Full text available at: http://dx.doi.org/10.1561/2300000012

Tactile Guidance for Policy Adaptation

Tactile Guidance for Policy Adaptation

Brenna D. Argall

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, 1015, Switzerland
brennadee.argall@epfl.ch

Eric L. Sauser

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, 1015, Switzerland
eric.sauser@epfl.ch

Aude G. Billard

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, 1015, Switzerland
aude.billard@epfl.ch

now
the essence of knowledge
Boston – Delft

Foundations and Trends® in Robotics

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is B. D. Argall, E. L. Sauser and A. G. Billard, Tactile Guidance for Policy Adaptation, Foundation and Trends® in Robotics, vol 1, no 2, pp 79–133, 2010

ISBN: 978-1-60198-436-4

Foundations and Trends® in Robotics

Volume 1 Issue 2, 2010

Editorial Board

Editorial Scope

Foundations and Trends® in Robotics will publish survey and tutorial articles in the following topics:

- Mathematical modelling
- Kinematics
- Dynamics
- Estimation Methods
- Robot Control
- Planning
- Artificial Intelligence in Robotics
- Software Systems and Architectures
- Mechanisms and Actuators
- Kinematic Structures
- Legged Systems
- Wheeled Systems
- Hands and Grippers
- Micro and Nano Systems
- Sensors and Estimation
- Force Sensing and Control
- Haptic and Tactile Sensors
- Proprioceptive Systems
- Range Sensing
- Robot Vision
- Visual Servoing
- Localization, Mapping and SLAM
- Planning and Control
- Control of manipulation systems
- Control of locomotion systems
- Behaviour based systems
- Distributed systems
- Multi-Robot Systems
- Human-Robot Interaction
- Robot Safety
- Physical Robot Interaction
- Dialog Systems
- Interface design
- Social Interaction
- Teaching by demonstration
- Industrial Robotics
- Welding
- Finishing
- Painting
- Logistics
- Assembly Systems
- Electronic manufacturing
- Service Robotics
- Professional service systems
- Domestic service robots
- Field Robot Systems
- Medical Robotics

Information for Librarians

Foundations and Trends® in Robotics, 2010, Volume 1, 4 issues. ISSN paper version 1935-8253. ISSN online version 1935-8261. Also available as a combined paper and online subscription.

Foundations and Trends® in
Robotics
Vol. 1, No. 2 (2010) 79–133

DOI: 10.1561/2300000012

now
the essence of knowledge

Tactile Guidance for Policy Adaptation

Brenna D. Argall[1], Eric L. Sauser[2] and Aude G. Billard[3]

[1] *Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, 1015, Switzerland, brennadee.argall@epfl.ch*
[2] *Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, 1015, Switzerland, eric.sauser@epfl.ch*
[3] *Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, 1015, Switzerland, aude.billard@epfl.ch*

Abstract

Demonstration learning is a powerful and practical technique to develop robot behaviors. Even so, development remains a challenge and possible demonstration limitations, for example correspondence issues between the robot and demonstrator, can degrade policy performance. This work presents an approach for policy improvement through a tactile interface located on the body of the robot. We introduce the *Tactile Policy Correction* (*TPC*) algorithm, that employs tactile feedback for the *refinement* of a demonstrated policy, as well as its *reuse* for the development of other policies. The TPC algorithm is validated on humanoid robot performing grasp positioning tasks. The performance of the demonstrated policy is found to improve with tactile corrections. Tactile guidance also is shown to enable the development of policies able

The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007–2013 — Challenge 2 — Cognitive Systems, Interaction, Robotics — under grant agreement n^o [231500]-[ROBOSKIN].

to successfully execute novel, undemonstrated, tasks. We further show that different modalities, namely teleoperation and tactile control, provide information about allowable variability in the target behavior in different areas of the state space.

Contents

1

Introduction

The realization of physical movement is fundamental to many robotics applications. Whether operating in industrial and laboratory settings, or within general society, physically embodied robots typically are tasked with the execution of physical actions, thus requiring algorithms for motion control. Over the years a variety of approaches for motion control have been proposed, with many resulting in impressive robot capabilities. The development of control paradigms becomes increasingly difficult however as robot and domain complexities grow, for example with high degree-of-freedom manipulators or interactions with compliant objects. Often traditional approaches that define explicit mathematical models of the world, and from these derive rules for control, struggle to scale with increasing complexity. Moreover, the development of a control paradigm for any robot platform is confounded by difficulties such as noisy sensors and inaccurate actuation.

