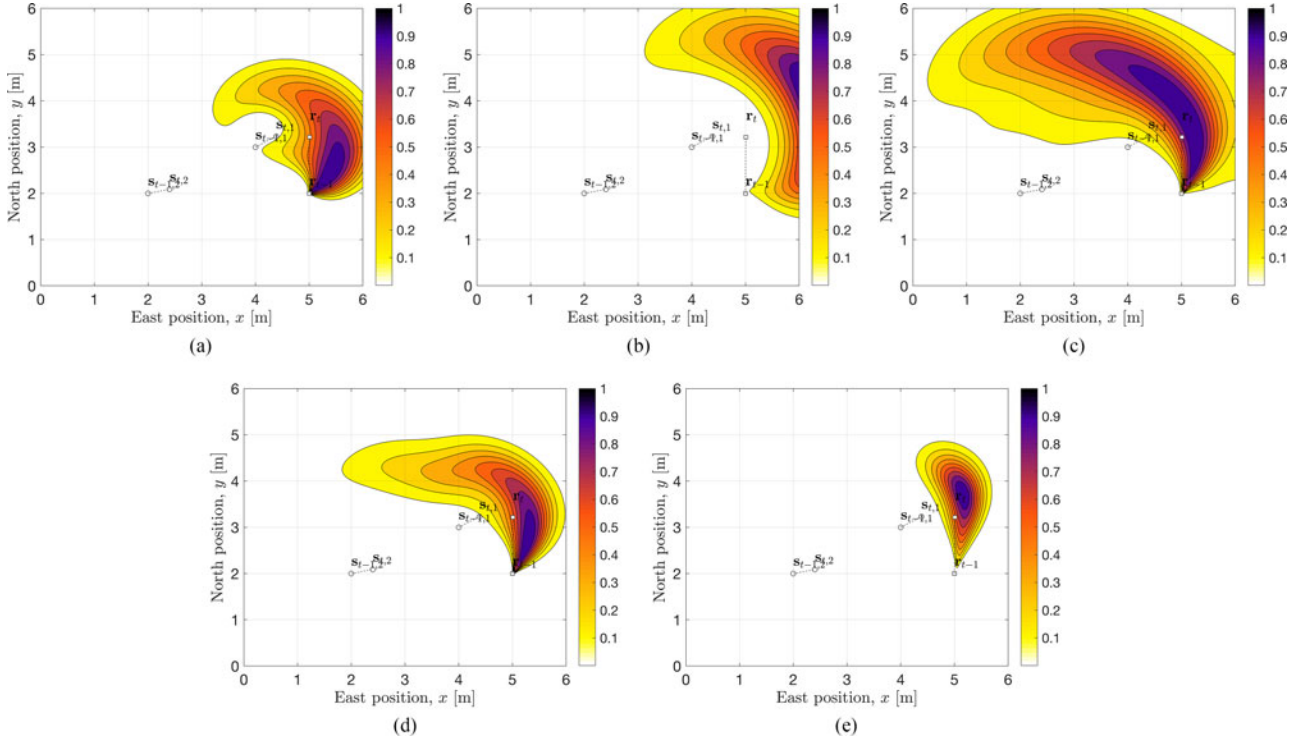


Fig. 1. Probabilistic anchoring of the observer for dynamic features: Map of the areas of likely observer positions for (a) FastSLAM assuming known data association; (b) FastSLAM with incorrect data association; (c) the RB-PHD filter, $\mathcal{L}^{\text{RB}}(\Omega_t | \mathbf{r}_t)$; (d) the SC-PHD filter, $\mathcal{L}(\Omega_t | \mathbf{r}_t)$; and (e) GEM-SLAM, $\mathcal{L}(\Omega_t | \mathbf{r}_t) p(\mathbf{y}_t)$.

corresponds to a confidence area of 5.35 m^2 . In practice, data association leads to false detection-to-feature assignments. The impact of false assignments is illustrated in Fig. 1(b), clearly showing that the ground truth observer position is outside of the confidence area. Therefore, FastSLAM leads to unreliable observer estimates for false data associations. The results for the RB-PHD filter are shown in Fig. 1(c) and correspond to a confidence area of 7.87 m^2 . The SC-PHD filter leads to an area of 6.22 m^2 (see Fig. 1(d)) by using the closed-form multi-detection evidence rather than the approximation by the RB-PHD filter. By fusing observer information inferred from the features with the observer reports, GEM-SLAM in Fig. 1(e) leads to a reduced confidence area of 2.24 m^2 .

F. Computational Complexity

The computational complexity of the GEM-SLAM realization in Algorithm 2 is $\mathcal{O}(M_t N_t I)$, i.e., linear in the feature detections, number of features and observer particles. Therefore, the complexity of GEM-SLAM is comparable to the RB-PHD filter and multi-hypothesis FastSLAM as shown in [53], as well as the SC-PHD filter [54]. As also mentioned in [53], all three PHD-filter approaches can be reduced to $\mathcal{O}(M_t I \log(N_t))$ using binary tree enhancements [25].

The crucial advantage of GEM-SLAM in terms of complexity is the required number of particles. As discussed in the previous subsection and highlighted in Fig. 1, GEM-SLAM needs to sample from a significantly reduced area of likely observer positions compared to the SC-PHD and RB-PHD filters as well as FastSLAM. Therefore, fewer particles are required for GEM-SLAM to achieve observer point estimates with comparable accuracy to the benchmark approaches.

VI. EXPERIMENTAL SETUP

A. Simulation Setup

The experimental setup is designed to decouple the GEM-SLAM performance from any specific detection algorithm. Hence, an ‘oracle localizer’ is used so that the Root Mean Square (RMS) localization error and rates of missing detections and false alarms can be controlled. The sensor path is simulated for 300 waypoints from (36) with $\gamma_0 = \pi$ rad and $\Delta_T = 0.1$ s, with $\sigma_{t,\gamma}^2 = 0.52 \text{ rad}^2$ and $\Sigma_{t,v_v} = 10^{-9} \mathbf{I}_4$. The observer reports are simulated from (38) with default values of $\sigma_{t,w_v} = 5$ m/s and $\sigma_{t,w_\gamma} = 0.02$ rad. The sources are simulated from (39) within a $50 \times 50 \times 3 \text{ m}^3$ surveillance volume with $\mathbf{Q}_t = \text{diag}[10^{-2}, 10^{-2}, 10^{-9}, 10^{-3}]$, survival probability, $p_s = 1$, and initial positions at the centre of three randomly selected quadrants inside the surveillance volume. The height of each feature is randomly drawn from a uniform distribution between $[1.5, 1.95]$ m. Feature detections are simulated from (6) with $\mathbf{R}_{t,m} = \text{diag}[25, 25, 9]$. False alarms are modelled by a Poisson process with a default FAR of $\lambda_c = 0$, and uniformly distributed alarms within the surveillance volume.

SLAM performance in static scenes is typically verified using open-source datasets, providing sensor data and accurate ground truth positions of the moving observer and nearby static features (see, e.g., [9, Table 46.1] for an overview). However, open-source datasets that provide ground truth positions of dynamic features are not currently available.

B. Experimental Description

Experiment 1: Evaluation for increasing RMS sensor report errors in velocity of $\sigma_{t,w_v} = 0, 5, 10$ m/s.

Experiment 2: Evaluation for increasing FARs between $\lambda_c = \{0, 3, 5\}$ with $\sigma_{t,w_v} = 5$ m/s.

Experiment 3: Evaluation for an increasing number of dynamic features, $N_{t,d} = \{0, 1, 3\}$. Dynamic features are simulated [55] for a constant velocity with an arbitrarily chosen example value of 1 m/s. The orientation is simulated as a random walk with process noise standard deviation of 0.35 rad.

