Tactile Exploration with Particle-Based Belief Entropy

Lara Brudermmüller*, Julius Jankowski†, Sylvain Calinon†, Marc Toussaint‡ and Nick Hawes*

*Oxford Robotics Institute, University of Oxford, United Kingdom, larab,nickh@robots.ox.ac.uk
†Idiap Research Institute, Ecole Polytechnique Fédérale de Lausanne, Switzerland, firstname.lastname@idiap.ch
‡Learning and Intelligent Systems Lab, TU Berlin, Germany, toussaint@tu-berlin.de

***Abstract*—This paper extends belief space control to continuous, particle-based belief states in the context of contact-rich manipulation under uncertainty with sparse tactile feedback. We answer the open question of how to quantify information gain for a continuous particle-based belief by proposing a new approximation for the entropy of a particle belief that also captures the object-robot interaction dynamics. Moreover, we address the challenge of the discontinuous and sparse nature of the measurement signal by proposing a sampling-based information-gathering controller that selects the next best action from a set of sampled candidate trajectories based on the approximated entropy of predicted future belief states. In robot experiments, we show that action selection based on the approximated particle entropy significantly improves the information-gathering process in terms of efficiency and success rate of a subsequent grasp.**

I. Introduction

Successful manipulation requires tactile feedback [5, 24], yet many robots only have access to limited tactile sensors alongside other sensors which may be noisy or potentially occluded when performing manipulation tasks. Therefore, *information-gathering* actions are crucial to improve the robot's understanding of the environment. Planning in such partially-observable and uncertain domains requires the robot to reason about the probability distribution over the underlying state, i.e. the *belief state*, and how it changes over time when the robot interacts with its environment. Such problems are generally formalised as partially-observable Markov decision processes (POMDPs) [14], and more recently as belief space control problems [22, 23]. Generating information-gathering action sequences requires planning in belief space, but the increased dimensionality of this space creates a more complex planning problem. This complexity is often reduced by making approximations, such as constraining the belief space to Gaussian distributions [21]. However, this is not suitable for contact-rich manipulation scenarios, as the contact dynamics are highly non-linear and multi-modal [23]. Other approaches reduce the complexity of the belief space control problem by discretising the state space, but this only supports reasoning about mild uncertainty [7, 16]. It hence remains challenging to find information-gathering plans in continuous and non-Gaussian belief spaces. Moreover a planner needs to handle the discontinuous and sparse nature of a contact measurement signal, i.e. contact or no-contact. Motivated by the above, this paper extends belief space control algorithms

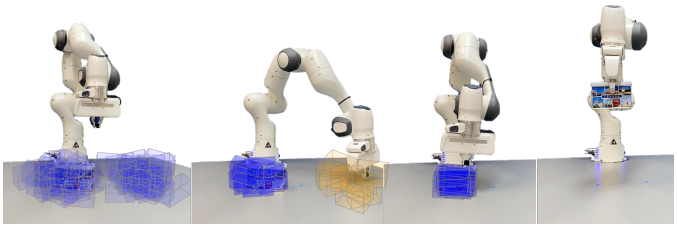


Fig. 1: Experimental setup of blind grasping. *From left to right:* initial bi-modal particle belief with uniform weights (■); information-gathering trajectory rejecting particle hypotheses (■); converged belief after contact; successful grasp.

to *continuous particle-based* belief states in the context of contact-rich manipulation with sparse tactile feedback. We make following contributions: *i)* a *new approximation for the entropy of a particle belief* that also captures the object-robot interaction dynamics; and *ii)* a *sampling-based information-gathering controller* that selects the next best action from a set of sampled trajectories based on the approximated entropy of predicted future belief states. In robot experiments (cf. setup in Fig. 1), we show that action selection based on the approximated particle entropy significantly improves the information-gathering process in terms of efficiency and success rate of a subsequent open-loop grasp.

II. Related Work

Dexterous manipulation under uncertainty is particularly challenging as state and action spaces are continuous, evaluating a physics model of the contact dynamics is computationally expensive, and observations from contact sensors are inherently discontinuous [16]. A common approach is to plan for a sequence of "move-until-touch" actions [7, 13, 19] to localise an object, followed by an open-loop grasp. The policy is split into an information-gathering (exploration) phase followed by a goal-directed (exploitation) phase. A similar strategy has been used in belief space control approaches for blind grasping [22, 23], where the exploration actions are found by optimizing a cost function that includes an information-theoretic measure of the belief state. Other works switch between exploration and exploitation phases based on the uncertainty of the belief state [18]. This trade-off between information gathering and goal-directed behaviour has been solved more formally within the context of POMDPs [16, 8, 10, 27]. Yet, all these approaches

are subject to at least one of the following two limitations: *i)* the problem space is typically subject to discretisation for means of tractability, which limits the applicability of such methods to cases with only mild uncertainty; and *ii)* most approaches ignore the dynamics of the object-robot interaction, by assuming either a static object [22, 29, 7, 13, 18] or only very limited dynamics [10]. While the first limitation is a common one in many POMDP and belief space control settings, the second limitation is particularly relevant for the computation of information-theoretic measures, such as the relative entropy, which typically does not account for the fact that the robot actions can also increase uncertainty, e.g. by pushing an object over. While the dynamics have been considered in the context of POMDPs [16, 8], the problem is typically solved in a discrete state space, which limits the complexity of the problem. Consequently, our work addresses the two limitations above by using a continuous non-parametric representation of the belief state and by incorporating the object-robot interaction dynamics in the planning process, which we capture in a new approximation for the entropy of a particle belief.

III. Problem Formulation

This paper addresses the problem of blindly grasping an object o using highly sparse tactile feedback via belief space control. In order to reach a state where grasping is likely to succeed, the robot needs to perform information-gathering actions w.r.t. the object location. Let $\boldsymbol{x}_k = (\boldsymbol{q}^r, \boldsymbol{q}^o) \in \mathcal{R}^{(n^r_{\text{dof}}+n^o_{\text{dof}})}$ describe the state of the underactuated robotic system, which includes the pose of the object $\boldsymbol{q}^o_k \in \mathbb{R}^{n^o_{\text{dof}}}$, and the robot configuration $\boldsymbol{q}^r_k \in \mathbb{R}^{n^o_{\text{dof}}}$ at time step k. We can only control the object configuration by executing robot control commands $\boldsymbol{u}_k \in \mathbb{R}^{n^r_{\text{dof}}}$. The system changes over time according to stochastic dynamics $\boldsymbol{x}_{k+1} \sim p(\cdot|\boldsymbol{x}_k, \boldsymbol{u}_k)$. Measurements $\boldsymbol{z}_k$ are taken by the system that follow a stochastic model $\boldsymbol{z}_k \sim p(\cdot|\boldsymbol{x}_k)$. Due to the uncertainty and partial observability of the system, the system state is a continuous random variable. The system's *belief* of its state $b_k{=}p(\boldsymbol{x}_k|\boldsymbol{u}_{0:k-1}, \boldsymbol{z}_{1:k}, b_0)$ depends on the history of actions and measurements taken, as well as the initial belief b_0.

