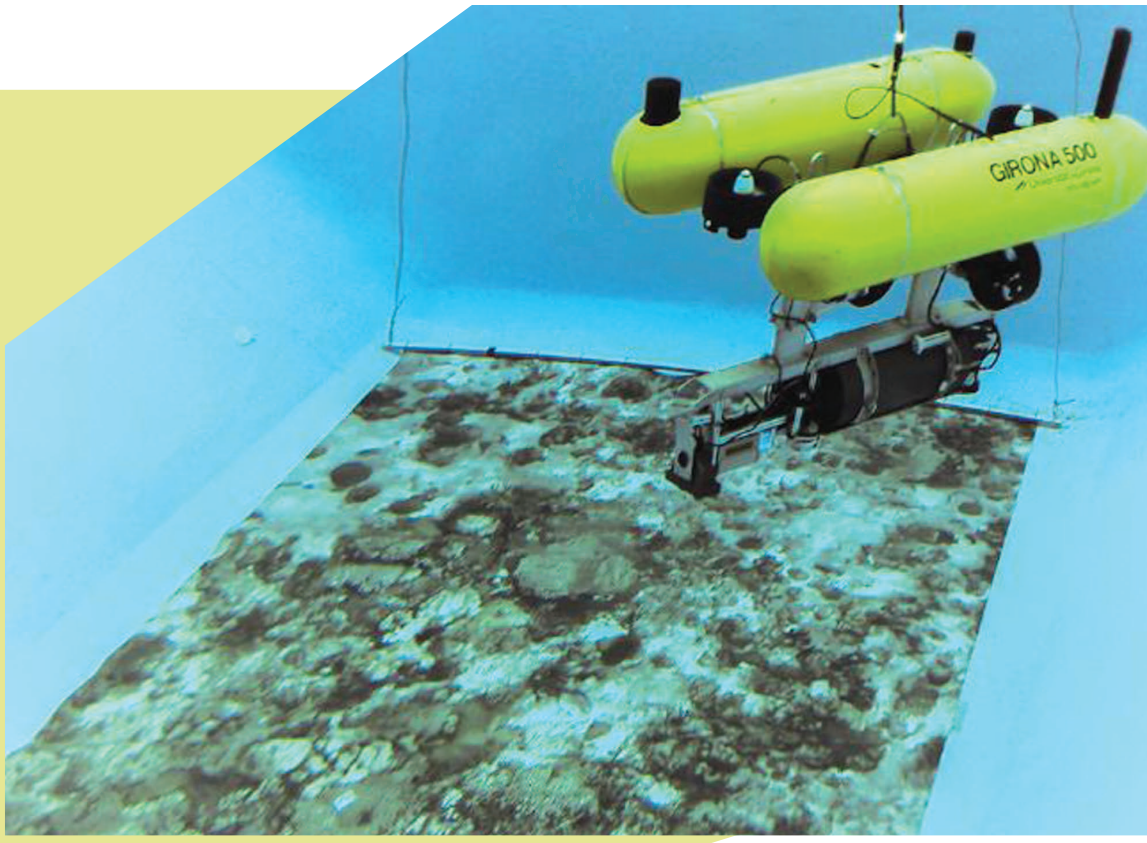


SLAM with SC-PHD Filters

An Underwater Vehicle Application

By Chee Sing Lee, Sharad Nagappa, Narcis Palomeras, Daniel E. Clark, and Joaquim Salvi

Digital Object Identifier 10.1109/MRA.2014.2310132
Date of publication: 7 May 2014

The random finite-set formulation for multiobject estimation provides a means of estimating the number of objects in cluttered environments with missed detections within a unified probabilistic framework. This methodology is now becoming the dominant mathematical framework within the sensor fusion community for developing multiple-target tracking algorithms. These techniques are also gaining traction in the field of feature-based simultaneous localization and mapping (SLAM) for mobile robotics. Here, we present one such instance of this approach with an underwater vehicle using a hierarchical multiobject estimation method for estimating both landmarks and vehicle position.

Underwater Robotics and SLAM

The subsea industry is increasingly interested in the use of autonomous underwater vehicles (AUVs) to perform inspection, maintenance, and light intervention tasks at submarine facilities. Of particular interest is the ability to have vehicles operating unattended for extended periods of time, which is a key to reducing operating costs. A critical issue in this is the vehicle's awareness of its

1070-9932/14/$31.00©2014IEEE

environment as well as its place within that environment, i.e., SLAM. To perform path-finding operations, surveys, or interventions, the vehicle must have a map of its surroundings as well as a firm knowledge of its own position relative to the local area. It is for these reasons that SLAM is widely described as a fundamental problem in mobile robotics.

The localization of AUVs is a particularly challenging area of research. The use of global positioning systems is limited to the surface of the water due to high levels of attenuation. An attitude and heading reference system is used to estimate the orientation of the vehicle, and the position is estimated using a dead reckoning algorithm with measurements from a pressure sensor and a Doppler velocity log (DVL). The DVL measures velocity with respect to the water or the sea bottom. It is possible to perform corrections due to drift in the sensors or noisy measurements by incorporating this estimation into a SLAM framework.

Distilled to its core, SLAM is a state-estimation problem: given a sequence of noisy sensor readings and odometry inputs, what is the most likely configuration of the robot's trajectory and environment? Early breakthroughs in SLAM were provided using the extended Kalman filter (EKF) to address this as a joint state estimation problem, coupling the positions of the vehicle and landmarks [1], [2]. In the decades following this foundational work, SLAM has grown into a vibrant field of research, with a great diversity of algorithms and application areas. Of particular note are unscented Kalman filter SLAM [3] as well as the widely acclaimed particle-based FastSLAM family of algorithms [4], [5].