C. Implementation Details

The GEM-SLAM realization in Section V-C is evaluated using $I = 100$ particles, and $L_b = 10$ newborn feature components per detection, $N_b = 10^{-4}$ and $\Sigma_b = \text{diag}[2, 2, 1]$. The initial sensor state is assumed known *a priori* with $\Sigma_{0,p} = \text{diag}[0.01, 0.01, 0.0025]$. The sensor pdf is approximated by evaluating the bimodal Kernel Density Estimate (KDE) over the x - and y -coordinates of the particles [56],

$$\hat{p}(\mathbf{r}_t | \mathbf{y}_{1:t}) = \frac{1}{2\pi I} \sum_{i=1}^I e^{-\frac{1}{2}(\mathbf{r}_t - \hat{\mathbf{r}}_t^{(i)})^T \mathbf{H}^{-1} (\mathbf{r}_t - \hat{\mathbf{r}}_t^{(i)})}, \quad (58)$$

where \mathbf{H} is the 2×2 symmetric, positive definite bandwidth.

To mitigate an exponential explosion, the feature GMM is pruned to a maximum of 250 components using [31, Table II] with merging and truncation thresholds, 0.05 and 10^{-7} respectively. Systematic resampling [49] is applied to the sensor particle to avoid particle depletion [57].

The sensor point estimate, $\hat{\mathbf{r}}_t$, is extracted as the weighted average of all particles. The number of features, \tilde{N}_t , is obtained as the weighted average of the number of features across all sensor particles. To extract feature point estimates, all feature GM means are transformed relative to $\hat{\mathbf{r}}_t$. The Estimation-Maximization (EM) algorithm in [58] is used to cluster the feature means into $\tilde{N}_t + 1$ clusters, with one diffuse cluster to absorb outliers [59, Section 10.1].

The GEM-SLAM results are compared against FastSLAM implemented as detailed in [25, Table 3.2]; the SC-PHD filter as per [43, Figure 6,7]; and the RB-PHD filter as in [28, Table 1] for a single-feature map. All algorithms are initialized with $J_0 = 0$, i.e., without knowledge of the initial feature states.

D. Performance Metrics

Accuracy of the sensor estimates, reports, and ground truth is evaluated using the Euclidean distance between two positions, $\mathbf{x}_t \triangleq [x_{t,x}, y_{t,x}, z_{t,x}]^T$ and $\mathbf{y}_t \triangleq [x_{t,y}, y_{t,y}, z_{t,y}]^T$:

$$d^{\text{eucl}}(\mathbf{x}_t, \mathbf{y}_t) = \|\mathbf{x}_t - \mathbf{y}_t\| \quad (59)$$

where $\|\cdot\|$ denotes the Euclidean norm. The OSPA distance [60], [61], $\Delta(\mathbf{X}_t, \mathbf{Y}_t)$, is used as a measure of feature accuracy, continuity, and false track initialization. Given $\mathbf{X}_t \triangleq \{\mathbf{x}_{t,1}, \dots, \mathbf{x}_{t,N}\}$ and $\mathbf{Y}_t \triangleq \{\mathbf{y}_{t,1}, \dots, \mathbf{y}_{t,M}\}$:

$$\Delta(\mathbf{X}_t, \mathbf{Y}_t) \triangleq \left[\frac{1}{N} \min_{\pi \in \Pi_N} \sum_{i=1}^M d_c(\mathbf{x}_{t,i}, \mathbf{y}_{t,\pi(i)})^p + (N - M)c^p \right]^{\frac{1}{p}}$$

where $1 \leq p < \infty$ is the OSPA metric order parameter, Π_N is the set of permutations of length M with elements from $\{1, \dots, N\}$, and $d_c(\mathbf{x}_{t,i}, \mathbf{y}_{t,\pi(i)}) \triangleq \min(c, d(\mathbf{x}_{t,i}, \mathbf{y}_{t,\pi(i)}))$ is the cutoff distance between two tracks with cutoff value, c .

The OSPA parameters were chosen as $c = 30$ m and $p = 2$ to emphasize detection errors as discussed in [62].

VII. RESULTS

A. Experiment 1 - Velocity Error

The average Euclidean distance between the ground truth observer path and GEM-SLAM estimates is shown in Fig. 2(a) for each setting of σ_{t,w_v} . The results are compared against the SC-PHD and RB-PHD filter, FastSLAM, and dead reckoning.

Across all three velocity error settings, GEM-SLAM achieves an average Euclidean distance of between 0.37 m, 0.41 m, and 0.25 m with standard deviation of 0.14 m, 0.17 m, and 0.11 m respectively. In comparison, dead reckoning results in an average localization error of 0.37 m with a standard deviation of 0.14 m for $\sigma_{t,w_v} = 0$ m/s. The error increases to 2.70 m with standard deviation of 1.46 m for $\sigma_{t,w_v} = 5$ m/s, and to 12.86 m with standard deviation of 6.48 m for $\sigma_{t,w_v} = 10$ m/s. GEM-SLAM hence results in observer localization improvements compared to dead reckoning of 84.81% and 98.05% for $\sigma_{t,w_v} = 5, 10$ m/s respectively.

The SC-PHD filter results in average Euclidean distances of 5.23 m for $\sigma_{t,w_v} = 0$ m/s, 5 m for $\sigma_{t,w_v} = 5$ m/s, and 8.48 m for $\sigma_{t,w_v} = 10$ m/s, with standard deviation of 1.95 m, 2.39 m and 4.1 m respectively. The RB-PHD filter results in a Euclidean distance of 5.18 m, 4.91 m, and 8.46 m and standard deviations of 1.92 m, 2.45 m, and 4.08 m respectively for the three settings of σ_{t,w_v} . Therefore, GEM-SLAM results in performance improvements of 92.93%, 91.8%, and 97.05% compared to the SC-PHD filter, as well as 92.85%, 91.64%, and 97.04% compared to the RB-PHD filter for $\sigma_{t,w_v} = 0, 5, 10$ m/s respectively.

The OSPA distances corresponding to the feature mapping accuracy of GEM-SLAM and the benchmark algorithms are shown in Fig. 2(b). The results indicate that GEM-SLAM results in almost constant average OSPA distances of 1 m, 1.09 m, and 0.9 m with standard deviations of 4.04 m, 4.05 m, and 4.06 m for $\sigma_{t,w_v} = 0, 5, 10$ m/s respectively. In comparison, the SC-PHD filter results in OSPA distances of 15.76 m, 15.95 m, and 17.98 m with standard deviations of 8.18 m, 8.20 m, and 7.96 m respectively for the three velocity error settings. The RB-PHD filter corresponds to OSPA distances of 15.89 m, 15.91 m, and 18.23 m, with 8.13 m, 8.12 m, 7.81 m standard deviation respectively. Therefore, GEM-SLAM outperforms the SC-PHD and RB-PHD filters by up to 16.99 m and 17.33 m respectively. These results are clearly correlated with the significant improvement in observer localization.

To provide further insight into the results, Fig. 3 shows the estimated scene map obtained by GEM-SLAM at $t = 100$ for $\sigma_{t,w_v} = 0$ m/s and for $\sigma_{t,w_v} = 10$ m/s. A comparison with the estimated map obtained by the SC-PHD filter is also provided. For the baseline scenario of $\sigma_{t,w_v} = 0$ m/s, the results in Fig. 3(a) show that the GEM-SLAM estimates of the observer trajectory and feature positions accurately reflect the ground truth. In contrast, the map estimates of the SC-PHD filter in Fig. 3(b) diverge from the true observer path, resulting in increasing errors in the feature estimates.