For a given sequence of control actions $\boldsymbol{u}_{0:k-1}$, the system may estimate information gain $\text{IG}(b_0, \boldsymbol{u}_{0:k-1})$ using its model of the dynamics and its sensor model. We formulate the problem of information gathering as trajectory optimization in belief space:

$$\max_{\boldsymbol{u}_{0:K-1}} \quad \text{IG}(b_0, \boldsymbol{u}_{0:k-1}). \tag{1}$$

Estimating the information gain is challenging as it involves predicting possible future measurements. Hence, in the following, we present our approach to estimating information gain for an information-gathering controller.

IV. Information-Gathering Controller

We propose a belief-based controller that samples candidate trajectories that are likely to make contact with the object according to the current belief. The controller selects the next best trajectory based on its expected information gain. In order to predict the expected information gain for each sampled trajectory, we need to anticipate future belief states given the current belief state and the trajectory. The idea of generating candidates and selecting the best one according to the information gain is similar to the approach in [7], and more generally to the concept of *predictive sampling* [9]. Yet, to apply these ideas to continuous belief space control for blind grasping we need to solve two open questions: *i)* how to quantify information gain for a continuous particle-based belief, and *ii)* how to consider object-robot interaction dynamics that possibly reduce or increase uncertainty.

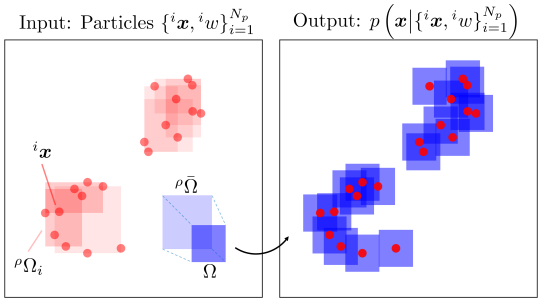


Fig. 2: *Computation of uniform kernel* Ω that is used to construct the probability density function p for a set of weighted particles (●). After fitting a tight bounding box ${}^\rho\Omega_i$ (■) to the $\rho = 6$ nearest neighbors of each particle, we scale the mean bounding box ${}^\rho\bar{\Omega}$ to obtain the uniform kernel Ω (■) that is used across all particles.

A. Particle-based Entropy Approximation

This section addresses the first open question by proposing a new approximation of the relative entropy for a particle-based belief. The differential entropy of a random, continuous variable $\boldsymbol{x}$ following a probability distribution $p(\boldsymbol{x})$ is defined as

$$\text{H}[\boldsymbol{x}] = -\int_{\mathcal{X}} p(\boldsymbol{x})\log(p(\boldsymbol{x}))d\boldsymbol{x}. \tag{2}$$

However, the entropy of a set of particles is not clearly defined. This is due to the fact that there is no unique way to define a probability density function that is parameterized by N_p weighted samples $\{{}^i\boldsymbol{x}, {}^iw\}_{i=1}^{N_p}$. Without considering dynamics, the particle-based belief may be treated as a discrete probability distribution with the weights representing the probability of each particle. The corresponding entropy is then given by $\hat{\text{H}}_w[\boldsymbol{x}] = -\sum_i {}^iw\,\log({}^iw)$. While this approximation captures information gained through observations, it does not capture the spatial density of particles, i.e. the local distance between the states of the particles. We thus propose to model the continuous belief with a uniform mixture distribution

$$b = p\left(\boldsymbol{x}\middle|\{{}^i\boldsymbol{x}, {}^iw\}_{i=1}^{N_p}\right) = \sum_i {}^iw\,\mathcal{U}(\boldsymbol{x} - {}^i\boldsymbol{x}, \Omega), \tag{3}$$

where $\Omega \subseteq \mathcal{X}$, the *kernel* of the uniform components, is a bounded region in state space with hypervolume $V(\Omega){=}\int_\Omega 1d\boldsymbol{x}$. The probability density of the uniform distribution $\mathcal{U}(\boldsymbol{x}, \Omega)$ is equal to ${}^1/_{V(\Omega)}$ if $\boldsymbol{x} \in \Omega$ and equal to

zero otherwise. We impose that all particles have the same shared kernel. The kernel Ω is computed in three steps, which are additionally illustrated in Fig. 2: 1) For each particle i, fit a tight bounding box ${}^{\rho}\Omega_i$ to the ρ nearest neighbors. 2) Compute the average bounding box ${}^{\rho}\bar{\Omega}$ by taking the mean size in each dimension. 3) Scale the bounding box based on the number of dimensions d and the number of nearest neighbors ρ with

$$\Omega = {}^{\rho}\bar{\Omega}\left(\sqrt[d]{\rho} - 1\right)^{-1}. \tag{4}$$

The kernel Ω centred at a particle state is the average support of the uniform distribution contributed by that particle. Hence, Ω encodes the density of particles and can be used for computing the entropy: the higher the density of the particles, the lower the hypervolume $V(\Omega)$. The approximation parameter ρ tunes the locality of the approximation of the belief based on particles. A small ρ allows us to approximate multi-modal beliefs by approximating the density locally. A large ρ takes into account the global density of the particle set and thus does not capture clusters of particles. This is in contrast to the work in [18], where all particles are used to compute a shared Gaussian kernel. Using the mixture distribution in Eq. (3), the entropy cannot be evaluated efficiently in closed-form due to overlaps of the uniform components. However, we compute an upper-bound $\hat{\mathrm{H}}[\boldsymbol{x}] \geq \mathrm{H}[\boldsymbol{x}]$ of the differential entropy in Eq. (2) analytically by using the inequality $\log(p(\boldsymbol{x})) \geq \log(p_i(\boldsymbol{x}))$, where $p_i(\boldsymbol{x})$ is the probability density contribution of the i-th component. As a result, the upper-bound of the differential entropy of the particles is given by

$$\hat{\mathrm{H}}[\boldsymbol{x}] = -\sum_i {}^{i}w \log({}^{i}w) + \log(V(\Omega)). \tag{5}$$