More recently, advances in set theoretic approaches have given rise to a new class of algorithms for multiple-object estimation based on random finite sets (RFSs). This article examines a vein of research that aims to bridge the fields of SLAM and multiobject estimation by introducing these methods to robotic applications. Already, these algorithms are gaining traction with researchers in the area of SLAM due to their ability to implicitly model the data association of landmarks and deal with high rates of false positives/negatives [6], [7]. Last year's IEEE Conference on Robotics and Automation included a workshop dedicated to the use of RFS methods in SLAM [8]–[10]. The work presented in this article is, to the best of our knowledge, the first application of RFS techniques in underwater robotics.

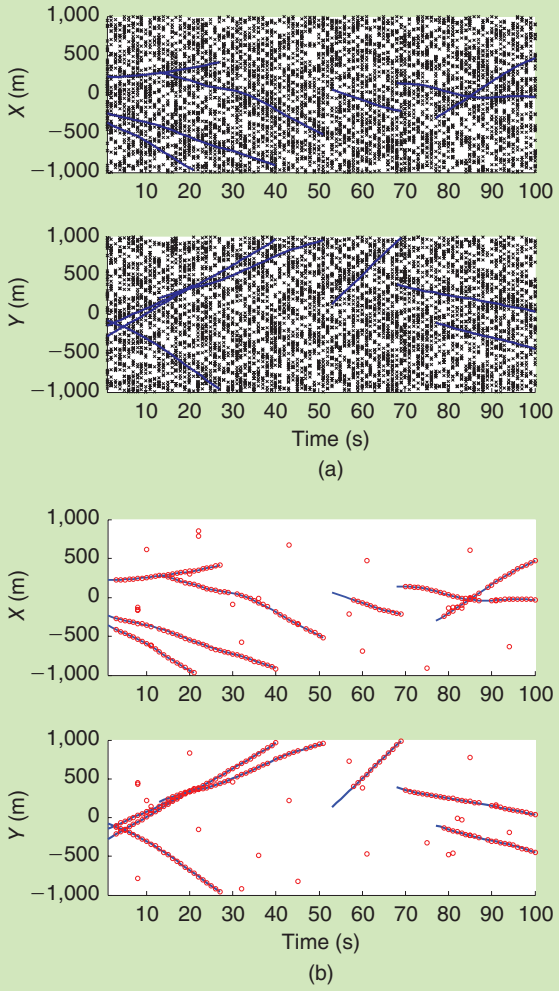


Figure 1. The tracking targets in clutter with the PHD filter. (a) A two-dimensional tracking scenario where targets move according to a constant-velocity model (blue solid lines), with new targets appearing randomly during the course of the scenario. At each time step, the number of clutter measurements is an order of magnitude greater than the number of true targets (black crosses). (b) The PHD filter output (red circles) plotted against ground truth trajectories (solid lines). Despite the high amount of measurement clutter, only a few false tracks are created, all of which are quite transient.

Probability Hypothesis Density Filters and RFSs

Prompted by the plethora of heuristic approximations to multitarget filtering, the probability hypothesis density (PHD) filter was derived as a principled approximation of the multitarget Bayes filter [11], [12]. It has been demonstrated to perform well under difficult measurement conditions characterized by 1) inaccurate measurements (noise), 2) failure to detect targets (missed detections), and 3) false detections (clutter). An example of what can be achieved is shown in Figure 1.

The traditional approach to multitarget filtering relies on using either a multitarget state-space or multiple single-target filters. Both approaches require the data-association problem to be addressed as well as complex heuristics to deal with missed detections and addition/removal of targets.

In contrast, the PHD filter adopts a flexible RFS representation for the multitarget state. Given a target space $\mathcal{X} \in \mathbb{R}^d$, an RFS on $\mathcal{X}$ is an unordered set of random vectors

$$\{x_1, \cdots x_n \in \mathcal{X}\},$$

where the cardinality of the set, n, is itself a random quantity. Adopting this representation, we can define a more sophisticated measurement model to handle the previously discussed measurement difficulties. For a particular target state x_i, let the probability of detection be $p_D(x_i)$, then the generated measurement is modeled as an RFS $\Upsilon(x_i)$ with the following probability density function:

$$\Upsilon(x_i) = \begin{cases} \phi & \text{with probability } 1 - p_D(x_i) \\ \{z_i\} & \text{with probability } p_D(x_i)g(z_i \mid x_i), \end{cases} \quad (1)$$

where $g(z_i \mid x_i)$ is the measurement likelihood defined by the sensor's noise characteristics. In addition, we define an RFS C, which constitutes the clutter measurements. Typically, we define the clutter to be a Poisson-distributed process. The entire measurement set is, thus, an RFS defined as

$$Z = \Upsilon(x_1) \cup \cdots \cup \Upsilon(x_i) \cup C. \quad (2)$$

With this measurement model in hand, we can apply the tools of multitarget calculus to derive a multiobject measurement likelihood, which weighs each detection against the likelihood of a missed detection or clutter in a principled fashion. This likelihood can be used to formulate a multiobject Bayes filter, but such a filter would be very difficult to implement for all but the most trivial scenarios. This is because it would have to propagate a multiobject probability density function, which is defined over all possible configurations for every cardinality of objects. The PHD filter is an approximation of this full multiobject filter, which propagates the first moment of the multiobject density function. This moment is known variously as the PHD or intensity function. The PHD can be roughly conceived as the expected value of an RFS. It is a nonnegative function on $\mathcal{X}$, where locations with high values correspond to likely locations of the objects. It also encodes the likely number of objects; the integral of the PHD over a certain region is the expected number of objects in that region.

