Robust Robotic Control from Noisy Perception

Anonymous Author(s)
Affiliation
Address
`email`

Abstract: Noise in visual perception has been a challenge for robot learning systems where optimal action policies are learned from intermediate representations. The problem is further exacerbated when the control task is unstable and fragile, in which case a small amount of noise in the feature space can lead to catastrophic failures of the policy learning algorithms. In this paper, we examine to what extent robot learning can be affected by perception noise in both stable and unstable control problems. We compare the performance of policies learned from intermediate representation (feature + ML), classic robust control methods derived using domain knowledge of system dynamics (feature + control), and policies learned directly from pixels (pixel + ML). We then propose a method to relieve the performance degradation caused by perception noise in feature + ML. Empirical results show significant improvements on policy performance in feature + ML with our approach and [JX: hopefully] it outperforms feature + control and pixel + ML.

Keywords: Robust Control, Noisy Perception

1 Introduction

Systems for visuomotor robotic control must implicitly or explicitly solve two important and closely intertwined problems. The first is visual perception, i.e., how to map raw high-dimensional visual observations o to their task-relevant latent causes a.k.a. the state representation s? The second is the task of learning optimal action policies $\pi(a|s)$ conditioned on those states.

In complex real-world settings, perception is an underconstrained problem. For example, a robot with a single camera cannot perceive depth, look around occlusions, or see spatially or temporally high resolution events. More precisely, the observations o are a non-invertible function of the relevant state s. As such, the perception outputs $\hat{s}$ cannot match the state s perfectly.

This imperfect perception problem may be represented formally as a partially observable Markov decision process (POMDP). The successes of reinforcement learning in the last few years have been demonstrated on fully observed tasks, and general methods for tackling POMDPs remain elusive.

In our work, we aim first to characterize the degree of susceptibility or robustness of different robotic control approaches to imperfect/partial perception, in various stable and unstable control problems. Then, we propose new approaches for synthesizing robust data-driven controllers.

Controller types.

- `feature + robust control` — plug the outputs $\hat{s}$ of perception into a robust controller. This requires domain knowledge of system dynamics, often linearized within a small operating range.
- `feature + ML` — train the perception module as before, but this time, train an ML policy on top of $\hat{s}$.
- `pixel + ML` — directly learn visuomotor policies that map from raw observations o to actions a.

Both `feature + ML` and `pixel + ML` can be trained either with imitation learning, or reinforcement learning.

Submitted to the 4th Conference on Robot Learning (CoRL 2020). Do not distribute.

There has been work on `feature + ML` systems from the computer vision literature. See [1] and [2], that show advantages from using generic mid-level visual representations such as edge detectors and segmentation maps when training action policies.

Questions.

- For varying levels of perception difficulty and varying types of control problems (e.g. stable vs unstable), how do the performances of the above controller types compare?
- How can we make the ML systems more robust to perception errors?

Solution directions and baselines.

- Directly apply existing methods for general POMDPs (not great, afaik). Basic candidate approaches: pure policy gradient (doesn't rely on Markov property) [3], deep recurrent Q networks [4]. Stochastic Latent Actor Critic (SLAC) and Deep Variational Reinforcement Learning (DVRL) are two somewhat related recent approaches for POMDPs that combine latent variable dynamics model learning and actor-critic policy learning — SLAC seems to work better). Other potentially interesting refs: QMDP-Net [5], and Yisong and Anima's POMDP PG work [6]. This type of method could apply to both `feature + ML` and `pixel + ML` controller types.
- For `feature + ML` systems, we could characterize the belief $b(s|o) = p(\hat{s}|o)$ before / during policy learning, so that we don't have to treat that as unknown within the POMDP. This amounts to producing well-calibrated uncertainty estimates in the perception module for the "feature + ML" pipeline – could use ensemble approaches for probabilistic state estimation, or even estimate $p(\hat{s}|o)$ post-hoc on held-out data after training the perception module a la Platt scaling. Then, just run standard RL on top of the beliefs. See [7] for a classic reference for this kind of "belief MDP".
- Depending on the setting, we may sometimes be able to act to better observe s. This could mean something as simple as adding a reward term for policy training, that incentivizes higher confidence predictions from the perception module. These techniques would also be independent of the controller type. This is related to the active perception literature.