In the face of such challenges, to develop robust control algorithms typically requires a significant measure of expertise and effort from the developer. The advancement of techniques that reduce the demands placed on a developer therefore are desirable. We introduce in this article an approach to policy development in which corrections

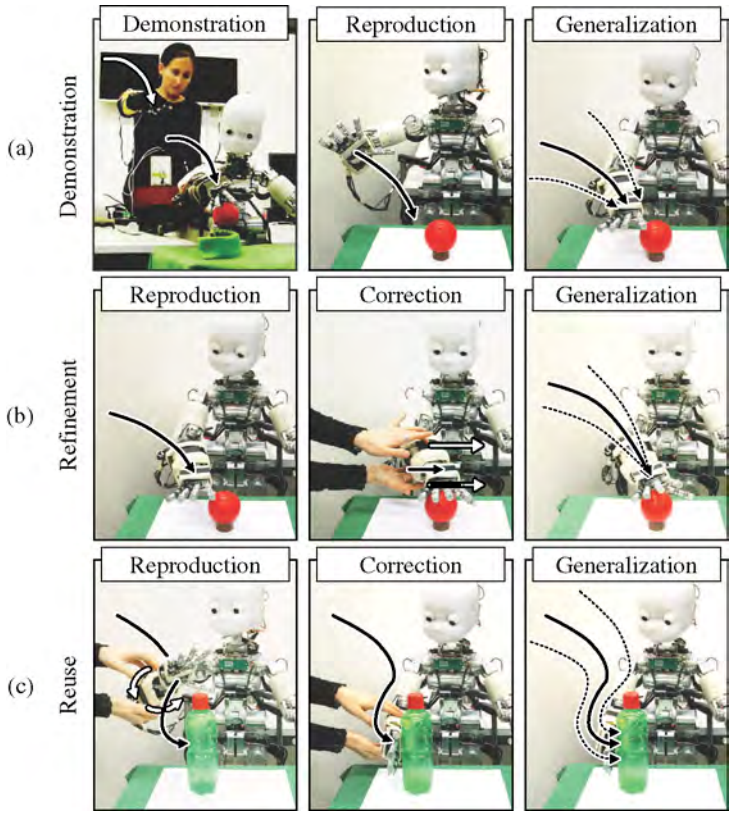

Fig. 1.1 Our approach of (a) task demonstration, followed by tactile correction of the learned policy for (b) refinement of the demonstrated behavior and (c) its reuse in the development of other policies. Black solid arrows indicate demonstrated or corrected executions, black dashed arrows generalization executions and white arrows human hand movement.

provided by a teacher through a tactile interface are used to adapt and improve a policy. Our *Tactile Policy Correction* (*TPC*) algorithm initially derives a policy via *Learning from Demonstration* (*LfD*) techniques (Figure 1.1a). Under LfD, a robot learner generalizes a policy from data recorded during the execution of a target behavior by a task expert. Our approach then has a human teacher provide policy corrections through a tactile interface located on the body of the robot. The corrections indicate relative adjustments to the robot pose, and thus to the policy predictions. The teacher provides corrections in order to accomplish one of two goals, and how corrections are incorporated into

the policy differs for each. The first goal is to *refine* a policy during execution, and thus to improve its performance based on execution experience (Figure 1.1b). The second goal is to assist in policy *reuse*, by guiding an existing policy towards accomplishing a different task (Figure 1.1c).

We validate our approach on a humanoid robot performing end-effector positioning tasks. We show that policies produced under our policy derivation technique are flexible with respect to variability seen between the teacher demonstrations, and furthermore that different teaching modalities (i.e., task demonstration, tactile correction) provide information about acceptable execution variability within different areas of the state space. The performance of a policy learned from demonstration is shown to improve after *refinement* through tactile corrections. Successful policy *reuse* also is validated. Through tactile guidance, executions with existing policies are iteratively adjusted towards producing new behaviors, with the result of policies able to execute alternate, undemonstrated, tasks. Tactile corrections thus *enable* the development of new policies, bootstrapped on the reuse of a policy learned from demonstration.

The remainder of this section reports on the related literature that supports this work. Section 2 introduces the TPC algorithm and presents our implementation in detail. Experimental setup and results are reported in Section 3. A discussion of our approach and findings are provided in Section 4, followed by concluding remarks.