This divergence is attributed to two factors. Firstly, by relying on only the feature information for observer localization, the SC-PHD filter leads to ambiguities in the estimated observer position when no features exist in the map, and can hence not be used to probabilistically anchor the observer. In our

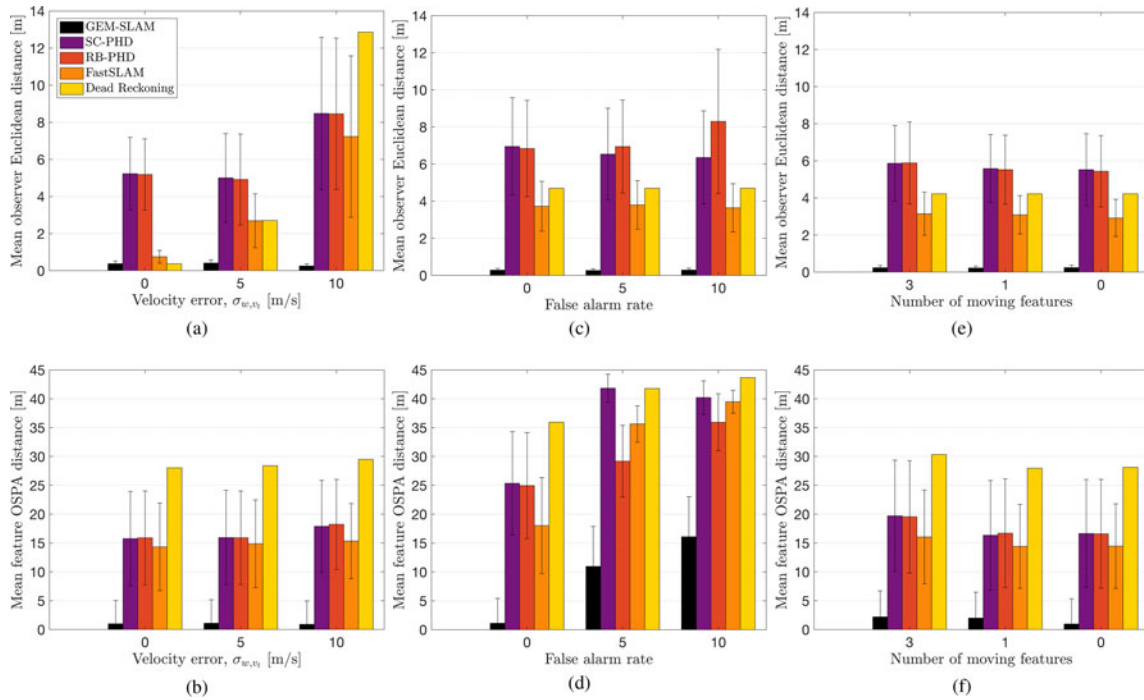


Fig. 2. Performance evaluation of GEM-SLAM and comparison to the SC-PHD and RB-PHD filter, FastSLAM, and dead reckoning. (a),(b) Experiment 1 for increasing reported velocity error; (c), (d) Experiment 2 for increasing FAR; (e),(f) Experiment 3 for an increasing number of moving features. Bars show the distance metrics and whiskers indicate the standard deviation across time. Top row: Average Euclidean distance of the observer state estimates compared to the ground truth across all time steps. Bottom row: Average OSPA distance of the multi-feature states.

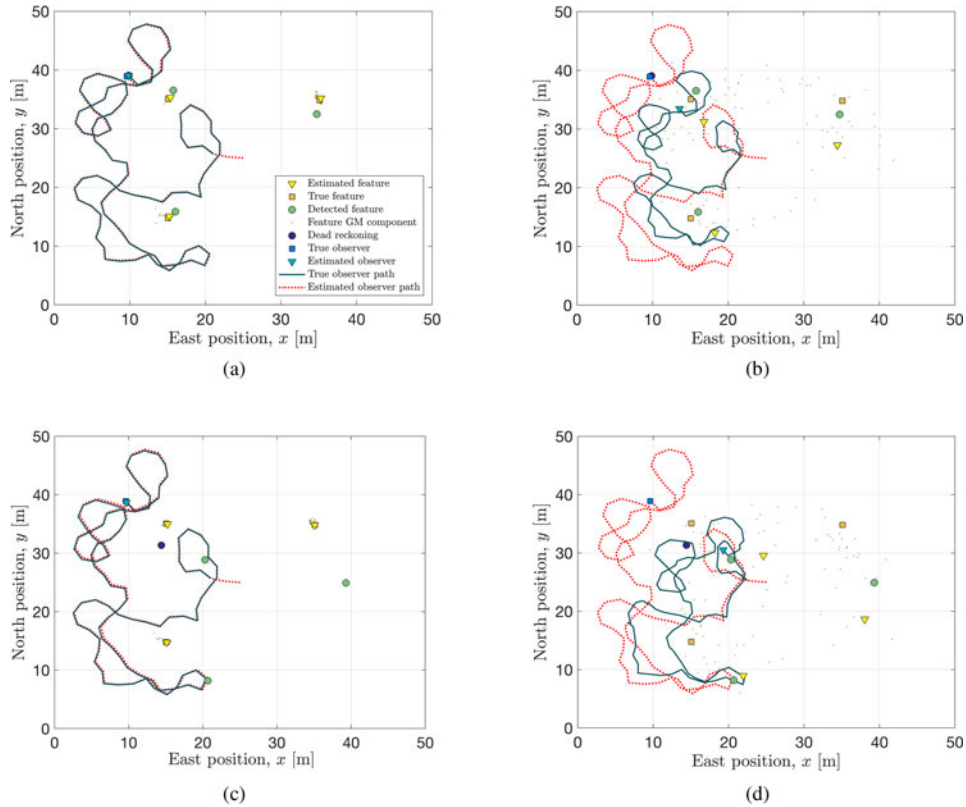


Fig. 3. Experiment 1: Map estimates after $t = 100$ time steps for one Monte Carlo iteration using $\sigma_{t,w_v} = 0$ m/s comparing (a) GEM-SLAM, and (b) the SC-PHD filter, and using $\sigma_{t,w_v} = 10$ m/s for (c) GEM-SLAM and (d) the SC-PHD filter.

experiments, all algorithms are initialized without prior knowledge of the feature positions. Features therefore need to be initialized from the observer-relative detections whilst simultaneously estimating the observer position itself. Ambiguity in the observer position can be resolved by optimally exploiting the observer reports, as demonstrated by the GEM-SLAM results. Secondly, uncertainty in the feature detections leads to uncertainty in observer localization. In the baseline scenario of $\sigma_{t,w_v} = 0$ deg, the observer reports are affected only by the orientation error of $\sigma_{t,w_\gamma} = 1$ deg. Report noise is therefore almost negligible, as indicated by the accurate dead reckoning estimates shown in Fig. 3(a) and 3(b). Nevertheless, the variance of prior importance sampling for the SC-PHD filter is determined by the process noise, i.e., $\sigma_{t,\gamma}^2 = 0.52$ rad², therefore leading to large estimation variance. In contrast, the use of the Rao-Blackwellized particle filter in GEM-SLAM leads to a reduced estimation variance, as both the report and process noise variance are explicitly used for importance sampling. The results are verified for a realistic velocity error of $\sigma_{t,w_v} = 10$ deg in Fig. 3(c) and 3(d).

The results in Fig. 2(a) also compare the GEM-SLAM performance against FastSLAM. The results highlight that FastSLAM results in a observer localization performance of 0.75 m, 2.69 m, and 7.23 m with standard deviation of 0.34 m, 1.45 m, and 4.35 m for $\sigma_{t,w_v} = 0, 5, 10$ m/s respectively. Observer localization of FastSLAM is therefore less robust to report uncertainty than GEM-SLAM. Nevertheless, despite the increasingly inaccurate observer estimates, FastSLAM's performance in feature mapping remains relative constant with OSPA distances of 14.33 m, 14.86 m, and 15.33 m and standard deviations of 7.59 m, 7.6 m, and 6.53 m for the three velocity error settings. The OSPA performance of FastSLAM is therefore comparable to that of the SC-PHD and RB-PHD filter for scenarios where uncertainty is dominated by observer report errors. Nevertheless, GEM-SLAM outperforms FastSLAM by up to 96.54% in observer localization accuracy and 94.13% accuracy in the feature map.