2 Related Work

[JX: TODO]

Classic Optimal Control [8, 9, 10]

Learning from Pixels

Learning from Representations

Learning with Noise If we limit the state to be partially observable (for example, using z observation in the stick balancing problem), the agent cannot determine the full state exactly based on the current observation; to plan optimal actions, it must integrate information over the past history of actions and observations. There are many works in the literature of POMDP are of close relevance. Hausknecht and Stone [11] replace the postconvolutional fully connected layer of DQN by a recurrent LSTM layer allows it to deal with partial observaiblity. QMDP-Net [5] incorporates two modules: A bayesian filter for updating the belief and a QMDP module to run value iteration at its core to select actions based on the current belief. However, it does not show robust extension to continuous-space tasks. Karkus et al. [5] study Markovian policies in episodic POMDPs and MDPs, with both discounted and discounted rewards. Guo et al. [12] proposes several methods for calibrating learned deep neural networks but they focus on classification problems.

Perception based estimation, planning, and control: We argue for *new interfaces, centered around uncertainty quantification and propagation, between perception and action that exploit tradeoffs between the robustness/performance of perceptual sensors and predictors and physical* *system dynamics, control tasks, and safety constraints*. Although we still lack a fundamental understanding of perception-based safety-critical control there is an extensive and active literature on integrating high-dimensional sensors, spanning several research communities, underlying our proposed research – while an exhaustive review is impossible, we attempt to summarize results most directly relevant to our proposal here. Classical approaches include the use of visual servoing [13, 14], where a system is given a set of target states in the form of reference images; however, visual servoing only works well when (a) the goal states are well defined and known *a priori*, and (b) the difference between current and goal images is small, and is further sensitive to occlusions, changes in lighting, and motion blur. Contemporary efforts in the robotics community has focused mainly on integrating camera measurements with inertial odometry via Extended Kalman Filter (EKF) [15, 16, 17], or Simultaneous Localization and Mapping (SLAM) algorithms in both ground [18] and aerial [19] vehicles. While these works focus solely on the estimation component, and do not consider downstream use of state estimates in control loops, the papers [20, 21, 22] all demonstrate techniques that use camera measurements to aid inertial position estimates to enable aggressive control maneuvers in unmanned aerial vehicles. We note however that all of these works can be viewed as taking extremely modular approaches, interfacing perceptual sensors and predictors with certainty equivalent controllers that do not account for any uncertainty in the sensing pipeline. The machine learning community has taken a more data-driven approach, with the earliest such example likely being [23], in which a 3-layer neural-network is trained to infer road direction from images. In Lambert et al. [24], a deep neural network is used to learn a map from image to system state, but no uncertainty quantification is made, restricting its downstream use to certainty equivalent planning and control. Contemporary approaches to vision based planning, typically relying on deep neural networks, include learning maps from image to trail direction [25], learning Q-functions for indoor navigation using 3D CAD images [26], and using images to specify waypoints for indoor robotic navigation [27].

Perception/action pipeline co-design: The approaches outlined above take a certainty equivalent approach in interconnecting perception and action: the planning and control components of the pipeline act take the estimates/predictions produced by the perception component as the true state of the world, and can be viewed as an extreme implementation of the modular philosophy described above wherein perception and action are designed independently and then interconnected via a rigid and limited interface. At the other extreme, one can consider the monolithic approach wherein perception and action are combined as a single component. In the seminal paper [28], the authors show that a monolithic end-to-end approach can be used to learn policies that map raw image observations directly to torques at the robot's motors to perform complex tasks. Similar end-to-end approaches have also been applied in the context of self-driving vehicles (SDVs) in [29, 23, 30, 31, 32], autonomous aerial vehicles (AAVs) in [33, 26], and in general robotics tasks in [34, 35, 36]. These methods have resulted in impressive empirical demonstrations, but lack interpretability, modularity, the ability to diagnose root causes of failure, and safety guarantees. *A novelty in our proposal will be to maintain the performance advantages of co-designed and end-to-end optimized pipelines, while maintaining the interpretability, modularity, and safety guarantees of more traditional pipelines.*

