# Learning Robot In-Hand Manipulation with Tactile Features

Herke van Hoof[1], Nutan Chen[2], Maximilian Karl[2], Tucker Hermans[3], Gerhard Neumann[1], Patrick van der Smagt[2,4], and Jan Peters[1,5]

## I. Introduction and Related Work

In-hand manipulation enables a robot to reconfigure objects that cannot be picked up in the desired configuration [1]. Most methods for in-hand manipulation rely on exact models of the hand and the objects [2, 3]. However, to manipulate unknown objects, systems should reactively adapt to the object rather than requiring an exact model. Adaptation can be realized through compliant hardware [4] and sensory feedback. Tactile and haptic feedback are especially useful for in-hand manipulation [5, 6], but commonly only work for fully actuated hands with known models. Exact models, however, are often not available for compliant robots and tactile sensors.

In our previous work [7], we have proposed learning control policies for unknown systems using non-parametric relative entropy policy search (NP-REPS). NP-REPS is a reinforcement learning method that combines smooth policy updates with non-linear control policies [8]. We evaluated this method on the task of learning to roll an object between the fingertips of the compliant ReFlex robot hand shown in Fig. 1.

Fig. 1: Using reinforcement learning and tactile feedback, a rolling primitive can be learned on an under-actuated robot hand.

When learning with high-dimensional tactile sensors, however, even a small amount of noise can disrupt the learner [9]. Such problems can be ameliorated by using low-dimensional representations, e.g. obtained by auto-encoders. We used such a learned representation to learn a tactile stabilization task shown in Fig. 2 [9].

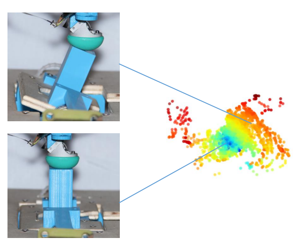

Fig. 2: A robot learns to manipulate using a learned feature space.

In related work, continuous RL has been used for reaching and grasping objects [10–12], as well as for the transportation of grasped objects [12–15]. However, these approaches have not been used to learn *in-hand manipulation*, i.e., changing the object's pose with respect to the hand. Kalakrishnan et al. [16] learned an in-hand task, but relied on trajectories with a fixed starting location.

In contrast to earlier work on learning representations for RL [10, 17, 18], we aim to learn feedback policies directly from the high-dimensional space in an iterative on-policy fashion, allowing the improving policy to generate more relevant samples for the reinforcement learner. Furthermore, we want to train our state encoders in a way that respects the transition dynamics of the controlled system

## II. Methods and Results

In this abstract, we first give a brief description of the NP-REPS algorithm. Than, we describe two experiments. In the first experiment, the robot learns an in-hand manipulation primitive for the under-actuated reflex hand. In the second experiment, we learn a tactile stabilization task using a low-dimensional representation of tactile features.

### A. Non-parametric Relative Entropy Policy Search

To update the control policy based on the learned representation, we use the NP-REPS algorithm [8]. This algorithm aims to maximize the expected rewards obtained by the policy while bounding the information loss of policy updates. This bound prevents overfitting and divergence of the learning process. Furthermore, this algorithm has the advantage of not requiring a parametric form of the value function or policy to be provided, reducing design effort, and is robust to noisy state transitions.

### B. Learning an In-Hand Manipulation Primitive

For the first experiment [7] we use the under-actuated ReFlex robot hand shown is Fig. 1. We desire our robot to perform a rolling task on a object grasped between two opposing fingers. To apply reinforcement learning, the six-dimensional state space is defined as a concatenation of the pressure at each finger with the proximal- and distal joint angles of each finger. As actions, the robot can set the continuous-valued velocity for the two motors. Rewards are given as a linear combination of the distance

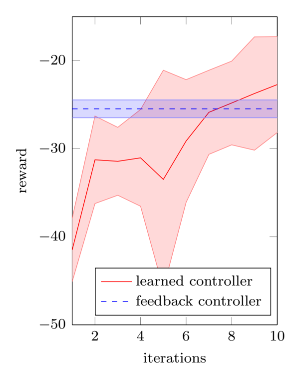


Fig. 3: Average reward and standard deviation across four learning trials.

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreements #610967 (TACMAN) and #270327 (CompLACS).

[1]TU Darmstadt, Computer Science Department.
{hoof, neumann, peters}@ias.tu-darmstadt.de
[2]Faculty for Informatics, TU München.
{nutan, karlma}@in.tum.de
[3]School of Computing, University of Utah.
thermans@cs.utah.edu
[4]fortiss, TUM Associate Institute.
[3]Max Planck Institute for Intelligent Systems.

to the desired goal position, deviation from the desired pressure, and a quadratic penalty for selecting large actions.

In each iteration, 10 roll-outs are samples and the controller is updated. The results of the experiment are shown in Fig. 3. The final policy learned by the algorithm performs better than a hand-coded baseline feedback controller.

