Learning active tactile perception through belief-space control

Jean-François Tremblay*, David Meger†, Francois R. Hogan+ and Gregory Dudek†

***Abstract*— We propose a method that autonomously learns tactile exploration policies by developing a generative world model that is leveraged to 1) estimate the object's physical parameters using a differentiable Bayesian filtering algorithm and 2) develop an exploration policy using an information-gathering model predictive controller. We evaluate our method on three simulated tasks where the goal is to estimate a desired object property (mass, height or toppling height) through physical interaction. We find that our method is able to discover policies that efficiently gather information about the desired property in an intuitive manner. Finally, we validate our method on a real robot system for the height estimation task, where our method is able to successfully learn and execute an information-gathering policy from scratch.**

I. Introduction

Robots deployed in unstructured environments must reason about the physical properties of novel objects — mass, height, friction — before reliably manipulating them. Unlike geometric shape, these properties are not recoverable from a single sensor reading; they require deliberate physical interaction. Psychology refers to such goal-directed touch as *exploratory procedures* [1]: lifting to judge mass, pressing to judge stiffness, pivoting to judge the center of mass. Designing such procedures by hand is brittle and object-class-specific. We instead pose the problem as one of active Bayesian inference: given a target property to estimate, learn an exploration policy that gathers maximally informative observations.

Our approach has two tightly coupled components. First, a *generative learning-based extended Kalman filter (EKF)* jointly estimates the robot/object state and the target property from a stream of proprioceptive and force/torque observations, without requiring ground-truth state supervision. Second, an *information-seeking model-predictive control (MPC)* controller uses the learned model to simulate belief trajectories forward and selects actions that minimise future uncertainty about the property. Trained end-to-end with no human demonstration or task-specific controller engineering, the system discovers diverse, intuitive behaviours tailored to each task.

Our three main contributions are: (i) a novel observation-likelihood loss for differentiable Kalman filtering that requires no state labels; (ii) an information-gathering controller that plans in belief space using the learned model; and (iii) experimental validation on three physics simulation benchmarks and a real 7-DoF robot arm. The complete, published version of this work is available in [2] and [3].

Work done while all authors were at Samsung Electronics. *J.-F. Tremblay is with Vention Inc., Montréal, Canada. †D. Meger and G. Dudek are with McGill University, Montréal, Canada. +F. R. Hogan is with Amazon Frontier AI & Robotics, Montréal, Canada. Corresponding Email: jft@cim.mcgill.ca

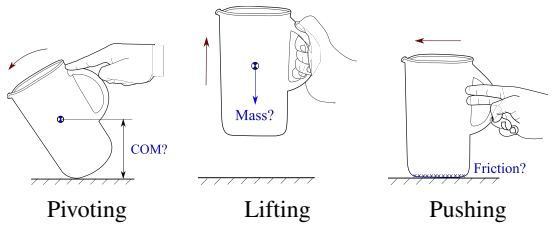

Fig. 1. Three exploratory strategies discovered by our method for estimating center of mass (COM), mass, and friction respectively.

II. Related Work

Classical state estimation appends unknown parameters to the Kalman filter state [4] to enable parameter estimation, but assumes known models. Recent work fuses deep learning with Bayesian filtering: Lee *et al.* [5] provides a differentiable filtering library covering both generative and discriminative models, though generative models there require full state supervision during training. Haarnoja *et al.* [6] presents a discriminative Backprop Kalman filter that replaces the observation model with a direct state predictor; such discriminative formulations cannot simulate the belief forward in time, which is essential for planning. On the active perception side, Denil *et al.* [7] use deep RL to determine "which object is heavier" via constrained pushes; SwingBot [8] uses a hand-engineered shaking routine before a swing-up task. Our method differs in learning unconstrained exploratory procedures for a 7-DoF arm with no human-specified action structure.