1.1 Background and Motivation

We begin with a discussion of policy development under *Learning from Demonstration* (*LfD*), followed by existing approaches to policy refinement and reuse within LfD.

1.1.1 Learning from Demonstration

Under LfD, teacher executions of a desired behavior are recorded and a policy is derived from the resultant dataset. LfD has seen success in a variety of robotics applications, and has the attractive characteristics of being an intuitive means for human teacher to robot learner knowledge

transfer, as well as being an accessible policy development technique for those who are not robotics-experts. There are many design decisions to consider when building an LfD system. These range from who executes the demonstrations and how they are recorded, to the technique used for policy derivation. Here we overview only those decisions specific to our particular system, and refer the reader to [2] and [8] for a full review of robot LfD.

When recording and executing demonstrations the issue of *correspondence* is key, where teacher demonstrations do not directly map to the robot learner due to differences in sensing or motion [21]. Correspondence issues are minimized when the learner records directly from its own sensors while under the control of the teacher. For example, under *teleoperation* the teacher remotely controls the robot platform (e.g. [27]), while under *kinesthetic control* the teacher touches the robot to guide the motion (e.g. [9]). Teleoperation requires an interface for the direct control of all degrees of freedom on the robot. By contrast, kinesthetic teaching requires a (passive or active) responsiveness to human touch, for example back-drivable motors or force–torque sensing in the joints. Both techniques are employed in our work.

Many approaches exist within LfD to derive a policy from the demonstration data [2], the most popular of which either directly approximate the underlying function mapping observations to actions, or approximate a state transition model and then derive a policy using techniques such as Reinforcement Learning [26]. Our work derives a policy under a variant of the first approach, where probabilistic regression techniques are used to predict a target robot pose based on world state, and a controller external to the algorithm selects an action able to accomplish this target pose. Our reason for splitting policy prediction into these two steps is tied to the mechanism by which the algorithm responds to tactile feedback (discussed in Section 2.1).

1.1.2 Policy Refinement and Reuse

Even with the advantages secured through demonstration, policy development typically is still non-trivial. To have a robot learn from its execution performance, or *experience*, therefore is a valuable policy

improvement tool for any development technique. Within the context of LfD specifically, execution experience can be used to overcome limitations in the demonstration dataset. One possible limitation is dataset sparsity, since demonstration from every world state is infeasible in all but the simplest domains. Other limitations include poor correspondence between the teacher and learner or deficiencies in the teacher, who may in fact provide suboptimal or ambiguous demonstrations. Here we consider policy *refinement* and policy *reuse* as two techniques to assist the development process, or equivalently to reduce the strain on the policy developer.

Within demonstration learning, a variety of approaches incorporate information gathered from experience in order to *refine* a policy. For example, execution experience is used to update reward-determined state values [15, 19, 25] and learned state transition models [1, 6]. Other approaches provide more demonstration data, driven by teacher-initiated demonstrations [9] as well as by learner requests for more data [11, 13]. In this work, we also provide more data, but using a different control mechanism than during the initial teacher demonstrations; specifically, teleoperation is used for the initial demonstration data, and a form of hybrid kinesthetic control when producing the refinement data.

Policy *reuse* under LfD occurs most frequently with behavior primitives, or simpler policies that contribute to the execution of a more complex policy. Hand-coded behavior primitives are used within tasks learned from demonstration [22], demonstrated primitives are combined into a new policy by a human [24] or automatically by the learning algorithm [3], and demonstrated tasks are decomposed into a library of primitives [7]. The focus of our approach is instead on adapting an existing policy to accomplish a *different* task, rather than incorporating the existing behavior as a subcomponent of a larger task.

1.1.3 Tactile Corrections

To enable policy refinement and reuse, the approach taken in this work is to provide *corrections* on a policy execution. Corrections have the advantage of providing guidance on a more suitable alternate prediction

for the policy, instead of requiring that this be inferred from an indication of prediction quality, as state reward does for example. Having directed feedback becomes particularly relevant when guiding a policy towards accomplishing a novel behavior.

Within LfD policy correction has seen limited attention, and most examples consider a human teacher selecting the correct prediction from a discrete set of actions with significant time duration [11, 22]. The target application domain for our work however has policies making continuous-valued predictions at a rapid rate, and both features complicate the individual selection of a single alternate prediction to serve as the correction. To address these challenges, we translate feedback from a tactile sensor into continuous-valued modifications of the current pose, as the robot executes. In contrast to other work with continuous-valued corrections [3], we offer corrective feedback online, instead of post-execution, and through a tactile interface, instead of a high-level computational language.