The approximation in Eq. (5) has a term that only depends on the particle weights, i.e. $\hat{\mathrm{H}}_w[\boldsymbol{x}] = -\sum_i {}^{i}w \log({}^{i}w)$, and a term that depends on the local density of the particles, i.e. $\hat{\mathrm{H}}_\Omega[\boldsymbol{x}] = \log(V(\Omega))$. We recover the mutual information $\mathrm{I}(\boldsymbol{x}_k; \boldsymbol{z}_k)$ of the system state $\boldsymbol{x}_k$ and the measurement $\boldsymbol{z}_k$, i.e. the amount of information obtained about $\boldsymbol{x}$ when observing $\boldsymbol{z}$ at time step k:

$$\begin{aligned}\mathrm{I}(\boldsymbol{x}_k; \boldsymbol{z}_k) &= \mathrm{H}[\boldsymbol{x}_k | \boldsymbol{x}_{k-1}, \boldsymbol{u}_{k-1}] - \mathrm{H}[\boldsymbol{x}_k | \boldsymbol{z}_k, \boldsymbol{x}_{k-1}, \boldsymbol{u}_{k-1}] \\ &= \hat{\mathrm{H}}_w[\boldsymbol{x}_{k-1}] - \hat{\mathrm{H}}_w[\boldsymbol{x}_k].\end{aligned} \tag{6}$$

As a result, the entropy difference between two consecutive time steps $\Delta\hat{\mathrm{H}}_{k-1,k} = \hat{\mathrm{H}}[\boldsymbol{x}_{k-1}] - \hat{\mathrm{H}}[\boldsymbol{x}_k]$ can be written as

$$\Delta\hat{\mathrm{H}}_{k-1,k} = \mathrm{I}(\boldsymbol{x}_k; \boldsymbol{z}_k) + \log\left(\frac{V(\Omega_{k-1})}{V(\Omega_k)}\right). \tag{7}$$

If the dynamics of the system keep the particle density constant, i.e. $\Omega_{k-1} = \Omega_k$, the entropy change is equal to the information gained by taking a measurement $\boldsymbol{z}$. A detailed derivation of the particle entropy approximation and quantitative experiments proving the validity of the approximation as an upper-bound to the true entropy of the particles can be found in Sec. A in the Appendix.

Limitation of the Differential Entropy: While the differential entropy measure is an attempt to extend the concept of entropy to continuous random variables, it has some limitations. The main one is that the differential entropy will become minus infinity if the probability distribution collapses in one of its dimensions. This problem is also reflected in the density-based term $\hat{\mathrm{H}}_\Omega[\boldsymbol{x}]$ in the entropy approximation in Eq. (5), as the hypervolume of the uniform kernel Ω will become zero if the particles collapse in one of the dimensions. Hence, in practice, we will use the weight-based term $\hat{\mathrm{H}}_w[\boldsymbol{x}]$ and the density-based term $\hat{\mathrm{H}}_\Omega[\boldsymbol{x}]$ as separate cost-terms in the cost function of the belief space control problem. Moreover, we reformulate the density-based term as the sum over the dimensions of the bounding box Ω instead of computing its hypervolume. More precisely, with Ω_i being the size of the bounding box in the i-th dimension, we reformulate the density-based term as

$$\hat{\mathrm{H}}_\Omega[\boldsymbol{x}] = \log\left(\frac{1}{d}\sum_{i=1}^{d}\Omega_i\right), \tag{8}$$

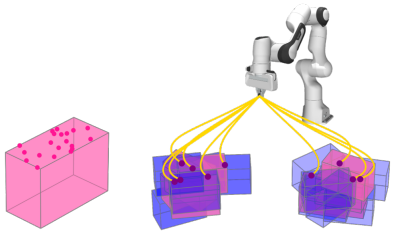
Algorithm 1: Prediction of Information Gain

Input: Belief $\{{}^{i}\boldsymbol{x}_0, {}^{i}w_0\}_{i=1}^{N_p}$, Controls $\boldsymbol{u}_{0:K-1}$
Output: Max. information gain $\Delta\hat{\mathrm{H}}$

1 $\hat{\mathrm{H}}_{w,0} \leftarrow -\sum_i {}^{i}w_0 \log({}^{i}w_0)$
2 **for** $k = 0, \ldots, K-1$ **do**
3 Sample transitions: ${}^{i}\boldsymbol{x}_{k+1} \sim p(\cdot|{}^{i}\boldsymbol{x}_k, \boldsymbol{u}_k), \quad \forall i$
4 Compute expected weight term
5 $E[\hat{\mathrm{H}}_{w,k+1}] \leftarrow$ (9)
6 Compute particle density:
7 $\hat{\mathrm{H}}_\Omega[\boldsymbol{x}_{k+1}] \leftarrow$ (8)
8 Compute expected entropy reduction:
  $\Delta\hat{\mathrm{H}}_{0,k+1} \leftarrow \hat{\mathrm{H}}_{w,0} - E[\hat{\mathrm{H}}_{w,k+1}] + \hat{\mathrm{H}}_\Omega[\boldsymbol{x}_0] + \hat{\mathrm{H}}_\Omega[\boldsymbol{x}_{k+1}]$
9 **end**
10 $\Delta\hat{\mathrm{H}} \leftarrow \max_k \Delta\hat{\mathrm{H}}_{0,k+1}$

B. Expected Information Gain

Given the new measure of the entropy of a particle-based belief state, we estimate the expected information gain of a candidate trajectory $\boldsymbol{u}_{0:K-1}$ by predicting future belief states and their associated expected relative entropy based on hypothetical measurements generated from the predicted belief states, as summarised in Alg. 1. While rolling out the dynamics for each candidate is straight-forward (*Alg. 1, lines 3-4*), updating the weights of the particles based on the measurement model is more challenging, as it requires predicting future measurements $\hat{\boldsymbol{z}}_{0:K-1}$ (*Alg. 1, lines 4-5*). A common approach to predicting measurements the robot observes during a candidate trajectory is to use the maximum likelihood estimate of the belief distribution [21, 17, 4, 23]. However, if the particle weights are all similar, as frequently happens after a resampling step within a particle filter, using exactly one state as a hypothesis for the real system state creates a possibly unreasonably strong bias towards a single point in the state space. Consequently, we propose to use every particle in the particle set as a hypothetical underlying state of the system to generate N_p hypothetical measurements in each time step. When a particle is used as a hypothetical state, the measurement is generated by a simple check if the robot is in contact at the given time step for that particle. Its probability is then computed using the measurement model



a) Sample target points b) Compute candidates

Fig. 3: *Candidate Sampling:* Starting from the current belief (■), particles are sampled using importance sampling (■). Uniformly sampled points within the object volume are processed through inverse kinematics to derive final robot configurations for candidate trajectories towards the object.