SC-PHD Filter

The derivation of the first moment filter involves an assumption on the nature of the prior distribution. For example, the PHD filter assumes that the prior is distributed according to a Poisson process, and the cardinalized PHD filter assumes an independently and identically distributed process. The single-cluster PHD (SC-PHD) filter used for SLAM assumes a single-cluster process. A cluster process is a hierarchical relationship between two point processes, where the outcome of a daughter process is conditioned on the outcome of the parent process. For example, it is the conditional relationship between the daughter and the parent that defines a cluster process. In SLAM, we consider the configuration of the map features to be the daughter process, which is conditioned on the state of the vehicle, which constitutes the parent process. As there is only one vehicle, there is only one cluster, so we are dealing with a single-cluster process. The full equations for the SC-PHD filter can be found in [13], but, for the sake of completeness, we will briefly summarize it here. Given a prior PHD for the vehicle position $\mathbf{X}$ and map $\mathbf{M}$

$$D_{k-1}(\mathbf{X}, \mathbf{M}) = s_{k-1}(\mathbf{X})\tilde{D}_{k-1}(\mathbf{M}|\mathbf{X}). \quad (3)$$

The first step in an iteration of the SC-PHD filter is the Chapman–Kolmogorov prediction

$$D_{k|k-1}(\mathbf{X}, \mathbf{M}) = \int s_{k-1}(\mathbf{X}')\pi_{k|k-1}(\mathbf{X}|\mathbf{X}')\tilde{D}_{k|k-1}(\mathbf{M}|\mathbf{X}')\,d\mathbf{X}', \quad (4)$$

where $\tilde{D}_{k|k-1}(\mathbf{M}|\mathbf{X})$ is the predicted PHD of the map

$$\tilde{D}_{k|k-1}(\mathbf{M}|\mathbf{X}) = \gamma_{k|k-1}(\mathbf{M}|\mathbf{X}) + \underbrace{\int \tilde{D}_{k-1}(\mathbf{M}'|\mathbf{X}')p_S(\mathbf{M}'|\mathbf{X}')\tilde{\pi}_{k|k-1}(\mathbf{M}|\mathbf{M}';\mathbf{X}')d\mathbf{M}'}_{\text{persistent}}. \quad (5)$$

Similar to most SLAM filters, the parent and daughter PHDs are convolved with Markov transition densities $\pi()$ and $\tilde{\pi}()$, respectively, to propagate them forward in time. However, there are also a couple of elements here that are not usually found in SLAM filters. The map prediction is composed of two parts. The additional term $\gamma_{k|k-1}(\mathbf{M}|\mathbf{X})$ is called the *birth intensity*, which models the PHD of newly appearing features. For practical purposes, the birth intensity is typically modeled from the current measurements [14]. Furthermore, the portion of the prediction corresponding to persistent targets incorporates a factor of $p_S(\mathbf{M}'|\mathbf{X}')$, which is a conditional probability distribution that models feature survival. The measurement update equations are

$$D_k(\mathbf{X}, \mathbf{M}) = \frac{s_{k|k-1}(\mathbf{X})L_{\mathbf{Z}_k}(\mathbf{X})}{\int s_{k|k-1}(\mathbf{X})L_{\mathbf{Z}_k}(\mathbf{X})d\mathbf{X}}\tilde{D}_k(\mathbf{M}|\mathbf{X}), \quad (6)$$

where $\tilde{D}_k(\mathbf{M}|\mathbf{X})$ is the updated PHD of the map

$$\tilde{D}_k(\mathbf{M}|\mathbf{X}) = \tilde{D}_{k|k-1}(\mathbf{M}|\mathbf{X}) \times \left[\underbrace{(1 - p_D(\mathbf{M}|\mathbf{X}))}_{\text{nondetection}} + \underbrace{\sum_{z \in Z_k} \frac{g(z|\mathbf{M},\mathbf{X})p_D(\mathbf{M}|\mathbf{X})}{\eta_z(\mathbf{M}|\mathbf{X})}}_{\text{detection}}\right] \quad (7)$$

$$\eta_z(\mathbf{X}) = k_k(z) + \tilde{D}_{k|k-1}(\mathbf{M}|\mathbf{X}) \times p_D(\mathbf{M}|\mathbf{X})g(z|\mathbf{M},\mathbf{X})d\mathbf{M} \quad (8)$$

$$L_{\mathbf{Z}_k}(\mathbf{X}) = \exp\left\{-\int p_D(\mathbf{M}|\mathbf{X})\tilde{D}_{k|k-1}(\mathbf{M}|\mathbf{X})d\mathbf{M}\right\} \times \prod_{z \in \mathbf{Z}_k} \eta_z(\mathbf{X}). \quad (9)$$

At first glance, these equations are understandably quite daunting, but we can tease out some meaning with closer inspection. In (6), we can see a Bayes' rule update for the parent PHD $s(\mathbf{X})$, where $L_{\mathbf{Z}_k}(\mathbf{X})$ is a multiobject observation likelihood. The same Bayes' rule structure can be seen inside the summation of (7). The remaining elements can be considered as a sort of bet hedging against the measurement difficulties previously discussed. The normalization term $\eta_z(\mathbf{X})$ incorporates a term $\kappa_k(z)$ that models the intensity of clutter measurements so that each received measurement is weighed against the possibility of being spurious. The updated map PHD is a sum of a detection and nondetection term. If a measurement is not received for a particular feature, it will remain in the updated map, weighted by a factor of $(1 - p_D(\mathbf{M}|\mathbf{X}))$.

Some additional modifications are necessary to adapt the SC-PHD filter from a tracking filter to a SLAM filter. For example, the probability of detection is conditional on the parent process since it is tied to the vehicle's field of view (FoV). Otherwise, we have found that the SC-PHD filter can be applied nearly out of the box, such as in the application described in this article.

Implementing the SC-PHD SLAM

To assess the suitability of the finite-set techniques for underwater navigational applications, we developed an implementation of an SC-PHD SLAM algorithm for use on an underwater robotic vehicle, the Girona 500. In this section, we will discuss a number of adaptations made to the algorithm as well as the experiment performed to evaluate its performance.