We posit that tactile feedback furthers many of the strengths of demonstration-based learning. Namely, humans already use touch to instruct other humans in certain contexts; for example when demonstrating a motion, like a tennis swing, that requires a particular position trajectory. To augment demonstration learning with tactile feedback therefore is one natural extension to the idea of teaching robots as humans teach other humans. Demonstration-based policy development also is accessible to those who are not robotics experts, and possibly operating robots outside of laboratory or industrial settings. Here the detection of tactile interactions can be critical for safe robot operation around humans, and so tactile sensing gains importance on a very fundamental level. These tactile sensing capabilities might then be additionally exploited, to transfer knowledge from human to robot for the purpose of behavior development.

Within the field of robot learning (including but not restricted to LfD), only a handful of works utilize human touch for the development of robot behaviors. For example, tactile feedback is detected in order to minimize resistance to movement during demonstration with an industrial arm [14], and to minimize the support forces provided by a teacher during humanoid behavior learning [20]. Tactile interactions between a

robotic pet-surrogate and elderly patients also are mapped to reward signals, that are used within a Reinforcement Learning paradigm to adapt behavior selection [29].

1.2 Our Approach

In summary, the approach presented in this paper employs *tactile corrections* to modify a policy learned through demonstration, for the purpose of both policy *refinement* and policy *reuse*.

Our target application domain is low-level motion control for high degree-of-freedom (DoF) robots. To specify a target behavior for each joint is complicated, and systems typically are under-constrained, resulting in, for example, many joint configurations mapping to a single end-effector pose. The ability to exploit previously learned domain knowledge for the development of new policy behaviors, i.e. policy reuse, thus is advantageous. Performance might suffer however if the reused policy provides only an approximation to the new target behavior. Moreover, while the use of demonstration for policy development is practical for many reasons, it is limited by the interface controlling the demonstration, the quality of which furthermore frequently degrades as the degrees of freedom to control increase. We aim to overcome policy deficiencies through refinement.

To accomplish both refinement and reuse, the policy incorporates new behavior examples. Instead of producing the examples from teacher demonstration however [9, 11, 13], which would be unable to improve upon limitations like a poor demonstration interface, we have the student respond online to corrections indicated by a teacher and treat the resultant trajectory as new training data. Providing explicit corrections has been seldom used within the LfD paradigm [11, 22], especially when the corrections are continuous-valued [3].

We provide corrections through a tactile interface. In addition to being a technique that is relatively unaddressed to date within the robot learning literature in general [20, 29], and the LfD literature in particular [14], we argue that information transfer through human touch is a natural extension of human demonstration, as an intuitive and effective mechanism for the transfer of knowledge from human to robot.

References

[1] P. Abbeel and A. Y. Ng, "Exploration and apprenticeship learning in reinforcement learning," in *Proceedings of the 22nd International Conference on Machine Learning (ICML'05)*, Bonn, Germany, 2005.

[2] B. Argall, S. Chernova, B. Browning, and M. Veloso, "A survey of robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 57, no. 5, pp. 469–483, 2009.

[3] B. D. Argall, "Learning Mobile Robot Motion Control from Demonstration and Corrective Feedback," PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, March 2009.

[4] B. D. Argall, E. L. Sauser, and A. G. Billard, "Tactile guidance for policy refinement and reuse," in *9th IEEE International Conference on Development and Learning (ICDL '10)*, Ann Arbor, Michigan, USA, 2010.

[5] P. Baerlocher and R. Boulic, "An inverse kinematics architecture enforcing an arbitrary number of strict priority levels," *International Journal of Computer Graphics*, vol. 20, 2004.

[6] J. A. Bagnell and J. G. Schneider, "Autonomous helicopter control using reinforcement learning policy search methods," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'01)*, Seoul, Korea, 2001.

[7] D. C. Bentivegna, "Learning from Observation Using Primitives," PhD thesis, College of Computing, Georgia Institute of Technology, Atlanta, GA, July 2004.

[8] A. Billard, S. Callinon, R. Dillmann, and S. Schaal, "Robot programming by demonstration," in *Handbook of Robotics*, (B. Siciliano and O. Khatib, eds.), New York, NY, USA: Chapter 59, Springer, 2008.

