Counter-Hypothetical Particle Filters for Single Object Pose Tracking

Elizabeth A. Olson Jana Pavlasek Jasmine A. Berry Odest Chadwicke Jenkins

***Abstract*— Particle filtering is a common technique for six degree of freedom (6D) pose estimation because of its ability to tractably represent belief over object pose. Due to the high-dimensional nature of 6D pose, this application of the particle filter is prone to particle deprivation and can lead to mode collapse of the underlying belief distribution in the importance sampling step. When this occurs, recovering belief in the region surrounding the true state is challenging since it is no longer represented in the probability mass formed by the particles. Previous methods mitigate this problem by reinvigorating particles in the predicted belief, but determining the frequency of reinvigoration has relied on hand-tuning abstract heuristics. In this paper, we estimate the necessary reinvigoration rate at each time step by introducing a *Counter-Hypothetical likelihood function*, which is used alongside the standard likelihood. Inspired by the notions of plausibility and implausibility from evidential reasoning, the addition of our Counter-Hypothetical likelihood function assigns a level of doubt to each particle. The competing cumulative values of confidence and doubt across the particle set are used to estimate the level of failure within the filter, in order to determine the portion of particles to be reinvigorated. We demonstrate the effectiveness of our method on the rigid body object 6D pose tracking task.**

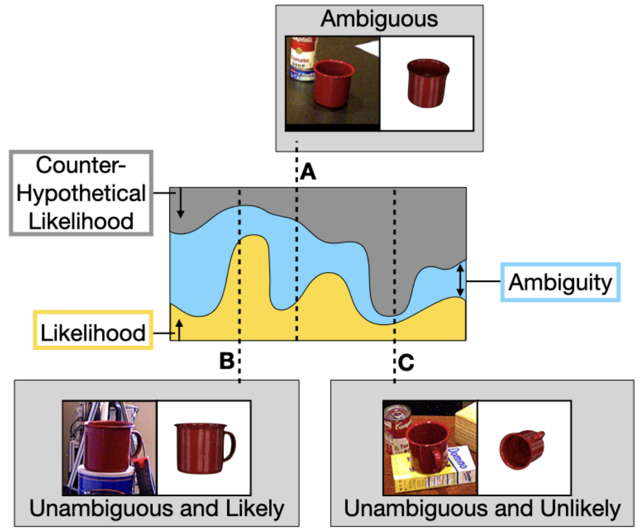

Fig. 1: To predict when the particle filter is in a failure mode, we estimate the doubt of each sample with a Counter-Hypothetical likelihood function. Inspired by Evidential Reasoning, we consider different factors for supporting a given estimate (the likelihood function) than those used for disproving the estimate (the Counter-Hypothetical likelihood function). The gap between these quantities represents ambiguity in the evidence. We illustrate how different observations and estimates of the same mug could cause various weightings in (A) the mug's handle is occluded, making the estimate ambiguous (B) the pose is unambiguously likely (C) the pose is unambiguously doubtful.

I. Introduction

Object pose tracking is crucial for autonomous manipulation tasks, but current methods struggle to remain reliable in occluded and cluttered scenes. Promising convolutional neutral networks have been trained to estimate 6D pose directly [1], [2], as well as measuring uncertainty within the likelihood of a particle filters for improved accuracy [3]. We propose a new way of further embedding deep learning into the particle filter, by learning to identify significant error in the filter's predictions to indicate if it is in failure mode. Particle filters typically can afford only a small number of samples when compared to the overall size of the continuous state space. Certain regions of the state space will contain no particles, making them unrepresented in the belief distribution. This phenomenon is called *particle deprivation*, and can occur due to poor initialization or the stochasticity of importance sampling. Regaining belief in these regions can be achieved by particle reinvigoration, however determining the portion of particles to be reinvigorated at a given iteration often requires tedious hand-tuning through trial and error. Methods that estimate the portion of particles that need to be reinitialized, such as ours, are known as *adaptive particle reinvigoration*.