$P(z_k|\boldsymbol{x}_k)$ from the particle filter. We marginalize over all particles and weight the contribution of each particle to the expected entropy reduction based on the particle weights at time step 0. Note that this procedure only concerns the weight term in the entropy approximation, as the density term only depends on the dynamics of the system and the particle set, but not on the measurements. Hence the expected entropy weight term at time step k is computed as

$$E[\hat{\mathrm{H}}_{w,k}] = \sum_j \left(-\sum_i {}^{i,j}\hat{w}_k \log({}^{i,j}\hat{w}_k) \right) {}^j w_0, \tag{9}$$

with ${}^{i,j}\hat{w}_k$ being the updated weight estimate of the ith particle at time step k when assuming the jth particle as the underlying state of the system, i.e.

$${}^{i,j}\hat{w}_k = \frac{{}^i w_0 \; P(z_k|{}^j\boldsymbol{x}_k)}{\sum_l {}^l w_0 \; P(z_k|{}^l\boldsymbol{x}_k)}. \tag{10}$$

The particle density H_Ω is computed using the predicted belief state $\boldsymbol{x}_{k+1}$ (*Alg. 1, lines 6-7*). Thus, for each time step we compute the expected entropy reduction as the sum of the expected entropy weight term and the particle density in the given time step (*Alg. 1, lines 8-9*). Eventually, the maximum expected information gain is computed as the maximum expected entropy reduction over all time steps. The best candidate trajectory is pruned at the time step with the maximum expected information gain, which avoids executing parts of the trajectory that gain no more information or even increase the belief entropy due to the dynamics of the system.

C. Belief-Dependent Candidate Sampling

While our approach does not make any assumptions about the specific form of candidate trajectories, we sample trajectories that are likely to result in contact events. Given the current belief state b_k at time step k, we sample N_{samples} particles via importance sampling. For each particle, we uniformly sample a point within the object volume and compute the inverse kinematics solution such that the robot's end-effector reaches the sampled point. Adopting the spline-based representation from [11] (cf. Appendix Sec. D), we construct candidate trajectories $\boldsymbol{u}^l_{0:K-1}$ that end in these robot configurations, as illustrated in Fig. 3.

TABLE I: Results of proposed approach compared to baseline.

Models	Success rate [%]	Avg. num. iterations
Proposed	**80**	10.5
Baseline	50	15

V. Experiments

We evaluate the proposed method in a real robot setup, where a Franka Emika Panda robot is tasked to localise and grasp a box in its workspace with solely sparse tactile feedback, given a prior bi-modal particle belief about the box pose. After every iteration of the information-gathering controller, the probability of grasp success is estimated on the current belief. If it is above a given threshold, the robot attempts an open-loop grasp of the object. The experimental setup is shown in Fig. 1. While the focus of this paper is on the information-gathering controller, the underlying assumption is that the robot is able to accurately track its belief state through a particle filter. We provide a novel measurement model and an adapted resampling strategy designed for sparse binary contact measurements and corresponding ablation studies concerning the particle filter, alongside an algorithmic overview and additional implementation details in the Appendix. We compare our information-gathering controller to an "uninformed" baseline that is the same as our system but without access to the information gain metric when choosing trajectories. This is a strong baseline, as opposed to a non-belief-informed random exploration baseline, such as uniformly sampling trajectories. We recorded 10 runs on the robot for each method with random initial box poses and a fixed maximum number of iterations (15). As metrics, we use the success rate of grasping and the average number of iterations until a successful grasp. The results, summarised in Table I, show that the proposed method significantly outperforms the uninformed baseline in terms of success rate and efficiency. We include a video summarising the experiments in the supplementary material.

VI. Conclusion

This paper proposes a novel information-gathering controller for contact-rich manipulation tasks based on the approximated entropy of a particle-based belief state. We have demonstrated the effectiveness of the proposed method in a real robot setup for blind grasping under uncertainty. The results show that the proposed entropy approximation is a suitable measure for the expected information gain of a candidate control sequence, significantly outperforming an uninformed baseline in terms of success rate and efficiency. Yet, we acknowledge several limitations alongside possible future directions. The real-time applicability is limited by the computational complexity of the information-gathering controller, as rolling out the dynamics for each candidate control trajectory is still computationally expensive. We believe that learning the contact dynamics from data, replacing the physics engine, could be a promising future direction. This would remove required knowledge about object properties, such as mass and friction, reduce the computational complexity of the information-gathering controller, and would also render

the dynamics stochastic, which possibly reflects the inherent multi-modality of the contact dynamics in a more realistic way. Last, we believe that an exciting future direction is to extend the information-theoretic approach by not only considering actions that generate sensory information about the environment, but also actions that funnel uncertainty about the system into smaller regions of the state space, such as in [3, 2] or more recently [12].

References

[1] Craig Corcoran and Robert Platt. A measurement model for tracking hand-object state during dexterous manipulation. In *2010 IEEE International Conference on Robotics and Automation*, pages 4302–4308. IEEE, 2010.

[2] Mehmet Remzi Dogar, Kaijen Hsiao, Matei T Ciocarlie, and Siddhartha S Srinivasa. Physics-based grasp planning through clutter. In *Robotics: Science and systems*, volume 8, pages 57–64, 2012.

[3] Michael A Erdmann and Matthew T Mason. An exploration of sensorless manipulation. *IEEE Journal on Robotics and Automation*, 4(4):369–379, 1988.

[4] Tom Erez and William D Smart. A scalable method for solving high-dimensional continuous pomdps using local approximation. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, pages 160–167, 2010.

[5] Nima Fazeli, Miquel Oller, Jiajun Wu, Zheng Wu, Joshua B Tenenbaum, and Alberto Rodriguez. See, feel, act: Hierarchical learning for complex manipulation skills with multisensory fusion. *Science Robotics*, 4(26), 2019.

[6] Gianluca Garofalo, Nico Mansfeld, Julius Jankowski, and Christian Ott. Sliding mode momentum observers for estimation of external torques and joint acceleration. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 6117–6123. IEEE, 2019.

[7] Paul Hebert, Thomas Howard, Nicolas Hudson, Jeremy Ma, and Joel W Burdick. The next best touch for model-based localization. In *2013 IEEE International Conference on Robotics and Automation*, pages 99–106. IEEE, 2013.

[8] Matanya Horowitz and Joel Burdick. Interactive non-prehensile manipulation for grasping via pomdps. In *2013 IEEE International Conference on Robotics and Automation*, pages 3257–3264. IEEE, 2013.