### C. Stabilization Task with Learned Features

In the second experiment [9], a 5 degree of freedom robot manipulates a pole through a SynTouch BioTac tactile sensor on the end-effector (see Fig. 2). The pole is on a platform which is able to rotate in roll and pitch. From a random initial position of the pole, the task for the robot is to move the pole to the center position. Thus, the reward is given by an exponential function of the negative squared distance. Actions consist of an increment in forward-backward and left-right position.

We attempted to learn the task using either raw tactile input or a representation learned using an autoencoder. In either case, the input was 228-dimensional tactile pattern data. To learn the representation, we used a variational auto encoder with a hidden layer with 512 neurons and a feature layer with three neurons. A new representation is learned after every iteration (10 roll-outs), with errors on recent data given a bigger weight in the loss function.

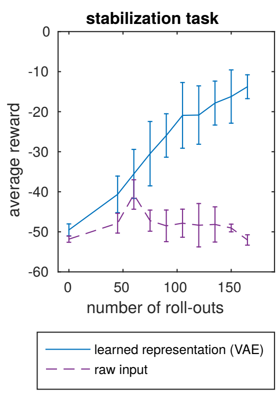

Fig. 4: Average reward and standard deviation across five learning trials.

The results of the experiment are shown in Fig. 4. Learning using the learned representation was successful, whereas the robot did not learn a useful policy based directly on raw input. We believe this effect is due to the effect of noise in high-dimensional input spaces [9].

## III. Conclusions and Future Work

In this abstract, we discussed reinforcement learning of two tactile skills with the NP-REPS algorithm. On an object rolling task with a six-dimensional state representation, the method was successful in finding a manipulation policy that outperforms a handcoded feedback controller. On a stabilization task with high-dimensional tactile inputs, policy improvement required a learned low-dimensional representation to cope with noise.

Compared to many of the existing methods for planning in-hand manipulations, our method has the advantage that we can apply it to unknown hands that are hard to model. The disadvantage, however, is that learning a policy requires time to interact with the system.

In future work, we would like to extend our method to learn to stabilize grasps from a random policy. The final policy should generalize to a wide range of initial configurations. Learning such a policy will require more data, and thus we are working on making our work applicable to larger data sets.

## References

[1] R.R. Ma and A.M. Dollar. On dexterity and dexterous manipulation. In *ICAR*, 2011.

[2] Z. Doulgeri and L. Droukas. On rolling contact motion by robotic fingers via prescribed performance control. In *ICRA*, 2013.

[3] A. Bicchi and R. Sorrentino. Dexterous manipulation through rolling. In *ICRA*, volume 1, 1995.

[4] L.P. Jentoft, Qian Wan, and R.D. Howe. Limits to compliance and the role of tactile sensing in grasping. In *ICRA*, 2014.

[5] H. Maekawa, K. Tanie, and K. Komoriya. Tactile sensor based manipulation of an unknown object by a multifingered hand with rolling contact. In *ICRA*, 1995.

[6] Q. Li, M. Meier, R. Haschke, H. Ritter, and B. Bolder. Rotary object dexterous manipulation in hand: a feedback-based method. *Int. J. Mechatronics and Automation*, 3(1), 2013.

[7] H. van Hoof, T. Hermans, G. Neumann, and J. Peters. Learning robot in-hand manipulation with tactile features. In *Humanoids*, 2015.

[8] H. van Hoof, J. Peters, and G. Neumann. Learning of non-parametric control policies with high-dimensional state features. In *AIStats*, 2015.

[9] H. van Hoof, N. Chen, M. Karl, P. van der Smagt, and J. Peters. Stable reinforcement learning with autoencoders for tactile and visual data. In *IROS*, 2016. Submitted.

[10] T. Lampe and M. Riedmiller. Acquiring visual servoing reaching and grasping skills using neural reinforcement learning. In *IJCNN*, 2013.

[11] O. Kroemer, R. Detry, J. Piater, and J. Peters. Combining active learning and reactive control for robot grasping. *Robot. and Auton. Syst.*, (9), 2010.

[12] O. Kroemer, C. Daniel, G Neumann, H. van Hoof, and J. Peters. Towards learning hierarchical skills for multiphase manipulation tasks. In *ICRA*, 2015.

[13] S. Levine, N. Wagener, and P. Abbeel. Learning contact-rich manipulation skills with guided policy search. In *ICRA*, 2015.

[14] Y. Chebotar, O. Kroemer, and J. Peters. Learning robot tactile sensing for object manipulation. In *IROS*, 2014.

[15] P. Pastor, M. Kalakrishnan, S. Chitta, E. Theodorou, and S. Schaal. Skill learning and task outcome prediction for manipulation. In *ICRA*, 2011.

[16] M. Kalakrishnan, L. Righetti, P. Pastor, and S. Schaal. Learning force control policies for compliant manipulation. In *IROS*, 2011.

[17] C. Finn, X.Y. Tan, Y. Duan, T. Darrell, S. Levine, and P. Abbeel. Learning visual feature spaces for robotic manipulation with deep spatial autoencoders. Technical Report 1509.06113, ArXiv, 2015.

[18] M. Watter, J. Springenberg, J. Boedecker, and M. Riedmiller. Embed to control: A locally linear latent dynamics model for control from raw images. In *NIPS*, 2015.