We introduce the Counter-Hypothetical Particle Filter (CH-PF) to counteract the problem of particle deprivation in high-dimensional state spaces. Our method proposes to quantify the confidence that the true state is unrepresented in the sample set. To this aim, we model the evidence *against* a particular hypothesis, termed the *Counter-Hypothetical* likelihood. It measures this weighting independently of the traditional likelihood through the lens of Evidential Reasoning (Dempster-Shafer Theory) [4]. This framework argues that the plausibility and implausiblity of proving an outcome can be based on different factors, and are not zero-sum due to potential overlap and ambiguity of the underlying evidence. Each particle is given both a likelihood weighting and a counter-hypothetical likelihood weighting, which are used to quantify the cumulative confidence and doubt across the sample set. The relationship between these values is used to reason about the likelihood that the true state is underrepresented in our sample set, and in turn to compute an adaptive rate of particle reinvigoration.

II. Related Work

A. Object Pose Estimation and Tracking for Robotics

Pose estimation and tracking have received considerable attention in the robotics community. In recent years, various works have demonstrated the capability of data-driven methods to provide discriminative pose estimates over a single

E. A. Olson, J. Pavlasek, J. A. Berry, and O. C. Jenkins are with the Robotics Department, University of Michigan, Ann Arbor, MI, USA, {lizolson, pavlasek, jasab, ocj}@umich.edu.

view [1], [2], [5], [6]. Data-driven methods have also been applied to pose tracking over a sequence of observations [7]. These methods have achieved impressive results but are prone to inaccuracies, particularly in challenging scenes such as those with heavy clutter. Probabilistic inference methods instead maintain belief over pose estimates in order to provide additional robustness for robotic manipulation applications. We focus on probabilistic inference for object pose estimation and tracking, specifically on particle filtering.

Particle filtering is an iterative inference algorithm which can represent an arbitrary nonparametric belief distribution using a set of weighted particles sampled from the state space [8]. Particle filtering is a common technique for 6D pose estimation and tracking [9], [10], [11], [12]. The sampling-based method allows for efficient approximation of the high-dimensional state space, and the nonparametric nature of the algorithm enables the representation of multiple competing hypotheses in the belief. More recently, deep convolutional neural networks have been applied to particle filtering for pose estimation. Deng et al. [3] create an observation model from autoencoder embeddings which is used within a Rao-Blackwellized particle filter. Though the main contribution of our work is the introduction of the Counter-Hypothetical likelihood function, we also demonstrate how this could could be learned in an end-to-end fashion.

B. Robust Particle Filtering

Many works have focused on mitigating particle deprivation. One approach is annealing, in which the distribution of importance weights is smoothed to avoid collapsing modes during importance sampling [13]. Annealing typically requires hand-tuning the annealing rate and schedule. In contrast, our work does not require modification to the sampling weights, but instead handles particle deprivation through reinvigoration.

In the global localization stage of Monte Carlo localization for mobile robots, particle deprivation is a common problem [8], motivating many works to reinitialize samples as needed. Preliminary work focused on sampling from an inverse distribution based on the sensor readings [14]. Similarly, Lenser and Veloso "reset" a subset of particles with samples drawn from the sensor distribution when the average probability of the current sample set is low [15]. Augmented Monte Carlo localization [16][17] extends this idea by performing particle reinvigoration from a uniform distribution at a rate proportional to the difference between the long- and short-term averages of the particle weights, instead of a fixed threshold. Recent work in localization leverages the estimates of a neural network by sampling from this proposal at a fixed rate and fuses the particles into the distribution through Importance Sampling [18]. These methods mitigate particle deprivation by resetting samples at a fixed rate, or by determining the rate through heuristics applied to the likelihood weights. Our method instead uses output from the trained Counter-Hypothetical likelihood to indicate when the particle filter may be in failure mode.

III. Counter-Hypothetical Particle Filter

We consider the problem of tracking a known object over time. Given a sequence of RGB images or RGB-D data, $z_{1:t}$, we seek to localize the pose, $x_t \in \mathcal{X}$, of an object at time t. We also model any motion to the system, caused by either user input or jittering, with $u_{1:t}$. Here $\mathcal{X}$ represents the space of 6D poses, comprising of 3D translation and 3D rotation.