[9] Taylor Howell, Nimrod Gileadi, Saran Tunyasuvunakool, Kevin Zakka, Tom Erez, and Yuval Tassa. Predictive sampling: Real-time behaviour synthesis with mujoco. *arXiv preprint arXiv:2212.00541*, 2022.

[10] Kaijen Hsiao, Leslie Pack Kaelbling, and Tomás Lozano-Pérez. Task-driven tactile exploration. In *Robotics: science and systems*, volume 12, 2010.

[11] Julius Jankowski, Lara Brudermüller, Nick Hawes, and Sylvain Calinon. Vp-sto: Via-point-based stochastic trajectory optimization for reactive robot behavior. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10125–10131. IEEE, 2023.

[12] Julius Jankowski, Lara Brudermüller, Nick Hawes, and Sylvain Calinon. Planning for robust open-loop pushing: Exploiting quasi-static belief dynamics and contact-informed optimization. *arXiv preprint arXiv:2404.02795*, 2024.

[13] Shervin Javdani, Matthew Klingensmith, J Andrew Bagnell, Nancy S Pollard, and Siddhartha S Srinivasa. Efficient touch based localization through submodularity. In *2013 IEEE International Conference on Robotics and Automation*, pages 1828–1835. IEEE, 2013.

[14] Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101(1-2):99–134, 1998.

[15] Michael C Koval, Nancy S Pollard, and Siddhartha S Srinivasa. Pose estimation for planar contact manipulation with manifold particle filters. *The International Journal of Robotics Research*, 34(7):922–945, 2015.

[16] Michael C Koval, Nancy S Pollard, and Siddhartha S Srinivasa. Pre-and post-contact policy decomposition for planar contact manipulation under uncertainty. *The International Journal of Robotics Research*, 35(1-3):244–264, 2016.

[17] Scott A Miller, Zachary A Harris, and Edwin KP Chong. Coordinated guidance of autonomous uavs via nominal belief-state optimization. In *2009 American Control Conference*, pages 2811–2818. IEEE, 2009.

[18] Ekaterina Nikandrova, Jonna Laaksonen, and Ville Kyrki. Towards informative sensor-based grasp planning. *Robotics and Autonomous Systems*, 62(3):340–354, 2014.

[19] Anna Petrovskaya and Oussama Khatib. Global localization of objects via touch. *IEEE Transactions on Robotics*, 27(3):569–585, 2011.

[20] Anna Petrovskaya, Oussama Khatib, Sebastian Thrun, and Andrew Y Ng. Bayesian estimation for autonomous object manipulation based on tactile sensors. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, pages 707–714. IEEE, 2006.

[21] Robert Platt, Russ Tedrake, Leslie Pack Kaelbling, and Tomas Lozano-Perez. Belief space planning assuming maximum likelihood observations. In *Robotics: Science and Systems*, 2010.

[22] Robert Platt, Leslie Kaelbling, Tomas Lozano-Perez, and Russ Tedrake. Simultaneous localization and grasping as a belief space control problem. In *International Symposium on Robotics Research*, volume 2, 2011.

[23] Robert Platt, Leslie Kaelbling, Tomas Lozano-Perez, and Russ Tedrake. Efficient planning in non-gaussian belief spaces and its application to robot grasping. In *Robotics Research*, pages 253–269. Springer, 2017.

[24] Alberto Rodriguez. The unstable queen: Uncertainty, mechanics, and tactile feedback. *Science Robotics*, 6(54), 2021.

[25] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. Intelligent Robotics and Autonomous Agents

series. MIT Press, 2005.

[26] Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ international conference on intelligent robots and systems*, pages 5026–5033. IEEE, 2012.

[27] Florian Wirnshofer, Philipp Sebastian Schmitt, Georg von Wichert, and Wolfram Burgard. Controlling contact-rich manipulation under partial observability. In *Robotics: Science and Systems*, 2020.

[28] Li Zhang and Jeffrey C Trinkle. The application of particle filtering to grasping acquisition with visual occlusion and tactile sensing. In *2012 IEEE International Conference on Robotics and Automation*, pages 3805–3812. IEEE, 2012.

[29] Claudio Zito, Valerio Ortenzi, Maxime Adjigble, Marek Kopicki, Rustam Stolkin, and Jeremy L Wyatt. Hypothesis-based belief planning for dexterous grasping. *arXiv preprint arXiv:1903.05517*, 2019.

APPENDIX

A. Entropy Approximation: Quantitative Evaluation

Figure 4 shows a quantitative evaluation of the performance of the proposed entropy approximation. While the approximation for a newly sampled set of particles can be noisy, it is important to note that the approximation captures the monotonic trend of the ground truth entropy. Even more importantly for reducing uncertainty, the approximation is deterministic and thus captures small changes to a given set of particles.

Upper-bound Entropy Approximation of a Uniform Mixture Distribution: In the following, we derive the upper-bound approximation of the differential entropy

$$\hat{\mathrm{H}}\left[\boldsymbol{x}\right] = -\int_{\mathcal{X}} p(\boldsymbol{x}) \log(p(\boldsymbol{x})) d\boldsymbol{x}, \tag{11}$$

where the probability density function is given as a mixture of uniform distributions

$$p(\boldsymbol{x}) = \sum_i {}^i w \, \mathcal{U}(\boldsymbol{x} - {}^i\boldsymbol{x}, \Omega). \tag{12}$$

Each uniform component is defined as

$$\mathcal{U}(\boldsymbol{x}, \Omega) = \begin{cases} \frac{1}{V} & \boldsymbol{x} \in \Omega \\ 0 & \text{otherwise} \end{cases}, \tag{13}$$

where $V = \int_\Omega 1 d\boldsymbol{x}$ is the hypervolume of the common kernel Ω. In the following, the uniform components are abbreviated with $\mathcal{U}_i(\boldsymbol{x}) = \mathcal{U}(\boldsymbol{x} - {}^i\boldsymbol{x}, \Omega)$.