Girona 500 Vehicle

The Girona 500 [Figure 2(a)] is an underwater robotic vehicle developed by the Underwater Robotics Research Center (CIRS) at the University of Girona. It is a triple-hull design, with each torpedo-shaped hull measuring 0.3 m in diameter and 1.5 m in length. Its overall dimensions are 1.5 m $\times$ 1.0 m $\times$ 1.0 m (L $\times$ W $\times$ H), and its weight is less than 200 kg. Its maximum operating depth is rated at 500 m. The triple-hull design provides a high separation between the vehicle's center of buoyancy and center of gravity, which, in turn, affords stability in pitch and roll. The Girona 500's design is highly modular. It can be equipped with between three and eight thrusters in various configurations to allow for motion in three to six redundant degrees of freedom. In addition to the typical navigational and survey sensors, it has space dedicated to mission-specific payloads, such as robotic arms for intervention tasks.

Hybrid Particle PHD SLAM

As the SC-PHD can be separated into a parent and a conditional daughter term, we adopt a hybrid particle and Gaussian mixture approach for implementing the filter on our underwater vehicle. A particle vehicle state allows us to cope with a nonlinear observation model, while a Gaussian mixture map state keeps the computational expenses in the map update manageable. The filter is updated using velocity measurements from the DVL and relative positional measurements from landmarks. The particles representing the vehicle state are propagated forward in time using a three-dimensional (3-D) constant velocity motion model.

The vehicle state is defined by a 12-dimensional vector consisting of Cartesian and angular positions, and their respective velocities

$$\mathbf{X} = [x, y, z, \dot{x}, \dot{y}, \dot{z}, \theta, \phi, \psi, \dot{\theta}, \dot{\phi}, \dot{\psi}]^T.$$

The Cartesian displacements x, y, and z are relative to the global reference frame, while the respective velocities are given in the vehicle reference frame.

Typically, this would require a particle state space consisting of positions and velocities in both linear coordinates and roll/pitch/yaw angles, for a total of 12 dimensions, which would require an enormous number of particles. To mitigate the curse of dimensionality, we first treat the roll and pitch angles and velocities from the inertial measurement unit (IMU) as a trusted source, thus eliminating one-third of the dimensions in the particle state space. However, the compass observations show significant inaccuracies, and, therefore, yaw values must remain in the filter state. We continue even further by implementing a pipelined approach for the linear velocities. We maintain a separate EKF to filter the vehicle-frame velocities based on a random-walk prediction and linear measurements from the DVL. These filtered velocities are then transformed to the global reference frame and used to predict the vehicle particles with the constant velocity model. In this manner, we need only to maintain a particle state space of five dimensions: $(x, y, z, \theta, \dot{\theta})$. Using this pipelined approach leads to estimation in a lower-dimension state space with the particles and lower variance on the estimated state [15]. Other options to improve the filter are the use of optimal sampling strategies [16] and multiple model filters [17] to better characterize the different behaviors of the vehicle.

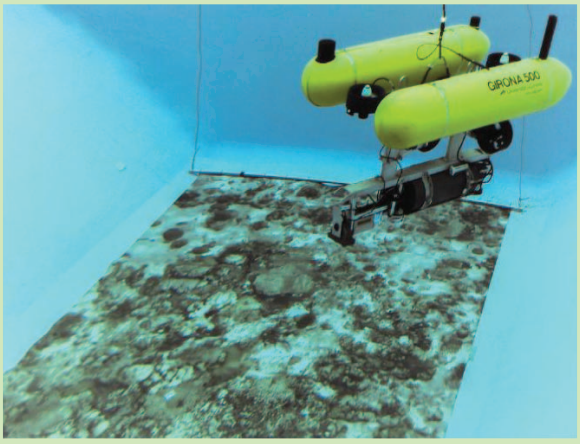
The reduced-dimensionality state is propagated forward through time using a constant-velocity motion model

$$\mathbf{X}_{k|k-1} = \begin{Bmatrix} \begin{pmatrix} x_{k-1} \\ y_{k-1} \\ z_{k-1} \end{pmatrix} + \mathbf{R}\begin{pmatrix} \dot{x}_{k-1} \\ \dot{y}_{k-1} \\ \dot{z}_{k-1} \end{pmatrix}\Delta t \\ \theta_{k-1} + \dot{\theta}_{k-1}\Delta t + n_\theta \dfrac{\Delta t^2}{2} \\ \dot{\theta}_{k-1} + n\theta\Delta t \end{Bmatrix}, \tag{10}$$

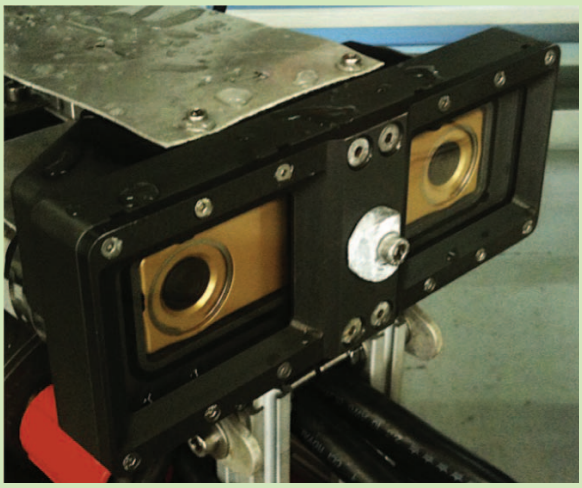
where $\mathbf{R}$ is the 3-D rotation matrix computed from the three orientation angles, n_θ is a process noise parameter on the yaw, and Δt is the length of the time interval.

The vehicle particles represent the parent state in the context of the SC-PHD filter, and, to each particle, we associate a Gaussian mixture, which represents the map, or daughter PHD conditioned on that particle's trajectory. Updating the filter with landmark observations thus consists of performing the PHD filter update for each map [18], which is then propagated upward to the parent process. This is reflected by the weight update of the parent particles in the single-cluster process. Figure 3 shows this hierarchical update.