The robotics community has recently begun to explore similar questions to those raised in this proposal, seeking to shift to an intermediate design philosophy. In the SDV community Zeng et al. [37], Liang et al. [38], the authors build on prior work on joint 3D detection, tracking, and prediction [39, 40], and propose an end-to-end learnable and interpretable motion planner, that takes as input LIDAR point clouds and an HD map, and produces intermediate representations in the form of 3D detections and their future trajectories. The final output representation is a space-time cost volume that represents the "goodness" of each location that the SDV can take: the planner then samples feasible trajectories and picks one that minimizes a cost function. This non-parametric cost-volume captures uncertainty and multi-modality in possible trajectories that can be taken by the ego-vehicle. Similarly, in Williams et al. [41] and related works, image and inertial data is mapped to a cost landscape, that is then optimized via a path integral based sampling algorithm. In the context of drone racing and agile perception based flight, recent work [42] combines a convolutional neural network (CNN) for perceptual sensing, which maps raw images into robust representations in the form of waypoints an desired speeds, which are then used by a planning module to generate short, minimimum-jerk trajectory segments and corresponding motor commands to reach the desired goal, and shows that this approach outperforms completely modular and completely monolothic design approaches in axes of interest. These preliminary results highlight three important points. First, that two independent sub-communities within robotics have converged to and discovered similar design pipelines highlights both the promise of our proposed approach, and the *need for a unified underlying theoretical and algorithmic framework*. Second, the *level at which representations are shared between perception and action components is an important design decision*: indeed, as described in [42], sharing discrete navigation commands as in [43, 44] improves robustness at the expense of performance, and conversely, representations at the level of direct control [35] can lead to highly agile control at the expense of high sample complexity and system fragility. Finally, uncertainty is not explicitly quantified or propagated between perception and action, nor do downstream planning and control objectives play a meaningful role in the perception design step.

To begin to address this gap, in [45] PI Matni proposes an approach that uses a jointly learned perception map and error profile to design a corresponding safe set and robust controller for the closed loop system, and shows that under suitable smoothness assumptions that the resulting perception-control loop has favorable generalization properties. This result shows how *uncertainty quantification of the perceptual sensor's output allows for a meaningful interaction between perception and control* – we will build upon and extend this insight. We end by noting that [45] draws inspiration from recent work that seeks to bridge the gap between linear control and estimation and learning theory. These assume a linear time invariant system, and derive finite-time guarantees for system identification and/or integrate learned models into control schemes with finite-time performance guarantees [46, 47, 48, 49, 50, 51, 52, 53, 54, 55] – we note that PIs Matni & Pappas are recognized as leading researchers in this area. *A major goal of this proposal is to carry over the successful principled integration of learning and control methods cited above to include learned perceptual sensors and predictors into safety-critical decision and control loops.*

References

[1] B. Zhou, P. Krähenbühl, and V. Koltun. Does computer vision matter for action? May 2019.

[2] A. Sax, J. O. Zhang, B. Emi, A. Zamir, S. Savarese, and J. Malik. Learning to navigate using Mid-Level visual priors. *CORL*, 2019.

[3] R. J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Mach. Learn.*, 8(3-4):229–256, May 1992.

[4] M. Hausknecht and P. Stone. Deep recurrent q-learning for partially observable mdps. In *2015 AAAI Fall Symposium Series*. aaai.org, 2015.

[5] P. Karkus, D. Hsu, and W. S. Lee. QMDP-Net: Deep learning for planning under partial observability. Mar. 2017.

[6] K. Azizzadenesheli, Y. Yue, and A. Anandkumar. Policy gradient in partially observable environments: Approximation and convergence. Oct. 2018.

[7] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Planning and acting in partially observable stochastic domains. *Artif. Intell.*, 101(1):99–134, May 1998.

[8] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada. Control barrier functions: Theory and applications. In *2019 18th European Control Conference (ECC)*, pages 3420–3431. IEEE, 2019.

[9] J. Doyle, K. Glover, P. Khargonekar, and B. Francis. State-space solutions to standard h 2 and h control problems. In *1988 American Control Conference*, pages 1691–1696. IEEE, 1988.

[10] Y. P. Leong and J. C. Doyle. Understanding robust control theory via stick balancing. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 1508–1514. IEEE, 2016.

[11] M. Hausknecht and P. Stone. Deep recurrent q-learning for partially observable mdps. In *2015 AAAI Fall Symposium Series*, 2015.