A. Particle Filtering

The particle filter is a Bayes filtering algorithm in which the belief distribution $bel(x_t)$ is a nonparametric distribution approximated by a particle set, $\mathbb{X}_t$:

$$\mathbb{X}_t = \{(x_t^1, \pi_t^1), (x_t^2, \pi_t^2), \ldots, (x_t^N, \pi_t^N)\} \tag{1}$$

Each particle, x_t^i has a corresponding weight, π_t^i. The predicted belief, $\widehat{bel}(x_t)$, is formed by applying action u_t to each particle in the previous sample set, $\mathbb{X}_{t-1}$. The particle set $\mathbb{X}_t$ is then generated through *importance sampling* from $\mathbb{X}_{t-1}$. Importance sampling allows us to sample from a target distribution, $\phi_{tar}(x_t)$, by sampling from a proposal distribution, $\phi_{prop}(x_t)$, with the precondition that $\phi_{tar}(x_t) > 0 \implies \phi_{prop}(x_t) > 0$. In the particle filter, the target distribution is the posterior belief, $bel(x_t)$, and the proposal distribution is the predicted belief, $\widehat{bel}(x_t)$. Samples from the proposal are drawn with replacement, where the probability of a particle being drawn is proportional to its weight, given by:

$$\pi_t^i = \frac{\phi_{tar}(x_t)}{\phi_{prop}(x_t)} = \frac{bel(x_t)}{\widehat{bel}(x_t)} = p(z_t|x_t^i) \tag{2}$$

We call the function which generates the importance weights the likelihood function, denoted:

$$\mathcal{L}(x_t^i) := p(z_t|x_t^i) \tag{3}$$

If the candidate particle set does not include samples close to the true value of the state, the probability of sampling values in this region is negligibly small, which is known as particle deprivation.

B. Particle Reinvigoration

A common approach to mitigating particle deprivation is *particle reinvigoration*, in which particles are drawn jointly from the predicted belief, $\widehat{bel}(x_t)$ and a candidate distribution, $\phi_{cand}(x_t)$. Choices of candidate distribution might include a uniform distribution over the region of interest of the state space, or a wide Gaussian distribution around an initial estimate. These distributions allow importance sampling to draw from outside of the sample set, reintroducing samples in underrepresented regions. However, sampling from the candidate distribution too frequently in place of the predicted belief will remove key information from the prior distribution from our Bayesian filter, while sampling from the prior distribution too frequently could lead to particle deprivation.

We represent the proportion of samples to be drawn from the $\phi_{cand}(x_t)$ distribution by α, where $0 \leq \alpha \leq 1$. The final particle set is defined as the union of particles drawn from each set:

$$\mathbb{X}_t = \{(x_t^1, \pi_t^1), , \dots, (x_t^{\alpha N}, \pi_t^{\alpha N})\} \bigcup \{(x_t^{\alpha N+1}, \pi_t^{\alpha N+1}), \dots, (x_t^N, \pi_t^N)\} \tag{4}$$

where $x_t^i \sim \phi_{cand}$ for $1 \leq i \leq \alpha N$, and $x_t^j \sim \widehat{bel}$ for $\alpha N < j \leq N$. Note that in practice, αN is constrained to be an integer. We aim to adaptively select the reinvigoration rate, α, at each iteration, such that α fluctuates in accordance with the portion of the true belief distribution we believe to be underrepresented by $\mathbb{X}_t$.

One method of achieving adaptive particle reinvigoration is Sensor Resetting Localization (SRL) [15]. This method defines a probability threshold, β, which represents a threshold on "good" unnormalized likelihood values. The reinvigoration rate is defined as:

$$\alpha = 1 - \left(\frac{1}{\beta N} \sum_{i=1}^{N} \mathcal{L}(x_t^i) \right) \tag{5}$$

The Counter-Hypothetical Particle Filter builds off this equation, using a Counter-Hypothetical likelihood to compute the reinvigoration rate instead of a probability threshold.

C. Counter-Hypothetical Resampling

With traditional Bayesian probability, the observed probability of a state being true and the observed probability of a state being false are zero-sum. The Counter-Hypothetical Particle Filter analyzes the change in performance of the particle filter when the ambiguity in the likelihood is considered, taking inspiration from Evidential Reasoning [4].