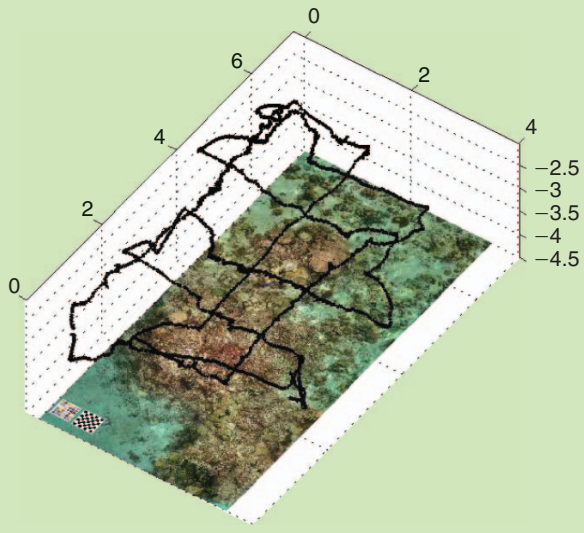
The SC-PHD SLAM algorithm bears some resemblance to a previous PHD filter SLAM solution: Rao–Blackwellized PHD (RB-PHD) SLAM [8]. Although the differences may be subtle, they are nevertheless important. The chief distinguishing factor between the two algorithms is the likelihood used to update the vehicle location (7). To evaluate the RB-PHD likelihood, Poisson approximations are required on both the prior and the posterior, and a further approximation is necessary on the number of map features. In contrast, the SC-PHD filter requires only a Poisson approximation on the prior. This has been shown to produce superior results in SLAM [13]. Moreover, the SC-PHD filter is not restricted to a particle filter implementation as Gaussian parent implementations have been proposed [19].



(a)



(b)



(c)

Figure 2. The elements of the underwater SC-PHD SLAM experiment. (a) The Girona 500 hovering above the test tank floor. The known pattern of the floor is used to estimate the ground-truth trajectory of the vehicle. (b) The housing for the Girona 500's stereo camera. (c) The camera-centered ground-truth trajectory constructed by matching the collected image with the priori seafloor pattern. (Photos courtesy of the Underwater Robotics Research Center at the University of Girona.)

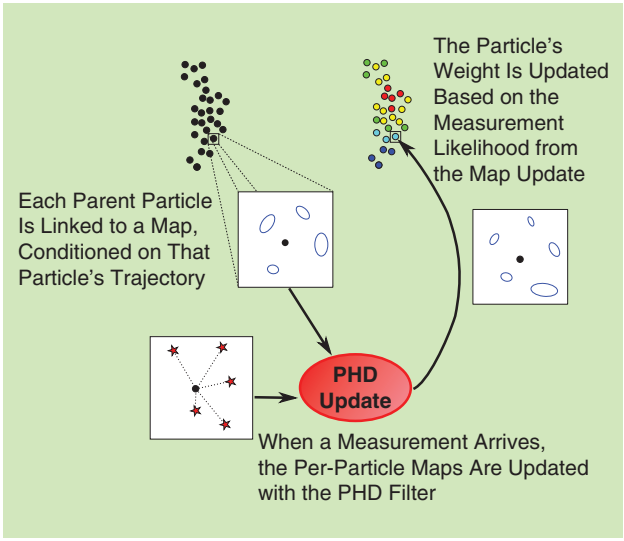


Figure 3. The SC-PHD SLAM process can be represented by the above hierarchy. The updates from landmark updates propagate upward and reweight particles representing the vehicle position.

Detecting Features of Interest

In the absence of position measurements, the increase in uncertainty on vehicle position is unbounded. This can be reduced by incorporating measurements with respect to static targets underwater. Typically, these targets will correspond to features in manmade structures, such as risers, pipes, anchor chains, or naturally occurring rock formations and/or vegetation. In our experiments, we utilize a stereo camera to detect and triangulate such points of interest with respect to the camera. While our experiments here rely solely on vision-based features, sonar is also an important sensing regime for underwater robotics, although feature detection can be more problematic with sonar imagery compared with vision [20].

Points of interest in the sensor FoV can correspond to physically meaningful structures with semantic meaning or merely an abstract set of points. For example, an underwater riser in the camera FoV may represent a point of interest. In this case, we can attribute a semantic tag to the structure and possibly other identifying information. In contrast to this, a lower-level analysis of the structure may decompose the riser into a set of corners, lines, and/or circles, with each of these representing a point of interest. The use of semantic tags is only possible in the presence of prior information. The vehicle is unable to assign tags to unrecognizable landmarks, and, in such instances, points of interest, such as points and lines, are more appropriate.

Our approach parameterizes features as points in $\mathbb{R}^3$ and involves the use of standard image feature detectors, such as speeded-up robust features (SURF) [21] to detect a small number of strong features or key points in the image. By detecting and matching key points in a pair of stereo images, we can approximately triangulate the position of key points in the camera coordinate system (Figure 4). For complete details of the observation process, see [22]. Due to changes in illumination and contrast, the set of detected key points can change significantly between images. This gives rise to a situation