[9] S. Calinon and A. Billard, "Incremental learning of gestures by imitation in a humanoid robot," in *Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI'07)*, Arlington, Virginia, USA, 2007.

[10] S. Calinon, F. D'halluin, D. G. Caldwell, and A. Billard, "Handling of multiple constraints and motion alternatives in a robot programming by demonstration framework," in *Proceedings of the IEEE-RAS International Conference on Humanoids Robots*, Paris, France, 2009.

[11] S. Chernova and M. Veloso, "Learning equivalent action choices from demonstration," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'08)*, Nice, France, 2008.

[12] D. Cohn, Z. Ghahramani, and M. Jordan, "Active learning with statistical models," *Artificial Intelligence Research*, vol. 4, pp. 129–145, 1996.

[13] D. H. Grollman and O. C. Jenkins, "Dogged learning for robots," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '07)*, Rome, Italy, 2007.

[14] G. Grunwald, G. Schreiber, A. Albu-Chaffer, and G. Hirzinger, "Programming by touch: The different way of human-robot interaction," *IEEE Transactions on Industrial Electronics*, vol. 50, no. 4, 2003.

[15] F. Guenter, M. Hersch, S. Calinon, and A. Billard, "Reinforcement learning for imitating constrained reaching movements," *RSJ Advanced Robotics, Special Issue on Imitative Robots*, vol. 21, pp. 1521–1544, 2007.

[16] R. Jakel, S. R. Schmidt-Rohr, M. Losch, and R. Dillmann, "Representation and constrained planning of manipulation strategies in the context of programming by demonstration," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '10)*, Anchorage, Alaska, USA, 2010.

[17] M. Kaiser, H. Friedrich, and R. Dillmann, "Obtaining good performance from a bad teacher," in *Programming by Demonstration vs. Learning from Examples Workshop at ML'95*, Tahoe City, California, USA, 1995.

[18] S. M. Khansari-Zadeh and A. Billard, "BM: An iterative algorithm to learn stable non-linear dynamical systems with gaussian mixture models," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '10)*, Anchorage, Alaska, USA, 2010.

[19] J. Kober and J. Peters, "Learning motor primitives for robotics," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '09)*, Kobe, Japan, 2009.

[20] T. Minato, Y. Yoshikawa, T. Noda, S. Ikemoto, H. Ishiguro, and M. Asada, "CB2: A child robot with biomimetic body for cognitive developmental robotics," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '07)*, San Diego, California, USA, 2007.

[21] C. L. Nehaniv and K. Dautenhahn, "The correspondence problem," in *Imitation in Animals and Artifacts*, (K. Dautenhahn and C. L. Nehaniv, eds.), Cambridge, MA, USA: Chapter 2, MIT Press, 2002.

[22] M. N. Nicolescu and M. J. Mataric, "Methods for robot task learning: Demonstrations, generalization and practice," in *Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'03)*, Melbourne, Victoria, Australia, 2003.

[23] P. K. Pook and D. H. Ballard, "Recognizing teleoperated manipulations," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '93)*, Atlanta, Georgia, USA, 1993.

[24] J. Saunders, C. L. Nehaniv, and K. Dautenhahn, "Teaching robots by moulding behavior and scaffolding the environment," in *First Annual Conference on Human-Robot Interactions (HRI '06)*, Salt Lake City, Utah, USA, 2006.

[25] M. Stolle and C. G. Atkeson, "Knowledge transfer using local features," in *Proceedings of IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning (ADPRL'07)*, USA: Honolulu, Hawaii, 2007.

[26] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, London, England: The MIT Press, 1998.

[27] J. D. Sweeney and R. A. Grupen, "A model of shared grasp affordances from demonstration," in *Proceedings of the IEEE-RAS International Conference on Humanoids Robots (Humanoids'07)*, Japan: Tokyo, 2007.

[28] N. Tsagarakis, G. Metta, G. Sandini, D. Vernon, R. Beira, F. Becchi, L. Righetti, J. Santos-Victor, A. Ijspeert, M. Carrozza, and D. Cald-well, "iCub: The design and realization of an open humanoid platform for cognitive and neuroscience research," *Advanced Robotics*, vol. 21, 2007.

[29] K. Wada and T. Shibata, "Social effects of robot therapy in a care house — change of social network of the residents for two months," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '07)*, Italy: Rome, 2007.