III. Method

We model interaction as a partially observable Markov decision process (POMDP) $(\mathcal{S}, \mathcal{M}, \mathcal{A}, p(s_{t+1}|s_t, m, a_t), \Omega, p(o_t|s_t), c)$. The latent state $s_t \in \mathbb{R}^n$ captures robot and contact dynamics; the scalar property $m \in \mathcal{M}$ (mass, height, etc.) is fixed throughout an episode but unknown. At each step the robot receives observation o_t (joint positions, velocities, wrist wrench) and executes action a_t (end-effector velocity). The goal is to estimate m accurately by the episode end.

a) Learning-based generative EKF: We learn neural dynamics and observation models

$$s_t = f_\theta(s_{t-1}, m, a_{t-1}) + \Sigma_\theta(s_{t-1}, m, a_{t-1})\, w_t, \quad (1)$$

$$o_t = h_\theta(s_t) + \Gamma_\theta(s_t)\, v_t, \quad (2)$$

with w_t, v_t standard Gaussian noise. By appending m to the state and running a differentiable EKF, we obtain the predicted belief $\bar{b}_t = \mathcal{N}(s_t, m; \bar{\mu}_t, \bar{\Sigma}_t)$ before incorporating o_t,

and b_t after incorporating o_t. Marginal beliefs for s and m are denoted b_t^s, $\bar{b}_t^s$ and b_t^m, $\bar{b}_t^m$ respectively. Training maximises a novel *observation evidence lower bound (ELBO)*

$$\mathcal{L}_o = \sum_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} \log p(o_t|\theta, s_t^i), \quad s_t^i \sim \bar{b}_t^s, \tag{3}$$

and a Gaussian property-estimation likelihood on b_t^m against the ground-truth m:

$$\mathcal{L}_m = -\sum_{t=1}^{T} \log b_t^m(m). \tag{4}$$

Because the observation model is generative (not discriminative), we can roll out the belief forward in time for planning.

b) Information-gathering controller: At execution time we run a receding-horizon MPC that selects the action sequence $a_{t:t+H}$ minimising predicted property uncertainty:

$$J(a_{t:t+H}) = \sum_{\tau=t}^{t+H} \sigma_\tau^m, \tag{5}$$

where σ_τ^m is the predicted standard deviation of the property marginal, simulated forward using the learned model. The cost J is optimised through a sampling-based method.

c) Training loop: Model and controller are trained jointly in a data-collection loop: the robot executes the current best policy, and the collected transitions are used to update the EKF network parameters.

IV. Experiments and Results

a) Simulation: We evaluate on three Robosuite [9] environments: *mass estimation* (tabletop pushing), *height estimation* (contacting an upright cylinder from above), and *toppling-height estimation* (pivoting an object to find its tipping point). Each episode uses a different randomised property value; the robot has $H = 10$ planning horizon. We compare against a model-free gated recurrent unit (GRU)-based TD3 baseline [10] with the same architecture capacity, rewarded for property estimation error. After $50{,}000$ interactions our method achieves lower mean absolute error (MAE) and more consistent policies across all three tasks. Qualitatively, the discovered strategies match intuitive exploratory behaviours: sliding contact for height, repeated lifts for mass, and angular pivoting for the toppling height.

b) Real robot: We deploy the height-estimation policy on a Franka Emika Panda arm equipped with a Robotous wrist force-torque sensor and a 3D-printed palm end-effector; joint angles and wrist wrench are the only sensing modalities. The policy is trained entirely from real interactions (no simulation transfer). After $50{,}000$ interaction steps the robot achieves a mean absolute height error of 1.19 cm, demonstrating that the method scales to noisy hardware without modification.

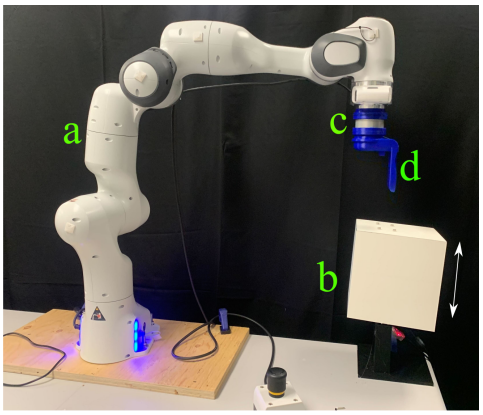

Fig. 2. Real robot setup for the height estimation experiment: (a) the Franka Emika Panda arm, (b) the motorised platform whose height changes each episode, (c) the Robotous RFT60-HA01 wrist force-torque sensor, and (d) the 3D-printed palm end-effector.