By inserting Eq. (12) into Eq. (11), one obtains

$$\hat{\mathrm{H}}\left[\boldsymbol{x}\right] = -\int_{\mathcal{X}} \left(\sum_i {}^i w \mathcal{U}_i(\boldsymbol{x}) \right) \log \left(\sum_j {}^j w \mathcal{U}_j(\boldsymbol{x}) \right) d\boldsymbol{x}. \tag{14}$$

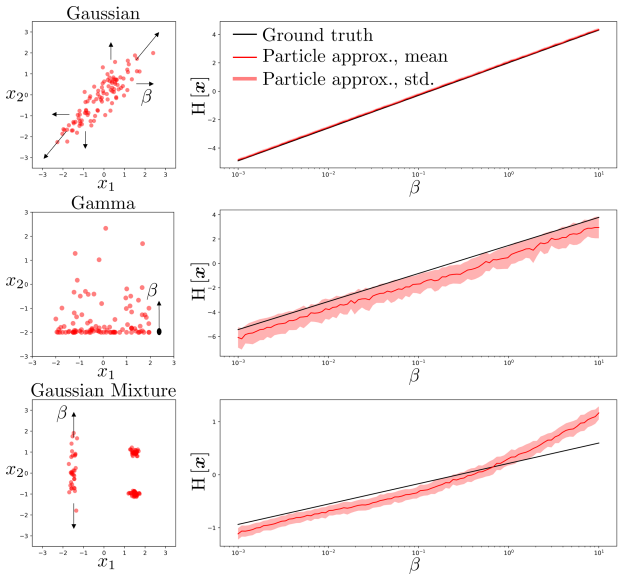


Fig. 4: Particle-entropy approximation for three different ground truth probability distributions. Comparing analytic differential entropy of the ground truth distribution against the differential entropy computed according to Eq. (5) after sampling the particles from the ground truth distribution. Parameter β scales the entropy of the ground truth distributions, shown on the left-hand side for β=1, as follows: **Top row:** The covariance matrix of the Gaussian distribution is scaled by β. **Middle row:** The rate parameter of the Gamma distribution is equal to β. **Bottom row:** The Gaussian mixture distribution has two fixed components on the right-hand side and one component on the left-hand side where β scales the variance in y-direction.

This can be rewritten by pulling the first sum out of the integral:

$$\hat{\mathrm{H}}\left[\boldsymbol{x}\right] = -\sum_i {}^i w \int_{\mathcal{X}} \mathcal{U}_i(\boldsymbol{x}) \log \left(\sum_j {}^j w \mathcal{U}_j(\boldsymbol{x}) \right) d\boldsymbol{x}. \tag{15}$$

Due to Eq. (13), the integration domain can be reduced to $\Omega_i = \Omega - \boldsymbol{x}_i$ and thus

$$\hat{\mathrm{H}}\left[\boldsymbol{x}\right] = -\sum_i {}^i w \int_{\Omega_i} \frac{1}{V} \log \left(\sum_j {}^j w \mathcal{U}_j(\boldsymbol{x}) \right) d\boldsymbol{x}. \tag{16}$$

The log-term can not be simplified any further and thus the integral can only be approximated, e.g. by numerical integration or via importance sampling. Instead, we use a lower-bound approximation of the integral for each individual component, i.e.

$$\int_{\Omega_i} \frac{1}{V} \log \left(\sum_j {}^j w \mathcal{U}_j(\boldsymbol{x}) \right) d\boldsymbol{x} \geq \int_{\Omega_i} \frac{1}{V} \log \left({}^i w \frac{1}{V} \right) d\boldsymbol{x}. \tag{17}$$

Note that this inequality becomes an equality if the uniform components do not overlap, i.e. $\Omega_i \cap \Omega_j = \emptyset, \forall i \neq j$. By

inserting the inequality (17) in Eq. 16, one obtains

$$\hat{\mathrm{H}}\left[\boldsymbol{x}\right] \leq -\sum_i {}^i w \int_{\Omega_i} \frac{1}{V}\left(\log {}^i w + \log\frac{1}{V}\right) d\boldsymbol{x}. \tag{18}$$

Since the integrand is now constant over the integral domain, the integral can be resolved and one obtains

$$\hat{\mathrm{H}}\left[\boldsymbol{x}\right] \leq -\sum_i {}^i w \left(\log {}^i w + \log\frac{1}{V}\right). \tag{19}$$

Further simplifications and using the equality $\sum_i {}^i w = 1$, we obtain the final inequality

$$\hat{\mathrm{H}}\left[\boldsymbol{x}\right] \leq -\sum_i {}^i w \log {}^i w + \log V. \tag{20}$$

B. Overview of Algorithmic Framework

Fig. 5 outlines the full proposed algorithmic framework for blind grasping under uncertainty. The robot iteratively refines its belief about the object pose through a particle filter and an information-gathering controller. The termination criterion for this exploration phase is based on the probability of grasp success, which is estimated on the updated belief after each iteration of the information-gathering controller. If the probability of grasp success is above a given threshold, the robot transitions to the grasping phase.

C. Particle Filter for Continuous Object Pose Estimation through Contact

The information-gathering controller builds on the ability to accurately track the current belief, as well as being able to predict future belief states. Given the continuous belief space control problem in Eq. (1), we use a particle filter to represent and track the belief state. Thus, the belief state is represented by a set of N_p particles $\{{}^i\boldsymbol{x}, {}^i w\}_{i=1}^{N_p}$, where each particle consists of a state sample ${}^i\boldsymbol{x}$ and a corresponding weight ${}^i w \in [0,1]$ subject to $\sum_i {}^i w{=}1$. As is standard for a particle filter, the belief is updated by applying a dynamics model and weighting the particles according to the likelihood of the observed measurements, defined by the measurement model [25]. For the dynamics model, assuming that we know the physical properties of the object, we simulate the contact dynamics between the robot and the object using a physics engine, such as MuJoCo [26][1]. While these dynamics are non-linear, they will be deterministic given the same initial conditions, as the physics engine is deterministic. In contrast to prior work [18, 15], we do not assume that the object pose is constant when the robot is in contact with it. We also note that almost all approaches to tactile exploration use richer sensor modalities, such as tactile sensors that typically provide the point of contact, as well as the contact normal, or force/torque sensors that provide precise measurements of contact forces. In contrast, we use only the torque sensors in the robot's joints, which we found in only one other work by [8] where the uncertainty is much smaller than in our scenario. This setting also implies that we cannot track our belief with standard measurement models [1], as they assume that contact force sensors perfectly determine whether the robot is in contact with the object. Similarly, we cannot use the measurement models from [20, 19] as they only work for tactile sensors. Therefore, we propose a new measurement model for state estimation from noisy binary contact feedback.