[12] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger. On calibration of modern neural networks. *arXiv preprint arXiv:1706.04599*, 2017.

[13] O. Tahri and F. Chaumette. Point-based and region-based image moments for visual servoing of planar objects. *IEEE Transactions on Robotics*, 21(6):1116–1127, 2005.

[14] P. I. Corke et al. *Visual Control of Robots: high-performance visual servoing*. Research Studies Press Taunton, UK, 1996.

[15] E. S. Jones and S. Soatto. Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. *The International Journal of Robotics Research*, 30(4):407–430, 2011.

[16] J. Kelly and G. S. Sukhatme. Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration. *The International Journal of Robotics Research*, 30(1):56–79, 2011.

[17] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis. Camera-imu-based localization: Observability analysis and consistency improvement. *The International Journal of Robotics Research*, 33(1):182–201, 2014.

[18] S. Lynen, T. Sattler, M. Bosse, J. A. Hesch, M. Pollefeys, and R. Siegwart. Get out of my lab: Large-scale, real-time visual-inertial localization. In *Robotics: Science and Systems*, 2015.

[19] S. Lynen, M. W. Achtelik, S. Weiss, M. Chli, and R. Siegwart. A robust and modular multi-sensor fusion approach applied to mav navigation. In *2013 IEEE/RSJ international conference on intelligent robots and systems*, pages 3923–3929. IEEE, 2013.

[20] G. Loianno, C. Brunner, G. McGrath, and V. Kumar. Estimation, control, and planning for aggressive flight with a small quadrotor with a single camera and imu. *IEEE Robotics and Automation Letters*, 2(2):404–411, 2016.

[21] S. Tang, V. Wüest, and V. Kumar. Aggressive flight with suspended payloads using vision-based control. *IEEE Robotics and Automation Letters*, 3(2):1152–1159, 2018.

[22] Y. Lin, F. Gao, T. Qin, W. Gao, T. Liu, W. Wu, Z. Yang, and S. Shen. Autonomous aerial navigation using monocular visual-inertial fusion. *Journal of Field Robotics*, 35(1):23–51, 2018.

[23] D. A. Pomerleau. Alvinn: An autonomous land vehicle in a neural network. In *Advances in neural information processing systems*, pages 305–313, 1989.

[24] A. Lambert, A. Shaban, A. Raj, Z. Liu, and B. Boots. Deep forward and inverse perceptual models for tracking and prediction. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 675–682. IEEE, 2018.

[25] A. Giusti, J. Guzzi, D. C. Cireşan, F.-L. He, J. P. Rodríguez, F. Fontana, M. Faessler, C. Forster, J. Schmidhuber, G. Di Caro, et al. A machine learning approach to visual perception of forest trails for mobile robots. *IEEE Robotics and Automation Letters*, 1(2):661–667, 2015.

[26] F. Sadeghi and S. Levine. Cad2rl: Real single-image flight without a single real image. *arXiv preprint arXiv:1611.04201*, 2016.

[27] S. Bansal, V. Tolani, S. Gupta, J. Malik, and C. Tomlin. Combining optimal control and learning for visual navigation in novel environments. *arXiv preprint arXiv:1903.02531*, 2019.

[28] S. Levine, C. Finn, T. Darrell, and P. Abbeel. End-to-end training of deep visuomotor policies. *The Journal of Machine Learning Research*, 17(1):1334–1373, 2016.

[29] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, et al. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*, 2016.

[30] A. Kendall, J. Hawke, D. Janz, P. Mazur, D. Reda, J.-M. Allen, V.-D. Lam, A. Bewley, and A. Shah. Learning to drive in a day. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8248–8254. IEEE, 2019.

[31] P. Drews, G. Williams, B. Goldfain, E. A. Theodorou, and J. M. Rehg. Aggressive deep driving: Model predictive control with a cnn cost model. *arXiv preprint arXiv:1707.05303*, 2017.

[32] Y. Pan, C.-A. Cheng, K. Saigol, K. Lee, X. Yan, E. Theodorou, and B. Boots. Agile autonomous driving using end-to-end deep imitation learning. *arXiv preprint arXiv:1709.07174*, 2017.

[33] S. Jung, S. Hwang, H. Shin, and D. H. Shim. Perception, guidance, and navigation for indoor autonomous drone racing using deep learning. *IEEE Robotics and Automation Letters*, 3(3): 2539–2544, 2018.