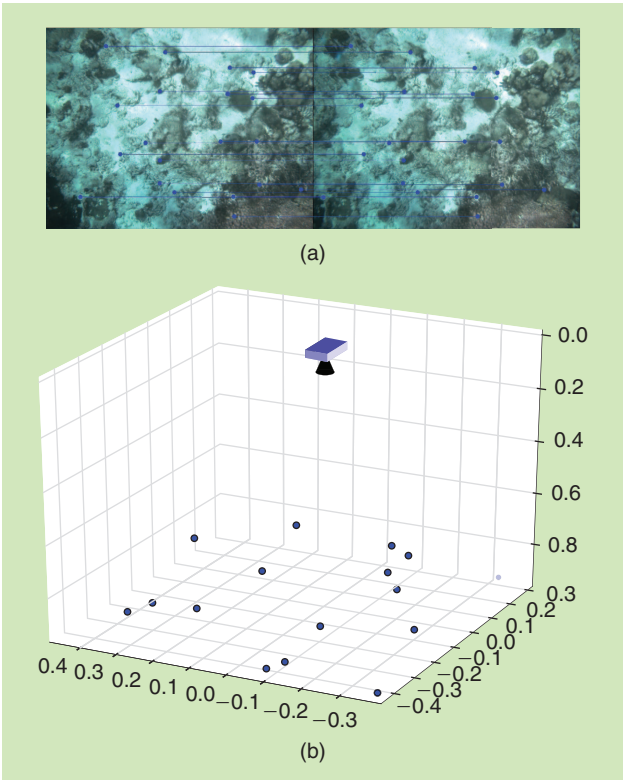


Figure 4. (a) The key points are matched between the left and right images. (b) Triangulation gives the 3-D coordinates of the points with respect to the camera. (Photo courtesy of the Underwater Robotics Research Center at the University of Girona.)

with numerous unstable key points with poor probability of detection. The more stable key points will be tracked by the filter, while key points that are visible only briefly will be modeled by the clutter RFS in the PHD filter.

Underwater SC-PHD SLAM

To test the SC-PHD SLAM algorithm, we conducted an experiment in the test tank at CIRS, which measures 8 m × 8 m × 5 m. The bottom of the test tank is overlaid with a known pattern simulating the sea floor [Figure 2(a)], and the vehicle was teleoperated in a lawnmower survey trajectory over this pattern.

The vehicle is equipped with a downward-looking stereo camera recording at approximately 10 frames/s. The SURF features are extracted from the camera images to provide measurements for the SC-PHD SLAM algorithm. For the purpose of the filter update, the map features are projected onto the camera image plane, and the likelihoods are evaluated in the image domain. This allows for better characterization of the measurement noise. Due to the fact that the appearance of the surface over which the vehicle is travelling is known a priori, it is also possible to obtain a ground-truth trajectory for this scenario by reprojecting the collected camera images onto the pattern. This is shown in Figure 2(c).

Determining a ground truth for the map is far less straightforward. As we have the image file on hand for the map mosaic, it is possible to run a feature extractor on the image

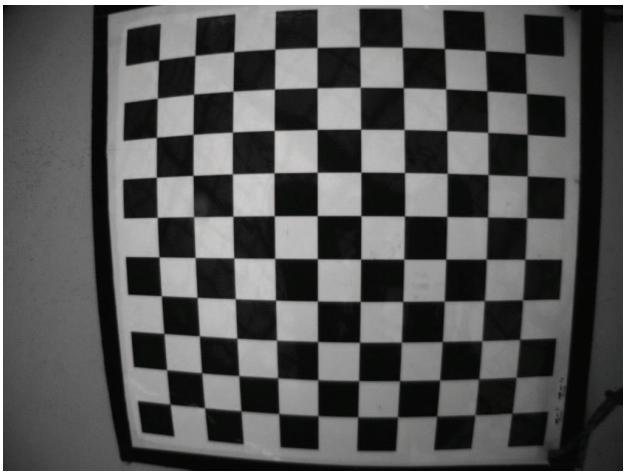


Figure 5. The distorted images from the downward-looking stereo camera. (Photo courtesy of the Underwater Robotics Research Center at the University of Girona.)

and predict where the most salient points of the environment would be. However, these features correspond to a top-down whole-scene view under ideal lighting conditions. It is not certain that the features extracted with this approach would be consistently observed from the various viewing angles, depths, and illumination conditions over the course of the vehicle's trajectory. Therefore, we instead opt with a vehicle-oriented approach for locating the map features. Using the ground-truth trajectory described previously, we project the SURF features from each camera frame onto the mosaic and match the projected features against features detected in the mosaic. Features that are successfully matched are used to construct the two-dimensional histogram observed in Figure 5.

The raw images from the camera exhibit a high degree of distortion. Although corrected for, this distortion can manifest as errors in triangulation if camera calibration is inexact. We account for this in the experiment by assigning a relatively high level of additive noise to the feature measurements.

The results of the experiment are shown in Figure 6. Without SLAM, localization is performed using an EKF filter using velocity measurements from the DVL. It is clear that there is significant drift in the estimated vehicle position when only the DVL is used to perform dead reckoning. Aside from accumulated error in the velocity readings, the dead reckoning estimate also suffers from inaccurate heading measurements. As mentioned previously, the IMU values are trusted completely, but, in reality, the sensor is affected by soft-metal effects due to the proximity to the test tank's concrete walls. This is evidenced by the significant deviation following the initial bend in the vehicle trajectory. In some portions of the trajectory, the IMU heading was off by up to 15°. For operations in open-sea settings, this effect should be far less pronounced as the sensor will be well separated from sources of interference.