B. Experiment 2 - False Alarms

The results for increasing FARs, as discussed in Section VI-B, are summarized in Fig. 2(c). The results highlight robustness of observer localization for GEM-SLAM with Euclidean distances of 0.28 m, 0.26 m, and 0.29 m with standard deviations of 0.1 m, 0.09 m, and 0.1 m for $\lambda_c = 0, 3, 5$ respectively. As only the FAR is varied in this experiment, dead reckoning results in an observer localization error of 4.69 m with 1.56 m standard deviation across all three settings of λ_c . GEM-SLAM results in an improvement in observer localization of at least 93.8% compared to dead reckoning. The performance improvements of GEM-SLAM compared to the SC-PHD filter correspond to 95.96%, 96.02%, and 95.44%. The performance improvements of GEM-SLAM compared to the RB-PHD filter is between 95.91% and 96.5%, and against FastSLAM between 92.03% and 93.14%.

GEM-SLAM also outperforms the benchmark algorithms in terms of feature mapping accuracy, with OSPA distance of 1.12 m, 10.97 m, and 16.11 m for $\lambda_c = 0, 3, 5$ respectively. These results correspond to an improvement of between 59.95% and 95.58% compared to the SC-PHD filter; between 55.16% and 95.51% compared to the RB-PHD filter; and between 59.20% and 93.78% compared to FastSLAM.

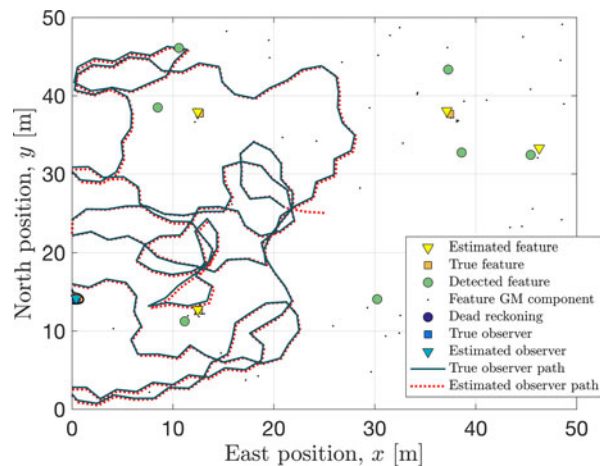


Fig. 4. GEM-SLAM scene map for $\lambda_c = 5$, showing that 7 detections, including 4 false alarms, cause a single false initialization.

The increasing OSPA distance with increasing λ_c is mainly attributed to temporary false feature initializations. To illustrate this point, Fig. 4 shows the estimated scene map at $t = 200$ for a single Monte Carlo run with $\lambda_c = 5$. The results show that from 7 detections, a single feature is falsely initialized at (46.29, 33.29) m. Nevertheless, the features are accurately mapped with estimates at (37.13, 38.08) m, (12.46, 12.75) m, and (12.44, 37.93) m compared to the ground truth at (37.46, 37.65) m, (12.57, 12.68) m, and (12.65, 37.73) m. False feature initialization could be reduced by reducing the number of birth components, J_b , or by decreasing the birth weight, w_b . Either setting reduces the contribution of birth components, thereby reducing the estimated number of features. As a trade-off, reducing the birth weight also results in slower initialization of features. Adverse affects of slow initialization on observer localization can be avoided if the initial positions of some of the features are known *a priori*.

C. Experiment 3 - Number of Dynamic Features

The observer localization performance of GEM-SLAM is compared to the benchmarks in Fig. 2(e), whilst Fig. 2(f) shows the corresponding OSPA results for feature mapping. The results highlight that GEM-SLAM achieves observer localization with accuracy of 0.23 ± 0.01 m on average across all three scenarios. Feature mapping is achieved with 2.2 m OSPA distance for a fully dynamic scene where all three features as well as the observer are moving. For a single moving feature, GEM-SLAM achieves an OSPA distance of 1.99 m, whilst a fully static scene results in 0.97 m feature mapping accuracy.

The SC-PHD and RB-PHD filter result in OSPA distances of 19.71 m and 19.55 m respectively, whilst FastSLAM results in an OSPA distance of 16.06 m for three moving features. A fully static scene results in OSPA distances of 16.65 m for the SC-PHD filter, 16.59 m for the RB-PHD filter, and 14.46 m for FastSLAM. GEM-SLAM therefore results in a feature mapping performance improvement for fully dynamic scenes of 88.83%, 88.74% and 86.30% compared to the SC-PHD filter, RB-PHD filter, and FastSLAM respectively.

VIII. CONCLUSION

This paper proposed a novel approach to SLAM, called GEM-SLAM, that is targeted at 1) dynamic scenes for moving

sensors and moving objects for 2) applications that enable perception beyond vision using sensors not conventionally used for SLAM. GEM-SLAM optimally fuses observer information inferred from feature detections with knowledge from the observer reports. GEM-SLAM was shown to generalize existing PHD-based state-of-the-art approaches, expressing the SC-PHD and RB-PHD filter as special cases. Specifically, we showed that the SC-PHD filter relies only on observer information inferred from the features to localize the observer, and is therefore affected by ambiguities that arise from uncertainty in feature detection. GEM-SLAM mitigates these ambiguities by optimally fusing observer information from the features with the observer reports. Moreover, we showed that the RB-PHD filter is based on a single-feature approximation of the multi-detection evidence. In contrast, GEM-SLAM exploits a closed form solution of the evidence. The proposed approach was evaluated and compared against the performance of the three benchmark algorithms in three experiments. The experiments demonstrated significant performance improvements and robustness against 1) report errors, 2) FAR, and 3) uncertainty due to dynamic features.

APPENDIX A GEM-SLAM DERIVATION

This appendix presents the derivation of the results in Section IV. In the following, the Probability Generating Functional (p.g.fl.) [30] is used as a tool to simplify manipulations of set integrals. The posterior p.g.fl., $G_{\mathcal{X}_t|\mathcal{Z}_t}[T]$, of a RFS, \mathcal{X}_t , with realization, \mathbf{X}_t , given RFS, \mathcal{Z}_t , with realization, \mathbf{Z}_t , is defined as [30]:

$$G_{\mathcal{X}_t|\mathcal{Z}_t}[T] \triangleq \int T^{\mathbf{X}_t} p(\mathbf{X}_t | \mathbf{Z}_{1:t}) \delta \mathbf{X}_t, \quad (\text{A.1})$$

where $T^{\mathbf{X}_t}$ is a test function over the support of \mathbf{X}_t :

$$T^{\mathbf{X}_t} = \begin{cases} 1, & \text{if } \mathbf{X}_t = \emptyset \\ \frac{1}{N_t!} \prod_{\mathbf{x}_t \in \mathbf{X}_t} T(\mathbf{x}_t), & \text{if } \mathbf{X}_t = \{\mathbf{x}_{t,1}, \dots, \mathbf{x}_{t,N_t}\}. \end{cases}$$

The p.g.fl. is related to the pdf via

$$p(\mathbf{X}_t | \mathbf{Z}_{1:t}) = \frac{\delta^n G_{\mathcal{X}_t|\mathcal{Z}_t}[T]}{\delta \mathbf{x}_{t,1} \dots \delta \mathbf{x}_{t,n}} \Big|_{T=0}. \quad (\text{A.2})$$

The PHD can be obtained from a p.g.fl. via the following relation:

$$\lambda(\mathbf{x}_t | \mathbf{Z}_{1:t}) = \frac{\delta G_{\mathcal{X}_t|\mathcal{Z}_t}[T]}{\delta \mathbf{x}_t} \Big|_{T=1}. \quad (\text{A.3})$$

The pdf and PHD can be obtained from the p.g.fl. by differentiating and setting the test function to 0 or 1 respectively [30], [39].