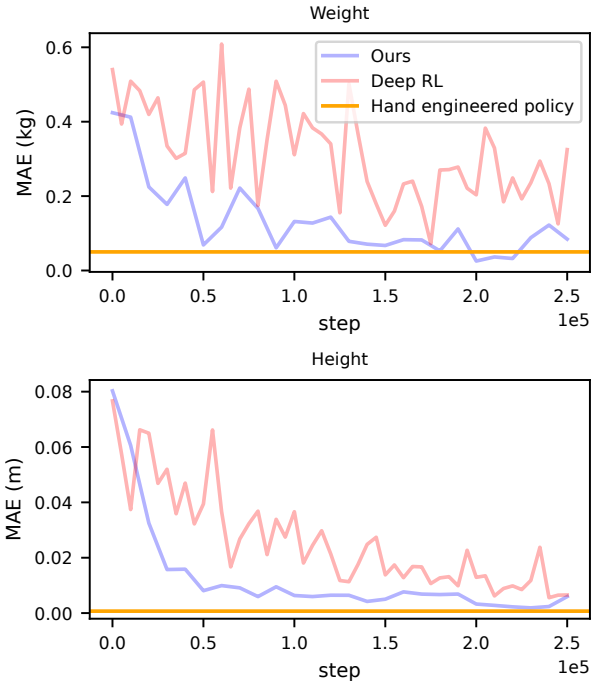

Fig. 3. Learning curves for two simulation tasks. Our method (reaches lower MAE than the RL baseline and does so with fewer environment interactions.

V. Conclusion

We presented an active tactile perception framework that autonomously discovers exploratory procedures by coupling a generative differentiable EKF with belief-space MPC. The method requires only scalar property labels during training and produces diverse, task-specific behaviours in both simulation and on a physical robot. Notably, the real-robot policy converges to a useful exploratory behaviour in under three hours of interaction time, highlighting the data efficiency of the approach. Future work will investigate relaxing the property supervision requirement and extending to multi-property estimation.

REFERENCES

[1] S. J. Lederman and R. L. Klatzky, "Hand movements: A window into haptic object recognition," *Cognitive Psychology*, vol. 19, no. 3, pp. 342–368, 1987.

[2] J.-F. Tremblay, D. Meger, F. R. Hogan, and G. Dudek, "Learning active tactile perception through belief-space control," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2025, pp. 8702–8708.

[3] J.-F. Tremblay, "Planning under uncertainty in the deep learning age," Ph.D. Thesis, McGill University, Montréal, Canada, 2026.

[4] R. F. Stengel, *Optimal control and estimation*. Dover Publications, 1994.

[5] M. A. Lee, B. Yi, R. Martín-Martín, S. Savarese, and J. Bohg, "Multimodal sensor fusion with differentiable filters," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 10 444–10 451.

[6] T. Haarnoja, A. Ajay, S. Levine, and P. Abbeel, "Backprop KF: Learning discriminative deterministic state estimators," in *Advances in Neural Information Processing Systems (NeurIPS)*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29, Curran Associates, Inc., 2016.

[7] M. Denil, P. Agrawal, T. D. Kulkarni, T. Erez, P. Battaglia, and N. De Freitas, "Learning to perform physics experiments via deep reinforcement learning," in *International Conference on Learning Representations (ICLR)*, 2017.

[8] C. Wang, S. Wang, B. Romero, F. Veiga, and E. Adelson, "Swingbot: Learning physical features from in-hand tactile exploration for dynamic swing-up manipulation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 5633–5640.

[9] Y. Zhu, J. Wong, A. Mandlekar, and R. Martín-Martín, "Robosuite: A modular simulation framework and benchmark for robot learning," in *arXiv preprint arXiv:2009.12293*, 2020.

[10] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *International Conference on Machine Learning (ICML)*, J. Dy and A. Krause, Eds., ser. Proceedings of Machine Learning Research, vol. 80, PMLR, 2018, pp. 1587–1596.