Introducing the seafloor pattern as a navigational reference improves matters. Applying the SC-PHD SLAM algorithm results in a better estimation of the trajectory. The multiple loop closures in the trajectory serve to keep the localization error bounded and eliminates the cumulative

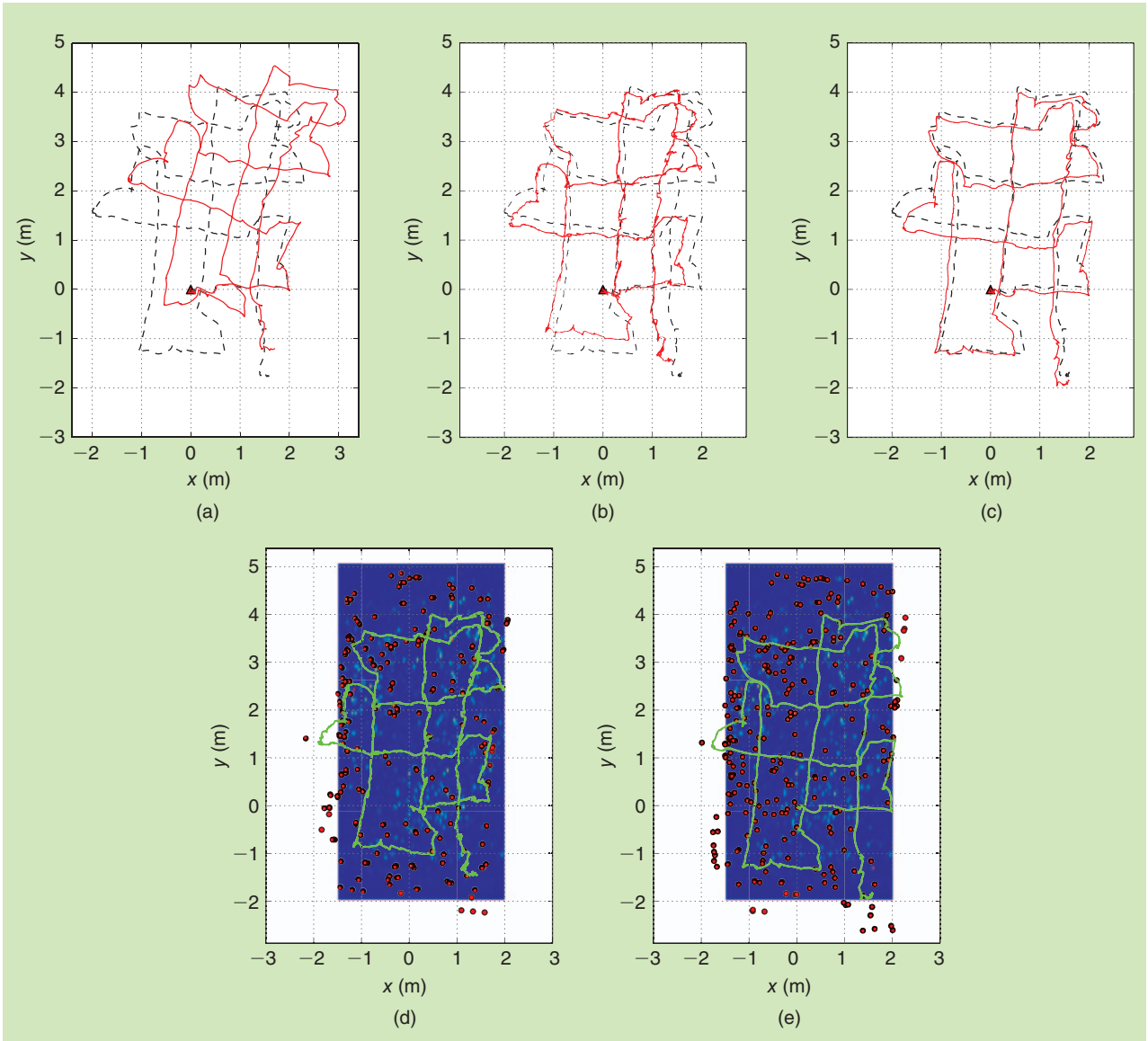


Figure 6. The SLAM results from the underwater experiment. The start of the trajectory is indicated by a triangle. SC-PHD SLAM exhibits a lower trajectory error compared to RB-PHD SLAM, while both SLAM methods are superior to odometry alone. The estimated landmarks for RB-PHD SLAM and SCPHD SLAM are overlaid on the ground truth histogram described in the "Underwater SC-PHD SLAM" section. Landmarks outside the histogram area originate from features detected in the test tank environment, outside of the mosaic. (a) The dead reckoning trajectory. (b) The RB-PHD SLAM trajectory. (c) The SCPHD SLAM trajectory. (d) The RB-PHD SLAM landmarks. (e) The SCPHD SLAM landmarks.

drift that occurs otherwise. The map state is taken from the map PHD belonging to the most highly weighted particle. As we have opted for a Gaussian mixture implementation for the map PHD, we can obtain the expected number of features by summing the weights of all of the mixture components; this is equivalent to taking the integral of the PHD over the entire map. We then select the most highly weighted mixture components as the locations of the map features. This is shown in Figure 5.

For comparison, we ran RB-PHD SLAM on our scenario using the single-feature approximation for computing the likelihood. Our results show that SC-PHD SLAM provides noticeably improved trajectory estimates.

Due to the limited computing capacity on board the vehicle, the SLAM algorithms are run offline. Clearly, autonomy is of paramount importance for robotics applications. We anticipate that, through a combination of algorithm optimization, improvements in available computing resources, and energy storage efficiency, real-time and online execution of SC-PHD SLAM is not too far in the future.

Outlook

We have observed here that PHD filter methods are suitable for application to the SLAM problem. There are, of course, some refinements that can be made to our implementation. For example, the suboptimal performance of the IMU in the

constricted test tank environment warrants a more conservative handling of its readings. Our research in this avenue will continue as we integrate the SC-PHD SLAM algorithm into a larger autonomous control framework, using its navigational estimates to inform higher-level tasks and missions.

The scenario presented here was in a high-light environment, where cameras may be used and a myriad of computer vision techniques can be brought to bear. Underwater vehicles operating at greater depths will be without this luxury and will need to rely on sonar, and feature detection will be even less straightforward. We expect that, given the PHD filter's capability to excel in difficult sensing regimes, it can rise to meet these challenges, and we are looking forward to seeing many interesting applications in the underwater arena.