A. GEM Prediction

The predicted p.g.fl. is found by multiplying both sides of (22) by $T^{\mathbf{X}_t}$, integrating over \mathbf{X}_t , and applying (A.1), i.e.,

$$G_{\mathcal{X}_t|\mathcal{Z}_{t-1}}[T] = \iint \prod_{\mathbf{x}_t \in \mathbf{X}_t} T(\mathbf{x}_t) p(\mathbf{X}_t | \mathbf{X}_{t-1}) \times p(\mathbf{X}_{t-1} | \mathbf{Z}_{1:t-1}) \delta \mathbf{X}_t \delta \mathbf{X}_{t-1}. \quad (\text{A.4})$$

Expanding \mathcal{X}_t into \mathbf{r}_t and \mathcal{S}_t :

$$G_{\mathcal{X}_t|\mathcal{Z}_{t-1}}[T] = \iint \left[\int G_{\mathcal{S}_t|\mathcal{S}_{t-1}}[T|\mathbf{r}_t] p(\mathbf{r}_t | \mathbf{r}_{t-1}) d\mathbf{r}_t \right] \times p(\mathcal{S}_{t-1} | \mathbf{r}_t, \mathbf{r}_{t-1}, \mathbf{\Omega}_{1:t-1}) p(\mathbf{r}_{t-1} | \mathbf{y}_{1:t-1}) \delta \mathcal{S}_{t-1} d\mathbf{r}_{t-1}. \quad (\text{A.5})$$

As shown in [30], for any RFS, $\mathcal{F} = F_1 \cup \dots \cup F_N$:

$$G_{\mathcal{F}}[T] = G_{F_1}[T] \dots G_{F_N}[T]. \quad (\text{A.6})$$

Hence, using (4) and (A.6) the p.g.fl., $G_{\mathcal{S}_t|\mathcal{S}_{t-1}}[T|\mathbf{r}_t]$, is:

$$G_{\mathcal{S}_t|\mathcal{S}_{t-1}}[T|\mathbf{r}_t] = G_{\mathcal{S}_t|\emptyset}[T|\mathbf{r}_t] \prod_{n=1}^{N_{t-1}} G_{\mathcal{S}_t|\mathcal{S}_{t-1,n}}[T|\mathbf{r}_t], \quad (\text{A.7})$$

where $G_{\mathcal{S}_t|\emptyset}[T|\mathbf{r}_t]$ is the p.g.fl. of newborn features, $G_{\mathcal{S}_t|\mathcal{S}_{t-1,n}}[T|\mathbf{r}_t]$ is the transition p.g.fl. from $t-1$ to t [30]:

$$G_{\mathcal{S}_t|\mathcal{S}_{t-1,n}}[T|\mathbf{r}_t] = p(\emptyset | \mathbf{s}_{t-1,n}) + \int T(\mathbf{s}_t) p(\mathbf{s}_t | \mathbf{s}_{t-1,n}) d\mathbf{s}_t. \quad (\text{A.8})$$

Using (A.7) and (A.8), the p.g.fl. in (A.5) can be written as

$$G_{\mathcal{X}_t|\mathcal{Z}_{t-1}}[T] = \iint p(\mathbf{r}_t | \mathbf{r}_{t-1}) p(\mathbf{r}_{t-1} | \mathbf{y}_{1:t-1}) \times G_{\mathcal{S}_t|\emptyset}[T|\mathbf{r}_t] G_{\mathcal{S}_{t-1}|\mathcal{O}_{t-1}}[G_{\mathcal{S}_t|\mathcal{S}_{t-1,n}}[T|\mathbf{r}_t]] d\mathbf{r}_t d\mathbf{r}_{t-1}, \quad (\text{A.9})$$

where $G_{\mathcal{S}_{t-1}|\mathcal{O}_{t-1}}[T|\mathbf{r}_t]$ is the multi-feature p.g.fl. at $t-1$ conditional on the RFS of feature detections, \mathcal{O}_{t-1} , with realization, $\mathbf{\Omega}_{t-1}$:

$$G_{\mathcal{S}_{t-1}|\mathcal{O}_{t-1}}[G_{\mathcal{S}_t|\mathcal{S}_{t-1,n}}[T|\mathbf{r}_t]] \triangleq \int \prod_{n=1}^{N_t} (G_{\mathcal{S}_t|\mathcal{S}_{t-1,n}}[T|\mathbf{r}_t]) p(\mathcal{S}_{t-1} | \mathbf{r}_t, \mathbf{r}_{t-1}, \mathbf{\Omega}_{1:t-1}) \delta \mathcal{S}_{t-1}.$$

Applying (A.3) to (A.9), results in (24).

B. GEM Update

Multiplying both sides of (10) by $T^{\mathbf{X}_t}$, and integrating over \mathbf{X}_t , the posterior p.g.fl., $G_{\mathcal{X}_t|\mathcal{Z}_t}[T]$, is found as

$$G_{\mathcal{X}_t|\mathcal{Z}_t}[T] = \frac{H_{\mathcal{Z}_t, \mathcal{X}_t}[T]}{H_{\mathcal{Z}_t, \mathcal{X}_t}[1]} \quad (\text{A.10})$$

where

$$H_{\mathcal{Z}_t, \mathcal{X}_t}[T] \triangleq \int \prod_{\mathbf{x}_t \in \mathbf{X}_t} T(\mathbf{x}_t) p(\mathcal{Z}_t | \mathbf{X}_t) p(\mathbf{X}_t | \mathbf{Z}_{1:t-1}) \delta \mathbf{X}_t.$$

Expanding \mathbf{X}_t by \mathcal{S}_t and \mathbf{r}_t and \mathcal{Z}_t by \mathbf{y}_t and $\mathbf{\Omega}_t$:

$$H_{\mathcal{Z}_t, \mathcal{X}_t}[T] = \int p(\mathbf{y}_t | \mathbf{r}_t) p(\mathbf{r}_t | \mathbf{y}_{1:t-1}) \times \int \left(\prod_{n=1}^{N_t} T(\mathbf{s}_{t,n}, \mathbf{r}_t) \right) p(\mathbf{\Omega}_t | \mathcal{S}_t, \mathbf{r}_t) p(\mathcal{S}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t-1}) \delta \mathcal{S}_t d\mathbf{r}_t, \quad (\text{A.11})$$

where the multi-detection likelihood function, $p(\mathbf{\Omega}_t | \mathbf{S}_t, \mathbf{r}_t)$ is derived using (A.2). The detection p.g.fl. required for the differentiation is obtained using (A.6) and (7), such that

$$G_{\mathcal{O}_t | \mathcal{S}_t} [T'] = G_{\mathcal{O}_t | \emptyset} [T'] \prod_{n=1}^{N_t} G_{\mathcal{O}_t | \mathcal{S}_{t,n}} [T']$$

$$G_{\mathcal{O}_t | \mathcal{S}_{t,n}} [T'] = p(\emptyset | \mathbf{r}_t, \mathbf{s}_{t,n}) + \int T'(\boldsymbol{\omega}_t) p(\boldsymbol{\omega}_t | \mathbf{r}_t, \mathbf{s}_{t,n}) d\boldsymbol{\omega}_t. \quad (\text{A.12})$$

Inserting (A.12) into (A.11) hence leads to:

$$H_{\mathcal{Z}_t, \mathcal{X}_t} [T] = \frac{\delta^{M_t}}{\delta \mathbf{\Omega}_t} F_{\mathcal{Z}_t, \mathcal{X}_t} [T], \quad (\text{A.13})$$

where

$$F_{\mathcal{Z}_t, \mathcal{X}_t} [T] = \int \hat{F}_{\mathcal{O}_t, \mathcal{S}_t} [T] p(\mathbf{y}_t | \mathbf{r}_t) p(\mathbf{r}_t | \mathbf{y}_{1:t-1}) d\mathbf{r}_t$$

$$\hat{F}_{\mathcal{O}_t, \mathcal{S}_t} [T] \triangleq G_{\mathcal{O}_t | \emptyset} [T'] G_{\mathcal{S}_t | \mathcal{O}_t} [T G_{\mathcal{O}_t | \mathcal{S}_{t,n}} [T']] \Big|_{T'=0}.$$