[34] S. Gupta, J. Davidson, S. Levine, R. Sukthankar, and J. Malik. Cognitive mapping and planning for visual navigation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2616–2625, 2017.

[35] G. Kahn, T. Zhang, S. Levine, and P. Abbeel. Plato: Policy learning using adaptive trajectory optimization. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3342–3349. IEEE, 2017.

[36] C. Richter and N. Roy. Safe visual navigation via deep learning and novelty detection. 2017.

[37] W. Zeng, W. Luo, S. Suo, A. Sadat, B. Yang, S. Casas, and R. Urtasun. End-to-end interpretable neural motion planner. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8660–8669, 2019.

[38] M. Liang, B. Yang, W. Zeng, Y. Chen, R. Hu, S. Casas, and R. Urtasun. Pnpnet: End-to-end perception and prediction with tracking in the loop. *arXiv preprint arXiv:2005.14711*, 2020.

[39] W. Luo, B. Yang, and R. Urtasun. Fast and furious: Real time end-to-end 3d detection, tracking and motion forecasting with a single convolutional net. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 3569–3577, 2018.

[40] A. Sauer, N. Savinov, and A. Geiger. Conditional affordance learning for driving in urban environments. *arXiv preprint arXiv:1806.06498*, 2018.

[41] G. Williams, P. Drews, B. Goldfain, J. M. Rehg, and E. A. Theodorou. Information-theoretic model predictive control: Theory and applications to autonomous driving. *IEEE Transactions on Robotics*, 34(6):1603–1622, 2018.

[42] E. Kaufmann, A. Loquercio, R. Ranftl, A. Dosovitskiy, V. Koltun, and D. Scaramuzza. Deep drone racing: Learning agile flight in dynamic environments. *arXiv preprint arXiv:1806.08548*, 2018.

[43] D. K. Kim and T. Chen. Deep neural network for real-time autonomous indoor navigation. *arXiv preprint arXiv:1511.04668*, 2015.

[44] A. Loquercio, A. I. Maqueda, C. R. Del-Blanco, and D. Scaramuzza. Dronet: Learning to fly by driving. *IEEE Robotics and Automation Letters*, 3(2):1088–1095, 2018.

[45] S. Dean, N. Matni, B. Recht, and V. Ye. Robust guarantees for perception-based control. *Learning for Decision and Control (L4DC), to appear*, 2020.

[46] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu. On the sample complexity of the linear quadratic regulator. *arXiv preprint arXiv:1710.01688*, 2017.

[47] Y. Abbasi-Yadkori and C. Szepesvári. Regret bounds for the adaptive control of linear quadratic systems. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 1–26, 2011.

[48] D. J. Russo, B. Van Roy, A. Kazerouni, I. Osband, and Z. Wen. A tutorial on thompson sampling. *Foundations and Trends on Machine Learning*, 11(1):1–96, July 2018. doi:10.1561/2200000070.

[49] M. Abeille and A. Lazaric. Improved regret bounds for thompson sampling in linear quadratic control problems. In *International Conference on Machine Learning*, pages 1–9, 2018.

[50] Y. Ouyang, M. Gagrani, and R. Jain. Control of unknown linear systems with thompson sampling. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1198–1205, Oct 2017. doi:10.1109/ALLERTON.2017.8262873.

[51] Y. Abbasi-Yadkori, N. Lazic, and C. Szepesvári. Model-free linear quadratic control via reduction to expert prediction. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3108–3117, 2019.

[52] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu. Regret bounds for robust adaptive control of the linear quadratic regulator. In *Advances in Neural Information Processing Systems*, pages 4192–4201, 2018.

[53] A. Cohen, T. Koren, and Y. Mansour. Learning linear-quadratic regulators efficiently with only $\sqrt{T}$ regret. *arXiv preprint arXiv:1902.06223*, 2019.

[54] H. Mania, S. Tu, and B. Recht. Certainty equivalent control of lqr is efficient. *arXiv preprint arXiv:1902.07826*, 2019.

[55] A. Rantzer. Concentration bounds for single parameter adaptive control. In *2018 Annual American Control Conference (ACC)*, pages 1862–1866. IEEE, 2018.