To describe the calculation of our reinvigoration rate, we first rewrite Equation (5):

$$\alpha = 1 - \frac{\sum_{i=1}^{N} \mathcal{L}(x_t^i)}{(\sum_{i=1}^{N} f(x_t^i)) + (\sum_{i=1}^{N} \mathcal{L}(x_t^i))} \tag{6}$$

where $f(x_t^i) := \beta - \mathcal{L}(x_t^i)$. With this notation, the numerator and right hand side of the denominator are an aggregate measurement of the likelihood of the sample set. The left hand side of the denominator, $\sum_{i=1}^{N} f(x_t^i)$, measures the poor performance across the sample set. In this way, calculating the rate of particle reinvigoration in SRL can be seen as simultaneously measuring the positive performance and poor performance of the sample set. However, the measure of poor performance, $f(x_t^i)$, is dependent on the measure of positive performance, as it is defined by $\mathcal{L}(x_t^i)$. Our approach replaces $f(x_t^i)$ with a Counter-Hypothetical likelihood, which is estimated independently of $\mathcal{L}(x_t^i)$.

We train the Counter-Hypothetical likelihood to measure how the observed image provides evidence *against* the hypothetical proposed state. Dempster-Shafer Theory allows for the evidence discounting an event, *generalized disbelief*, to be quantified independently of the evidence associated with supporting an event, *generalized belief* [19]. Generalized belief is a measurement of all evidence that undeniably supports an event, which is bounded from above by plausibility, which includes the ambiguity in the belief. Similarly, generalized disbelief quantifies the evidence that works to disprove an event, while implausiblity is an upper bound that considers ambiguity. We posit that this framework is apt for our application of evaluating object poses based on images. The occlusions and geometric symmetries present suggest that there is ambiguity in how a given pose is supported or unsupported by the evidence, motivating us to measure these quantities independently. We consider the likelihood function to be analogous to Dempster-Shafer Theory's notion of generalized belief, and therefore introduce a Counter-Hypothetical likelihood to function in accordance with generalized disbelief. We do not directly use generalized belief to reason about the true underlying probability, a common misstep when applying this method [20], we merely take inspiration from Evidential Reasoning by measuring doubt independently of confidence.

We introduce a function to reason about the confidence of a state counter to our given hypothesis, the *Counter-Hypothetical likelihood*, $\mathcal{C}(x_t)$. By comparing the quantities of the (unnormalized) likelihoods, $\mathcal{L}(x_t^i)$ and $\mathcal{C}(x_t^i)$, across the proposal distribution, we can reason about the cumulative confidence and doubt in our sample set. To this end, we redefine α, the reinvigoration ratio, as follows by modifying Equation (6):

$$\alpha = 1 - \frac{\sum_{i=1}^{N} \mathcal{L}(x_t^i)}{\sum_{i=1}^{N} \mathcal{C}(x_t^i) + \sum_{i=1}^{N} \mathcal{L}(x_t^i)} \tag{7}$$

We then sample αN particles from ϕ_{cand} in accordance with our doubt in the set, and sample the remaining $(1 - \alpha)N$ particles from $\widehat{bel}$ based on our confidence in the set. As such, the Counter-Hypothetical likelihood quantifies our notion of generalized disbelief, which controls the amount of particle reinvigoration to be performed.

IV. Experiments

To evaluate the proposed Counter-Hypothetical likelihood, we measure key performance metrics on the YCB-Video Dataset, a benchmarked real-world dataset [1]. We implement a standard particle filter to estimate the 6D pose of a given object across the video sequences. Our results test on both RGB and RGB-D data. To test the proposed modification, we design a Counter-Hypothetical likelihood function, in which the function is learned. The baselines are built upon the same particle filter, but each are modified to capture a standard practice for mitigating particle deprivation. For the baselines description, please see Section II.

A. Implementation Details

All particle filter variants in the experiment use the same initialization, particle set size of 50, and RGB/RGB-D likelihood functions to weight the particles. The candidate reinvigoration distribution is a wide Gaussian distribution sampled off of a region of interest produced by PoseCNN [1], which is also used for the initialization. To estimate the likelihood of a sample, we use the pretrained encoder network provided

	RGB						RGB-D					
	Occluded		Non-Occluded		All		Occluded		Non-Occluded		All	
	ADD	ADD-S	ADD	ADD-S	ADD	ADD-S	ADD	ADD-S	ADD	ADD-S	ADD	ADD-S
Annealing [13]	40.0	76.7	55.5	78.6	48.7	77.7	47.8	63.3	77.0	91.7	64.2	79.2
SRL [15]	42.6	**81.2**	60.0	79.1	50.6	80.0	53.1	81.3	76.6	92.1	66.3	87.3
Aug. MCL [17]	**44.7**	80.6	57.9	78.9	52.1	79.6	52.4	78.2	**77.3**	**92.3**	66.4	86.1
MCL+E2E [18]	41.2	62.3	**67.6**	**87.7**	**56.0**	76.5	49.2	65.4	75.1	92.0	63.7	80.3
CH-PF (ours)	40.2	80.9	57.2	81.4	49.8	**81.2**	**59.3**	**83.1**	75.1	92.2	**68.2**	**88.2**