Using (A.3), (A.10) and (A.13):

$$\lambda(\mathbf{x}_t | \mathbf{Z}_t) = \frac{\delta}{\delta \mathbf{x}_t} \frac{\delta^{M_t}}{\delta \mathbf{\Omega}_t} F_{\mathcal{Z}_t, \mathcal{X}_t} [T] \Big|_{T=1} \cdot \quad (\text{A.14})$$

The numerator is obtained as

$$\frac{\delta}{\delta \mathbf{x}_t} \frac{\delta^{M_t}}{\delta \mathbf{\Omega}_t} F_{\mathcal{Z}_t, \mathcal{X}_t} [T]$$

$$= \frac{d}{d\mathbf{r}_t} \int \frac{\delta^{M_t}}{\delta \mathbf{\Omega}_t} \frac{\delta}{\delta \mathbf{s}_t} \hat{F}_{\mathcal{O}_t, \mathcal{S}_t} [T] p(\mathbf{y}_t | \mathbf{r}_t) p(\mathbf{r}_t | \mathbf{y}_{1:t-1}) d\mathbf{r}_t$$

$$= \frac{\delta^{M_t}}{\delta \mathbf{\Omega}_t} \frac{\delta}{\delta \mathbf{s}_t} \hat{F}_{\mathcal{O}_t, \mathcal{S}_t} [T] p(\mathbf{y}_t | \mathbf{r}_t) p(\mathbf{r}_t | \mathbf{y}_{1:t-1}). \quad (\text{A.15})$$

The partial derivative, $\frac{\delta^{M_t}}{\delta \mathbf{\Omega}_t} \frac{\delta}{\delta \mathbf{s}_t} \hat{F}_{\mathcal{O}_t, \mathcal{S}_t} [1]$, was derived in [30]:

$$\frac{\delta^{M_t}}{\delta \mathbf{\Omega}_t} \frac{d}{d\mathbf{s}_t} \hat{F}_{\mathcal{O}_t, \mathcal{S}_t} [1] = \mathcal{L}(\mathbf{\Omega}_t | \mathbf{r}_t) \lambda(\mathbf{s}_t | \mathbf{r}_t, \mathbf{Z}_{1:t}) \quad (\text{A.16})$$

with $\mathcal{L}(\mathbf{\Omega}_t | \mathbf{r}_t)$ as defined in (30). Inserting (A.16) into (A.15):

$$\frac{\delta}{\delta \mathbf{x}_t} \frac{\delta^{M_t}}{\delta \mathbf{\Omega}_t} F_{\mathcal{Z}_t, \mathcal{X}_t} [T] \Big|_{T=1}$$

$$= \mathcal{L}(\mathbf{\Omega}_t | \mathbf{r}_t) \lambda(\mathbf{s}_t | \mathbf{r}_t, \mathbf{Z}_{1:t}) p(\mathbf{y}_t | \mathbf{r}_t) p(\mathbf{r}_t | \mathbf{y}_{1:t-1}). \quad (\text{A.17})$$

The denominator in (A.14) is obtained in a similar way, i.e.,

$$\frac{\delta^{M_t}}{\delta \mathbf{\Omega}_t} F_{\mathcal{Z}_t, \mathcal{X}_t} [1] = \int \mathcal{L}(\mathbf{\Omega}_t | \mathbf{r}_t) p(\mathbf{y}_t | \mathbf{r}_t) p(\mathbf{r}_t | \mathbf{y}_{1:t-1}) d\mathbf{r}_t \quad (\text{A.18})$$

Inserting (A.17) and (A.18) into (A.14) yields (29).

APPENDIX B RB-PHD REDERIVED

The multi-detection evidence can be expressed as:

$$p(\mathbf{\Omega}_t | \mathbf{r}_t) = \int p(\mathbf{\Omega}_t | \mathbf{r}_t, \mathbf{S}_t) p(\mathbf{S}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t-1}) \delta \mathbf{S}_t \quad (\text{B.1})$$

Alternatively, $p(\mathbf{\Omega}_t | \mathbf{S}_t, \mathbf{r}_t)$ can be expressed as [63]

$$p(\mathbf{\Omega}_t | \mathbf{S}_t, \mathbf{r}_t) = \sum_{\mathbf{W} \subseteq \mathbf{X}_t} p(\mathbf{W} | \mathbf{S}_t, \mathbf{r}_t) p(\mathbf{\Omega}_t - \mathbf{W} | \emptyset, \mathbf{r}_t).$$

Applying (13) to $p(\mathbf{\Omega}_t - \mathbf{W} | \emptyset, \mathbf{r}_t)$ and with IID detections:

$$p(\mathbf{\Omega}_t | \mathbf{S}_t, \mathbf{r}_t) \quad (\text{B.2})$$

$$= e^{-N_{t,c}} \sum_{\mathbf{W} \subseteq \mathbf{X}_t} \prod_{\mathbf{w}_t \in \mathbf{W}} p(\mathbf{w}_t | \mathbf{S}_t, \mathbf{r}_t) \prod_{\mathbf{z}_t \in (\mathbf{\Omega}_t - \mathbf{W})} \kappa(\mathbf{z}_t | \mathbf{r}_t),$$

where $\kappa(\mathbf{z}_t | \mathbf{r}_t) \triangleq p(\mathbf{z}_t | \emptyset, \mathbf{r}_t)$. Inserting (B.2) into (B.1):

$$p(\mathbf{\Omega}_t | \mathbf{r}_t) = e^{-N_{t,c}} \sum_{\mathbf{W} \subseteq \mathbf{X}_t} \prod_{\mathbf{z}_t \in (\mathbf{\Omega}_t - \mathbf{W})} \kappa(\mathbf{z}_t | \mathbf{r}_t)$$

$$\times \int \prod_{\mathbf{w}_t \in \mathbf{W}} p(\mathbf{w}_t | \mathbf{S}_t, \mathbf{r}_t) p(\mathbf{S}_t | \mathbf{r}_t, \mathbf{\Omega}_{1:t-1}) \delta \mathbf{S}_t. \quad (\text{B.3})$$

Applying (12) to (B.3), and reducing the resulting sum to a single feature and single detection as in [29] results in (34).

APPENDIX C NOMENCLATURE

t	Time index
Δ_t	Time delay between $t-1$ and t
M_t	Number of feature detections
$N_t, N_{t t-1}$	Updated / predicted number of features
$N_{t,c}$	Number of false alarms
\mathbf{r}_t	Observer state at t
\mathbf{y}_t	Observer reports at t
\mathcal{S}_t	Feature RFS at t
\mathbf{S}_t	Realization of \mathcal{S}_t
$\mathbf{s}_{t,n}^a$	Single-feature state in global reference frame
$\mathbf{s}_{t,n}$	Single-feature state relative to \mathbf{r}_t
$p_b(\cdot)$	Probability of feature birth
$p_d(\cdot)$	Probability of feature detection
$p_s(\cdot)$	Probability of feature survival
\mathcal{O}_t	Feature detection RFS
$\mathbf{\Omega}_t$	Realization of \mathcal{O}_t
$\boldsymbol{\omega}_{t,m}$	Single-feature detection
\mathcal{X}_t	SLAM RFS
\mathbf{X}_t	Realization of \mathcal{X}_t
\mathcal{Z}_t	SLAM measurement RFS
\mathbf{Z}_t	Realization of \mathcal{Z}_t
$p(\cdot)$	Probability density function
$\lambda(\cdot)$	Probability hypothesis density
$\lambda_b(\cdot)$	PHD of newborn features
$\lambda_m(\cdot)$	PHD of surviving features
$\lambda_d(\cdot)$	PHD of detected features
$G(\cdot)$	Probability Generating Functional
T	Test function
$\ell(\cdot)$	Single-detection evidence
$\mathcal{L}(\cdot)$	Multi-detection evidence
$\kappa(\cdot)$	False alarm likelihood
λ_c	Maximum number of false alarms
$f(\cdot)$	Dynamical model of observer state
$h(\cdot)$	Observer report model
$d(\cdot)$	Dynamical model single-source state
$\bar{g}(\cdot)$	Feature-to-detection transformation
$P(\cdot)$	Feature persistence process