Acknowledgments

We would like to recognize the generous support of the European Commission through the FP7-ICT-2011-7 project PANDORA—Persistent Autonomy through Learning, Adaptation, Observation and Re-planning (Ref 288273), the Engineering and Physical Science Resarch Council through their grant for Sustained Autonomy through Coupled Plan-based Control and World Modelling with Uncertainty (EP/J012432/1), and the Catalan government through the FI-DGR grant.

References

[1] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," in *Proc. 2nd Conf. Uncertainty Artificial Intelligence*, 1986, pp. 435–461.

[2] J. Leonard and H. Durrant-Whyte, "Simultaneous map building and localization for an autonomous mobile robot," in *Proc. IEEE/RSJ Int. Workshop Intelligent Robots Systems Intelligence Mechanical Systems*, Nov. 1991, vol. 3, pp. 1442 –1447.

[3] R. Martinez-Cantin and J. Castellanos, "Unscented SLAM for large-scale outdoor environments," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, Aug. 2005, pp. 3427–3432.

[4] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM: A factored solution to the simultaneous localization and mapping problem," in *Proc. AAAI Nat. Conf. Artificial Intelligence*, Edmonton, Canada, 2002, pp. 593–598.

[5] 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. 16th Int. Joint Conf. Artificial Intelligence*, Acapulco, Mexico, 2003, pp. 1151–1156.

[6] J. Mullane, B.-N. Vo, M. Adams, and B.-T. Vo, *Random Finite Sets for Robot Mapping and SLAM—New Concepts in Autonomous Robotic Map Representations* (Series Springer Tracts in Advanced Robotics). Berlin Heidelberg, Germany: Springer-Verlag, 2011, vol. 72.

[7] C. S. Lee, D. Clark, and J. Salvi, "SLAM with dynamic targets via single-cluster PHD filtering," *IEEE J. Select. Topics Signal Processing*, vol. 7, no. 3, pp. 543–552, June 2013.

[8] J. Mullane, S. Keller, and M. Adams. (2012, May). Random set versus vector based SLAM in the presence of high clutter. Presented at IEEE ICRA 2012: Workshop on Stochastic Geometry in SLAM. [Online]. Available: http:// www.cec.uchile.cl/~martin/icra_2012_workshop_stoch_geom_slam 4_5_12/wkshp_final_papers/martin_adams_ICRA12_SLAM_Wkshp.pdf

[9] D. Moratuwage, B.-N. Vo, S. Wijesoma, and D. Wang, "Extending the Bayesian RFS SLAM framework to multi-vehicle SLAM," in *Proc. 12th Int. Conf. Control Automation Robotics Vision*, 2012, pp. 638–643.

[10] D. Clark, C. Lee, and S. Nagappa. (2012). Singlecluster PHD filtering and smoothing for SLAM applications. Presented at ICRA. [Online]. Available: http://www.cec.uchile.cl/~martin/icra_2012_workshop_stoch_geom_slam_4_5_12/wkshp_final_papers/Daniel_Clark_ICRA12_Workshop_final.pdf

[11] R. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.

[12] D. Clark and R. Mahler, "Generalized PHD filters via a general chain rule," in *Proc. IEEE 15th Int. Conf. Information Fusion*, 2012, pp. 157–164.

[13] C. Lee, D. Clark, and J. Salvi, "SLAM with single cluster PHD filters," in *Proc. IEEE Int. Conf. Robotics Automation*, 2012, pp. 2096–2101.

[14] B. Ristic, D. 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.

[15] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods* (Series Springer Texts in Statistics), 2nd ed. New York: Springer-Verlag, 2004.

[16] R. van der Merwe, A. Doucet, N. de Freitas, and E. Wan. (2000, Dec.). The unscented particle filter. in *Advances in Neural Information Processing Systems (NIPS13)*, T. G. D. T. K. Leen and V. Tresp, Eds. Cambridge, MA: MIT Press. [Online]. Available: http://cslu.cse.ogi.edu/publications/ps/merwe00a.ps.gz

[17] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter*. Norwood, MA: Artech House, 2004.

[18] B.-N. Vo and W.-K. Ma, "The gaussian mixture probability hypothesis density filter," *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4091–4104, 2006.

[19] A. Swain and D. Clark, "The single-group PHD filter: An analytic solution," in *Proc. Int. Conf. Data Fusion*, 2011, pp. 1–8.

[20] N. Hurtos, X. Cufí, Y. Petillot, and J. Salvi, "Fourier-based registrations for two-dimensional forward-looking sonar image mosaicing," in *Proc. IEEE/RSJ, Int. Conf. Intelligent Robots Systems*, Oct. 2012, pp. 5298–5305.

[21] H. Bay, A. Ess, T. Tuytelaars, and L. van Gool. (2008, June). Speeded-up robust features. *Comput. Vis. Image Underst.* [Online]. *110(3),* pp. 346–359. Available: http://dx.doi.org/10.1016/j.cviu.2007.09.014

[22] S. Nagappa, N. Palomeras, C. Lee, N. Gracias, D. Clark, and J. Salvi, "Single cluster PHD SLAM: Application to autonomous underwater vehicles using stereo vision," in *Proc. IEEE OCEANS Conf.*, Bergen, Norway, June 2013, pp. 1–9.

Chee Sing Lee, Computer Vision and Robotics Group, University of Girona, Spain. E-mail: cslee@eia.udg.edu.

Sharad Nagappa, Computer Vision and Robotics Group, University of Girona, Spain. E-mail: snagappa@eia.udg.edu.

Narcis Palomeras, Computer Vision and Robotics Group, University of Girona, Spain. E-mail: npalomer@eia.udg.es.

Daniel E. Clark, Joint Research Institute in Signal and Image Processing, Heriot-Watt University, Edinburgh, United Kingdom. E-mail: d.e.clark@hw.ac.uk.

Joaquim Salvi, Computer Vision and Robotics Group, University of Girona, Spain. E-mail: qsalvi@eia.udg.edu.