TABLE I: AUC scores for varying modifications of a 6D single object pose tracking particle filter on the YCB Video Dataset. Our presented method, CH-PF, outperforms other methods focused on alleviating particle deprivation in instances of occlusion, when the information from the likelihood function is the most informative (RGB-D data).

by PoseRBPF [3]. This work provides RGB-D encoders for the crop of the observation image and for the rendered candidate pose. This likelihood function is trained as an auto-encoder, in which the network learns to reconstuct a pose rendering. Unlike the original PoseRBPF paper, we do not initialize with ground truth, or use a Rao-Blackwellized implementation. We instead filter across the full 6D, and allow for poor initialization, to test the performance of these adaptive particle reinvigoration methods.

To obtain the weightings for the Counter-Hypothetical likelihood, we use a similar architecture as PoseRBPF and also use synthetic data. However, the training strategy in PoseRBPF is designed to encode different poses with symmetry similarly, in order to maintain a multi-modal distribution in their belief. For the Counter-Hypothetical implementation, we instead train the embeddings to focus on absolute error between an observed object and a rendering of a candidate pose. Half of the pairs of training data is generated such that the estimated pose is a very small perturbation from the ground truth pose of the observation image, while the other half is paired with a randomly generated pose estimate. These categories become the labels for the contrastive loss, so embeddings of very close poses are similar, and anything else has significant mismatch. This demonstrates how the Counter-Hypothetical likelihood function does not need a different architecture or data, but is merely more intentionally trained to convey “red flags" of noticeable error.

V. Results

We present results with the absolute and symmetric pointwise matching errors between the estimated and ground truth pose (commonly referred to as ADD and ADD-S respectively). These errors are typically analyzed by viewing the Area Under the Curve (AUC) score of each method. Full quantitative results are shown in Table I.

This analysis shows that reinvigorating with regressed pose at a fixed rate works well for non-occluded environments. However, we observe the presence of occlusions negatively affects the performance of the hybrid method that is strongly biased by the output of the neural network (MCL+E2E). Selected qualitative results for the Counter-Hypothetical Particle Filter method are shown in Figure 2. We visualize how our algorithm is capable of increasing particle reinvigoration when the filter is converging to an incorrect estimate, and lessening the reinvigoration when it is not needed.

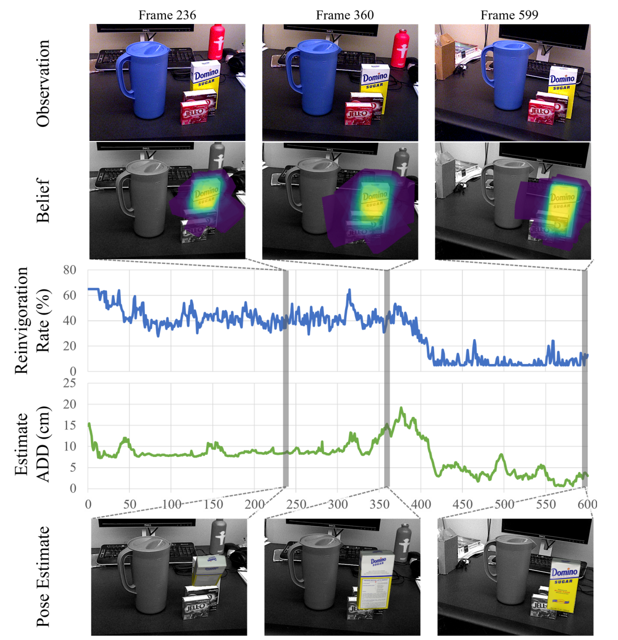

Fig. 2: Selected qualitative results for the Counter-Hypothetical Particle Filter. The belief of the sugar box converges to a local maximum in early frames (left). CH-PF applies a higher reinvigoration rate to mitigate this. The error in the estimate briefly increases (middle), but the belief eventually converges to the correct estimate (right).