B_t	Feature birth process
K_t	False alarm process
$D(\cdot)$	Feature detection process
\mathbf{u}_t	Control input
\mathbf{v}_t	Process noise on dynamical model of observer state
\mathbf{w}_t	Observer report noise
$\mathcal{N}(\cdot)$	Normal distribution
$\mathcal{N}^w(\cdot)$	Wrapped normal distribution
$\vartheta(\cdot)$	Wrapping operator
$\mathcal{U}(\cdot)$	Uniform distribution
$\int \cdot d\mathbf{w}$	Integral over random variable, \mathbf{w}
$\int \cdot \delta \mathbf{W}$	Set integral over set \mathbf{W}
k_t	Detection-to-feature assignment
\mathbf{p}_t	Linear subspace of observer state
$\tilde{\mathbf{r}}_t$	3D observer position
γ_t	Observer orientation
v_t	Observer speed
$y_{t,v}$	Reported observer speed
$y_{t,\gamma}$	Reported observer orientation
\mathbf{F}_t	Observer dynamical model
\mathbf{h}	Observer report model
$\mathbf{v}_{t,\mathbf{v}}$	Process noise on \mathbf{p}_t
$\Sigma_{t,\mathbf{v},\mathbf{v}}$	Covariance of $\mathbf{v}_{t,\mathbf{v}}$
$\sigma_{t,v}^2$	Velocity component of $\Sigma_{t,\mathbf{v},\mathbf{v}}$
$v_{t,\gamma}$	Process noise on γ_t
$\sigma_{t,\gamma}^2$	Variance of $v_{t,\gamma}$
$w_{t,v}$	Report noise on v_t
σ_{t,w_v}^2	Variance of $w_{t,v}$
$w_{t,\gamma}$	Report noise on γ_t
σ_{t,w_γ}^2	Variance of $w_{t,\gamma}$
\mathbf{D}_t	Specific feature dynamical model
$\mathbf{n}_{t,n}$	Feature process noise
\mathbf{Q}_t	Covariance of $\mathbf{n}_{t,n}$
$g(\cdot)$	Cartesian-to-spherical transformation
$\mathbf{e}_{t,m}$	Feature detection error
$\mathbf{R}_{t,m}$	Covariance of $\mathbf{e}_{t,m}$
$\mathbf{\Gamma}(\cdot)$	Rotation matrix
J_t	Number of feature GM components
J_b	Number of feature birth GM components
$w_t, w_{t t-1}$	Updated / predicted GM weight
$w_{b,t}$	GM weight of birth component
$\mathbf{m}_t, \mathbf{m}_{t t-1}$	Updated / predicted GM mean
$\Sigma_t, \Sigma_{t t-1}$	Updated / predicted GM covariance
$\mathbf{m}_{b,t}$	GM mean of birth component
$\Sigma_{b,t}$	GM covariance of birth component
\mathbf{K}_t	Kalman gain
\mathbf{S}_t	Kalman innovation covariance
\mathbf{G}_t	Measurement Jacobian
I	Number of observer particles
$\alpha_t, \hat{\alpha}_t$	Normalized / unnormalized importance weight
$\hat{\gamma}_t$	Observer orientation particle
$\mu_{t,\gamma}$	Mean of proposal distribution for γ_t
$\varsigma_{t,\gamma}$	Variance of proposal distribution for γ_t
$\boldsymbol{\mu}_{t,\mathbf{p}}$	Mean of posterior pdf of \mathbf{p}_t
$\varsigma_{t,\mathbf{p}}$	Covariance of posterior pdf of \mathbf{p}_t

REFERENCES

- [1] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Boston, MA, USA: Artech House, 2004.
- [2] D. J. Salmond, "Tracking in uncertain environments," Univ. Sussex, Brighton, U.K., Roy. Aerosp. Establishment, Farnborough, U.K., Tech. Memo. AW 121, 1989.
- [3] S. Blackman, *Multiple Target Tracking With RADAR Applications*. Norwood, MA, USA: Artech House, 1986.
- [4] L. George and A. Mazel, "Humanoid robot indoor navigation based on 2D bar codes: Application to the NAO robot," in *Proc. IEEE-RAS Int. Conf. Human. Robots*, Oct. 2013, pp. 329–335.
- [5] C. Fischer and H. Gellersen, "Location and navigation support for emergency responders: A survey," *IEEE Pervasive Comput.*, vol. 9, no. 1, pp. 38–47, Jan. 2010.
- [6] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robot. Autom. Mag.*, vol. 13, no. 2, pp. 99–110, 2006.
- [7] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping: Part II," *IEEE Robot. Autom. Mag.*, vol. 13, no. 3, pp. 108–117, Sep. 2006.
- [8] G. Dissanayake, S. Huang, Z. Wang, and R. Ransinghe, "A review of recent developments in simultaneous localization and mapping," in *Proc. Int. Conf. Ind. Inf. Syst.*, Aug. 2011, pp. 477–482.
- [9] C. Stachniss, J. J. Leonard, and S. Thrun, *Simultaneous Localization and Mapping*. Berlin, Germany: Springer, 2016, pp. 1153–1176.
- [10] C. Cadena *et al.*, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1309–1332, Dec. 2016.
- [11] C. Evers, A. H. Moore, and P. A. Naylor, "Acoustic simultaneous localization and mapping (a-SLAM) of a moving microphone array and its surrounding speakers," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Shanghai, China, Mar. 2016, pp. 6–10.
- [12] C. Evers, A. H. Moore, and P. A. Naylor, "Localization of moving microphone arrays from moving sound sources for robot audition," in *Proc. Eur. Signal Process. Conf.*, Budapest, Hungary, 2016, pp. 1008–1012.
- [13] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, Jun. 2007.
- [14] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments," *Int. J. Robot. Res.*, vol. 31, no. 5, pp. 647–663, 2012.
- [15] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *Int. J. Robot. Res.*, vol. 34, no. 3, pp. 314–334, 2015.
- [16] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. Cambridge, MA, USA: MIT Press, 2005.
- [17] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.
- [18] S. Lowry *et al.*, "Visual place recognition: A survey," *IEEE Trans. Robot.*, vol. 32, no. 1, pp. 1–19, Feb. 2016.
- [19] G. Grisetti, C. Stachniss, and W. Burgard, "Nonlinear constraint network optimization for efficient map learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 3, pp. 428–439, Sep. 2009.
- [20] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," *IEEE Trans. Robot.*, vol. 24, no. 6, pp. 1365–1378, Dec. 2008.
- [21] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *Int. J. Robot. Res.*, vol. 31, no. 2, pp. 216–235, 2012.
- [22] F. Lu and E. Milios, "Globally consistent range scan alignment for environment mapping," *Auton. Robots*, vol. 4, no. 4, pp. 333–349, Oct. 1997.
- [23] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "Fast-SLAM: A factored solution to the simultaneous localization and mapping problem," in *Proc. AAAI Nat. Conf. Artif. Intell.*, 2002, pp. 593–598.
- [24] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges," in *Proc. Int. Joint Conf. Artif. Intell.*, Acapulco, Mexico, 2003, pp. 1151–1156.
- [25] M. Montemerlo and S. Thrun, *FastSLAM* (Ser. Springer Tracts in Advanced Robotics). vol. 27, Berlin, Germany: Springer, 2007.
- [26] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice* (Ser. Statistics for Engineering and Information Science). New York, NY, USA: Springer, 2001.