Smooth Measurement Model: The measurement model is used to refine the belief state after rolling out the dynamics model with control action $\boldsymbol{u}_k$ through the observed measurement $\boldsymbol{z}_k$. Given the sparse and noisy nature of the contact signal, the binary detection model of an imperfect contact measurement process, i.e. contact or no-contact, proposed in [7] does not account for the fact that the robot might be close to the object without actually making contact. We provide more technical details concerning the actual measurements within our experiments, which are based on the torques measured in the robot's joints, in the appendix. Importantly, unlike other works [15, 10], we do not assume that the contact measurements are discriminative, i.e. accurately distinguish between contact and no-contact. Therefore, we propose a measurement model that incorporates proximity estimates to smooth the discontinuous measurement signal. We model the probability of positive contact such that it decays exponentially with the minimum distance between the robot and the object. While this is similar to the observation model formulated by Zito et al. [29], it does not assume that the robot makes contact with its end-effector, but rather anywhere on the robot's body. The likelihood of a contact measurement $z_k = 1$ given the particle's pose ${}^i\boldsymbol{x}_k$ is given by:

$$P(z_k = 1 | {}^i\boldsymbol{x}_k) = \alpha_{\mathrm{tp}} + (\alpha_{\mathrm{tp}} - \alpha_{\mathrm{fp}}) \exp(-\gamma \cdot d(\boldsymbol{q}_k^r, {}^i\boldsymbol{q}_k^o)), \tag{21}$$

where α_{tp} and α_{fp} are the true and false positive rates of a contact measurement respectively, and γ is the distance-based decay rate of the probability of contact. Based on the assumption that the contact point on the object is the one closest to the robot, the distance $d(\boldsymbol{q}_k^r, \boldsymbol{q}_k^o)$ is computed as the minimum Euclidean distance between the robot configuration ${}^i\boldsymbol{q}_k^r$ and the object configuration ${}^i\boldsymbol{q}_k^o$. The probability of a no-contact measurement is inferred by $1 - P(z_k = 1 | {}^i\boldsymbol{x}_k)$. In our experiments, we show that the proximity-based measurement model leads to more accurate belief tracking with lower chance of particle depletion. Yet, the downside is that it is computationally more expensive than the binary model, as it requires additional distance computations.

Resampling Strategy: While for a sparse binary contact sensor, smoothing measurements with proximity estimates helps spreading the contact observations over a non-infinitesimal region, the sparse nature of the contact signal still leads to a higher likelihood of particle starvation, a problem that has been particularly recognised in the context of contact sensing [15, 28]. Particle starvation occurs when there are no particles in the vicinity of the true state. This often links to a loss of diversity in the particle set, which can lead to a divergence of the particle filter. A common strategy to mitigate particle starvation is resampling, e.g. through importance

[1]The simulation mirrors the impedance control of the robot, which regulates the interaction forces with the environment.

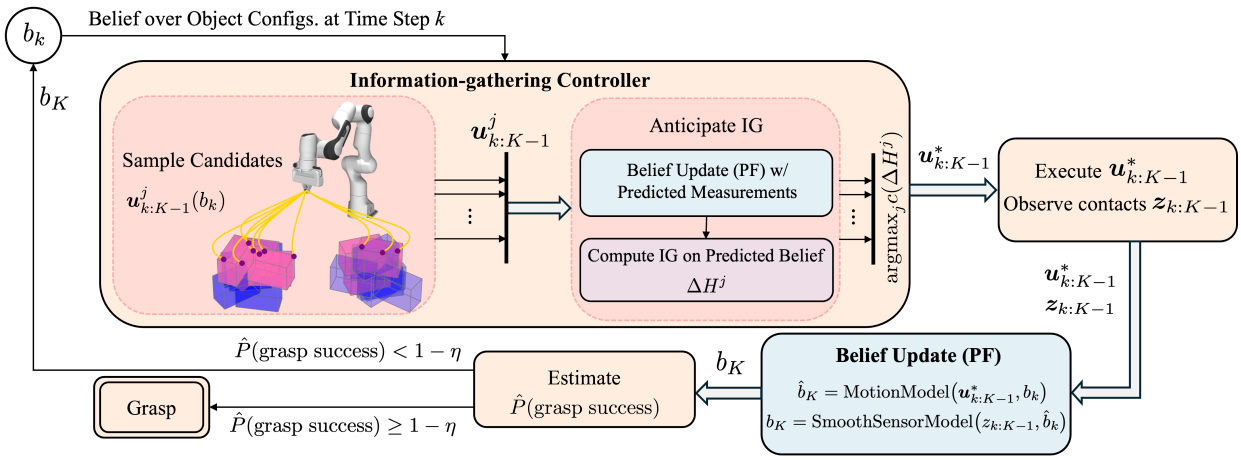


Fig. 5: Proposed pipeline for blind grasping under uncertainty. The robot localises the object through a particle filter and an information-gathering belief space controller. The robot transitions to the grasping phase if the probability of grasp success is above a given threshold.

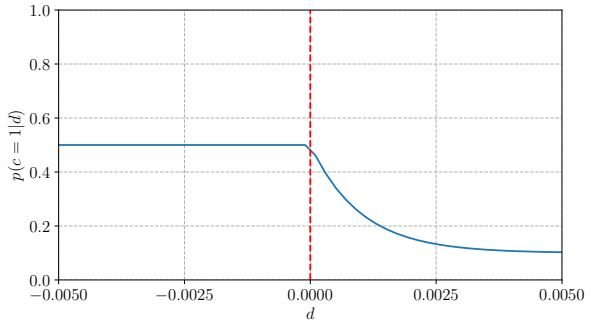


Fig. 6: The probability distribution for a positive contact measurement for the proximity-based $P(z_k|\boldsymbol{x}_k)$ measurement model. The distance on the x-axis is measured as the closest Euclidean distance between the object surface and the robot's last four links.

sampling [25]. Yet in the context of contact-rich manipulation, resampling in the case when contact is observed is more challenging as the likelihood of sampling particles from the contact manifold, i.e. contact points with zero-distance between object and robot, is very low [15]. While Koval et al. [15] offer methods of how to construct such contact manifolds, they are still difficult to construct in more general and high-dimensional spaces. We thus propose a resampling strategy that is based on the estimated proximity of the object particles to the current robot pose and the contact signal. If the measurement was a contact measurement, we resample M new particles based on importance sampling with added noise and compute the probability of the new samples matching the last observation via our smooth measurement model. We then replace the particles that have low weights with the new best samples in terms of the probability of matching the last observation. If the measurement was a no-contact measurement, we perform standard importance sampling. Resampling is triggered if one or more particles have a low weight.