VI. Conclusion

This work aims to improve the accuracy of particle filters tracking the 6D pose of rigid objects by adapting the rate of particle reinvigoration based on the estimated incompleteness of the current belief distribution. We propose independently estimating the potential error in each samples through a novel Counter-Hypothetical likelihood function. This modification allows us to reason over the cumulative doubt in our particle set, and use this estimate to apply particle reinvigoration as needed. This paper demonstrates the effectiveness of this modification in overcoming poor initialization when compared to other resampling techniques and particle filter implementations similarly focused at maintaining particle diversity. We also show how this alteration inspired by Evidential Reasoning allows for more adaptive resource allocation within the particle filter to improve the overall performance.

REFERENCES

[1] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, "PoseCNN: A convolutional neural network for 6D object pose estimation in cluttered scenes," in *Robotics: Science and Systems (RSS)*, 2018.

[2] J. Tremblay, T. To, B. Sundaralingam, Y. Xiang, D. Fox, and S. Birchfield, "Deep object pose estimation for semantic robotic grasping of household objects," in *Conference on Robot Learning (CoRL)*, 2018.

[3] X. Deng, A. Mousavian, Y. Xiang, F. Xia, T. Bretl, and D. Fox, "PoseRBPF: A Rao-Blackwellized particle filter for 6D object pose tracking," in *Robotics: Science and Systems (RSS)*, 2019.

[4] G. Shafer, *A mathematical theory of evidence*. Princeton university press, 1976, vol. 42.

[5] C. Wang, D. Xu, Y. Zhu, R. Martín-Martín, C. Lu, L. Fei-Fei, and S. Savarese, "DenseFusion: 6D object pose estimation by iterative dense fusion," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3343–3352.

[6] Y. He, H. Huang, H. Fan, Q. Chen, and J. Sun, "FFB6D: A full flow bidirectional fusion network for 6D pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 3003–3013.

[7] C. Wang, R. Martín-Martín, D. Xu, J. Lv, C. Lu, L. Fei-Fei, S. Savarese, and Y. Zhu, "6-PACK: Category-level 6D pose tracker with anchor-based keypoints," in *International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 10 059–10 066.

[8] D. Fox, W. Burgard, F. Dellaert, and S. Thrun, "Monte carlo localization: Efficient position estimation for mobile robots," *AAAI/IAAI*, vol. 1999, no. 343-349, pp. 2–2, 1999.

[9] Z. Sui, O. C. Jenkins, and K. Desingh, "Axiomatic particle filtering for goal-directed robotic manipulation," in *International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 4429–4436.

[10] Z. Sui, Z. Zhou, Z. Zeng, and O. C. Jenkins, "SUM: Sequential scene understanding and manipulation," in *International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 3281–3288.

[11] S. Li, S. Koo, and D. Lee, "Real-time and model-free object tracking using particle filter with joint color-spatial descriptor," in *International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 6079–6085.

[12] C. Choi and H. I. Christensen, "Robust 3d visual tracking using particle filtering on the special euclidean group: A combined approach of keypoint and edge features," *The International Journal of Robotics Research*, vol. 31, no. 4, pp. 498–519, 2012.

[13] J. Deutscher, A. Blake, and I. Reid, "Articulated body motion capture by annealed particle filtering," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, vol. 2. IEEE, 2000, pp. 126–133.

[14] S. Thrun, D. Fox, W. Burgard, and F. Dellaert, "Robust monte carlo localization for mobile robots," *Artificial intelligence*, vol. 128, no. 1-2, pp. 99–141, 2001.

[15] S. Lenser and M. Veloso, "Sensor resetting localization for poorly modelled mobile robots," in *International Conference on Robotics and Automation (ICRA)*, vol. 2, 2000, pp. 1225–1232.

[16] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2005.

[17] J.-S. Gutmann and D. Fox, "An experimental comparison of localization methods continued," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 1. IEEE, 2002, pp. 454–459.

[18] N. Akai, T. Hirayama, and H. Murase, "Hybrid localization using model-and learning-based methods: Fusion of monte carlo and e2e localizations via importance sampling," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 6469–6475.

[19] J. Y. Halpern, *Reasoning about uncertainty*. MIT press, 2017.

[20] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan kaufmann, 1988.