- [27] B. Siciliano and O. Khatib, *Springer Handbook of Robotics*. Berlin, Germany: Springer-Verlag, 2008.
- [28] J. S. Mullane, B.-N. Vo, M. D. Adams, and B.-T. Vo, *Random Finite Sets for Robot Mapping and SLAM (Ser. Springer Tracts in Advanced Robotics)*, vol. 72. Berlin, Germany: Springer, 2011.
- [29] J. S. Mullane, B.-N. Vo, M. D. Adams, and B.-T. Vo, "A random-finite-set approach to Bayesian SLAM," *IEEE Trans. Robot. Autom.*, vol. 27, no. 2, pp. 268–282, Apr. 2011.
- [30] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.
- [31] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4091–4104, Nov. 2006.
- [32] N. Atanasov, M. Zhu, K. Daniilidis, and G. J. Pappas, "Localization from semantic observations via the matrix permanent," *Int. J. Robot. Res.*, vol. 35, nos. 1–3, pp. 73–99, 2016.
- [33] H. Deusch, S. Reuter, and K. Dietmayer, "The labeled multi-Bernoulli SLAM filter," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1561–1565, Oct. 2015.
- [34] S. Reuter, B. T. Vo, B. N. Vo, and K. Dietmayer, "The labeled multi-Bernoulli filter," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3246–3260, Jun. 2014.
- [35] A. Swain and D. E. Clark, "The single-group PHD filter: An analytic solution," in *Proc. Conf. Inf. Fusion*, Jul. 2011, pp. 1–8.
- [36] C. S. Lee, S. Nagappa, N. Palomeras, D. E. Clark, and J. Salvi, "SLAM with SC-PHD filters: An underwater vehicle application," *IEEE Robot. Autom. Mag.*, vol. 21, no. 2, pp. 38–45, Jun. 2014.
- [37] J. Housineau, D. E. Clark, S. Ivekovic, C. S. Lee, and J. Franco, "A unified approach for multi-object triangulation, tracking and camera calibration," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2934–2948, Jun. 2016.
- [38] I. Schlangen *et al.*, "Marker-less stage drift correction in super-resolution microscopy using the single-cluster PHD filter," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 1, pp. 193–202, Feb. 2016.
- [39] R. P. S. Mahler, "Statistics' 101 for multisensor, multitarget data fusion," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 19, no. 1, pp. 53–64, Jan. 2004.
- [40] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes (vol. I: Elementary Theory and Methods)*. Berlin, Germany: Springer, 2003.
- [41] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes (vol. II: General Theory and Structure)*. Berlin, Germany: Springer, 2007.
- [42] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York, NY, USA: McGraw-Hill, Inc., 1991.
- [43] C.-S. Lee, D. E. Clark, and J. Salvi, "SLAM with dynamic target via single-cluster PHD filtering," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 3, pp. 543–552, Jun. 2013.
- [44] I. J. Cox, "A review of statistical data association techniques for motion correspondence," *Int. J. Comput. Vis.*, vol. 10, no. 1, pp. 53–66, 1993.
- [45] M. Montemerlo and S. Thrun, "Simultaneous localization and mapping with unknown data association using FastSLAM," in *Proc. Int. Conf. Robot. Autom.*, vol. 2, Sep. 2003, pp. 1985–1991.
- [46] S. Gannot and A. Yeredor, "The Kalman filter," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Berlin, Germany: Springer-Verlag, 2008, ch. 8, pt. B.
- [47] B. Ristic, D. E. Clark, B.-N. Vo, and B.-T. Vo, "Adaptive target birth intensity for PHD and CPHD filters," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 2, pp. 1656–1668, Apr. 2012.
- [48] K. V. Mardia and P. E. Jupp, *Directional Statistics*. Hoboken, NJ, USA: Wiley-Blackwell, 1999.
- [49] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.
- [50] J. Traa and P. Smaragdus, "A wrapped Kalman filter for azimuthal speaker tracking," *IEEE Signal Process. Lett.*, vol. 20, no. 12, pp. 1257–1260, Dec. 2013.
- [51] D. J. Salmund, "Mixture reduction algorithms for point and extended object tracking in clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 45, no. 2, pp. 667–686, Apr. 2009.
- [52] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statist. Comput.*, vol. 10, no. 3, pp. 197–208, 2000.
- [53] M. D. Adams, B.-N. Vo, R. P. S. Mahler, and J. S. Mullane, "SLAM gets a PHD: New concepts in map estimation," *IEEE Robot. Autom. Mag.*, vol. 21, no. 2, pp. 26–37, Jun. 2014.
- [54] C.-S. Lee, "Simultaneous localization and mapping using single cluster probability hypothesis density filters," Ph.D. dissertation, Universitat de Girona, Girona, Spain, 2014.
- [55] X.-R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. Part I: Dynamic models," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1333–1364, Oct. 2003.
- [56] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," *Annals Statist.*, vol. 38, no. 5, pp. 2916–2957, 2010.
- [57] J. Vermaak, A. Doucet, and P. Perez, "Maintaining multimodality through mixture tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2003, vol. 2, pp. 1110–1116.
- [58] I. D. Gebru, X. Alameda-Pineda, F. Forbes, and R. Horaud, "EM algorithms for weighted-data clustering with application to audio-visual scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 12, pp. 2402–2415, Dec. 2016.
- [59] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer, 2006.
- [60] B. Ristic, *Particle Filters for Random Set Models*, 1st ed. New York, NY, USA: Springer-Verlag, 2013.
- [61] B. Ristic, B.-N. Vo, and B.-T. Vo, "A metric for performance evaluation of multi-target tracking algorithms," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3452–3457, Jul. 2011.
- [62] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3447–3457, Aug. 2008.
- [63] R. P. S. Mahler, "Multitarget Markov motion models," *Proc. SPIE*, vol. 3720, pp. 47–58, 1999.



Christine Evers (SM'16) received the B.Sc. degree in electrical engineering and computer science from Jacobs University Bremen, Bremen, Germany, in 2005, and the M.Sc. degree in signal processing and communications and the Ph.D. degree from the University of Edinburgh, Edinburgh, U.K., in 2006 and 2010, respectively. From 2009 to 2010, she was a Research Fellow at the University of Edinburgh. From 2010 to 2014, she was as a Senior Systems Engineer in Selex ES, Edinburgh, U.K. In 2014, she joined Imperial College London, London, U.K., as a

Research Associate. She is currently an EPSRC Research Fellow in the Department of Electrical and Electronic Engineering, Imperial College London. In 2017, she was awarded a fellowship by the U.K. Engineering and Physical Research Council to advance her research on acoustic signal processing and scene mapping for socially assistive robots. Her research interests include statistical signal processing for speech and audio applications, including sound source localization and tracking, acoustic simultaneous localization and mapping for robot audition, blind speech dereverberation, and sensor fusion. She is member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing.



Patrick A. Naylor (SM'07) received the B.Eng. degree in electronic and electrical engineering from the University of Sheffield, Sheffield, U.K., and the Ph.D. degree from Imperial College London, London, U.K. His research interests include speech, audio, and acoustic signal processing. He has worked in particular on adaptive signal processing for speech dereverberation, blind multichannel system identification and equalization, acoustic echo control, speech quality estimation and classification, single and multichannel speech enhancement, and speech production modeling with particular focus on the analysis of the voice source signal. In addition to his academic research, he enjoys several fruitful links with industry in the U.K., USA, and Europe. He is a member of Academic Staff in the Department of Electrical and Electronic Engineering, Imperial College London. He is the Past-Chair of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing and the Director of the European Association for Signal Processing. He was an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS and a Senior Area Editor of the IEEE TRANSACTIONS ON AUDIO SPEECH AND LANGUAGE PROCESSING.

tion modeling with particular focus on the analysis of the voice source signal. In addition to his academic research, he enjoys several fruitful links with industry in the U.K., USA, and Europe. He is a member of Academic Staff in the Department of Electrical and Electronic Engineering, Imperial College London. He is the Past-Chair of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing and the Director of the European Association for Signal Processing. He was an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS and a Senior Area Editor of the IEEE TRANSACTIONS ON AUDIO SPEECH AND LANGUAGE PROCESSING.