TABLE II: Ablation results of particle filter

	Binary	Smooth	Smooth & Resampled
MSE	0.40 ± 0.97	0.35 ± 0.87	0.45 ± 0.61
Share of AP	0.19 ± 0.20	0.36 ± 0.28	1.00 ± 0.00

Simulation Experiments: We evaluate the proposed smooth sensor model and resampling strategy for sparse binary contact measurements in a 2D simulated environment. In an ablation study, we compare our smooth sensor model to the binary sensor model from Hebert et al. [7] without our resampling strategy and the smooth sensor model with our resampling strategy. We show a qualitative comparison in Fig. 7. We report the mean and standard deviation of the weighted average of the *mean squared error (MSE)* using the particle weights and the *share of active particles (AP)* after the last time step in Table II. The results have been generated in 10k experiment runs with the ground truth object, robot trajectory and initial belief sampled from the same distribution as shown in Fig 7. We only used 15 particles in the experiment in order to make the particle depletion more likely. The results show that the smooth sensor model outperforms the binary sensor model in terms of belief tracking, as well as much higher particle diversity, i.e. a higher share of active particles with weights that are not close to zero. While it is straightforward that the resampling leads to all particles being active, the results show that the diversity is also increased through the resampling strategy.

D. Spline-based Trajectory Representation

We adopt the spline-based representation from the work of [11] for constructing reference trajectories from a parameter $\boldsymbol{\theta}$. The spline-based representation maps a phase variable $s \in [0, 1]$ and the parameter $\boldsymbol{\theta}$ to a reference position

$$\boldsymbol{u}(s) = \boldsymbol{\Phi}_{\boldsymbol{\theta}}(s)\boldsymbol{\theta} + \boldsymbol{\Phi}_{\boldsymbol{\lambda}}(s)\boldsymbol{\lambda}. \tag{22}$$

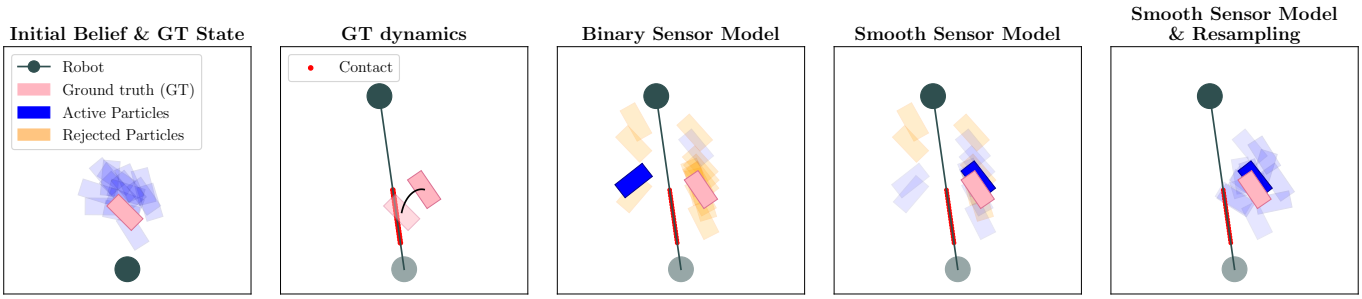


Fig. 7: *Ablation Study:* We compare the performance of the binary sensor model without resampling (left), the smooth sensor model without resampling (middle), and the smooth sensor model with resampling (right). The smooth sensor model with resampling shows the best performance in terms of belief tracking and particle diversity.

The parameter $\boldsymbol{\theta}{=}\boldsymbol{\theta}_{1:N_{\text{via}}}$ contains N_{via} via-points the resulting trajectory $\boldsymbol{u}(s)$ will pass through, i.e. $\boldsymbol{u}(s{=}{}^{n}/_{N_{\text{via}}}){=}\boldsymbol{\theta}_n$, $\forall n{=}1{:}N_{\text{via}}$. The parameter $\boldsymbol{\lambda}{=}\left[\boldsymbol{u}_0^\top, \dot{\boldsymbol{u}}_0^\top, \dot{\boldsymbol{u}}_{K-1}^\top\right]^\top$ contains information about the first set-point $\boldsymbol{u}(s{=}0){=}\boldsymbol{u}_0$, the velocity of the first set-point $\dot{\boldsymbol{u}}(s{=}0){=}\dot{\boldsymbol{u}}_0$ and the velocity of the final set-point $\dot{\boldsymbol{u}}(s{=}1){=}\dot{\boldsymbol{u}}_{K-1}$, which are assumed to be given. The velocities of the the first and last set-points are set to zero in our experiments.

E. Implementation Details

Contact Measurement Signal: The robot's torque sensors provide information about the torques acting on the robot's joints. We use this information to detect whether the robot is in contact with the object and to estimate the object pose. As the torque sensors from the Franka robot arm are quite noisy, we cannot directly use them as observations, but rather transform the signal into a binary contact/no-contact signal. More precisely, we filter the signal from measurements due to gravity, friction and the actual actuation of the robot joints through an observer [6]. Then, we compute the norm of the torques measured in the first five joints of the robot and use a threshold to determine whether the robot is in contact with the object. This binary measurement $z_k \in \{0, 1\}$ is then used as an observation in the particle filter. We only use the first five joints as the last two joints are particularly noisy and do not provide useful information about the contact state.

Low-level Control: We ensure moderate contact forces throughout the robot operation via an impedance controller on the low-level. While the control gains are higher during the localisation phase, we reduce the gains in the moment of grasping to allow for a more robust grasp. In parallel, we keep track of the contact forces acting on the robot's end effector by projecting the torques measured in the robot's joints onto the end effector frame. If the contact forces exceed a given threshold, the robot stops the current action early and transitions to the particle filter update phase.

Initial Belief: The initial belief about the object pose is represented as a set of particles, where each particle is a 6D pose of the object. The particles are sampled from a Gaussian mixture model with two components. For each component, we use a standard deviation of $\sigma_{\text{pos}}{=}0.05$ [m] for the box position and a standard deviation of $\sigma_{\text{ori}}{=}0.4$ [rad] for the yaw orientation. All weights are set to ${}^{1}/_{N_p}$, where $N_p{=}100$ is the number of particles.

Hyperparameters: In all experiments, the threshold on the probability of grasp success is set to 0.75. The number of candidates sampled in each iteration of the predictive sampling-based planner is set to $N_{\text{samples}}{=}20$. The parameters for the proximity-based measurement model are set to $\alpha_{tp} = 0.5$, $\alpha_{fp} = 0.1$ and $\gamma = 1000$. The probability distribution for a positive contact measurement for these parameters is shown in the Appendix in Fig. 6. Moreover, in the experiments, we set the maximum number of iterations, i.e. a maximum number of localising actions, to 15. After this number of iterations, the robot stops the localisation phase and transitions to the grasping phase, regardless of the probability of grasp success.