Robot Learning from Demonstration in the Force Domain

Leonel Rozo, Pablo Jiménez and **Carme Torras**
Institut de Robòtica i Informàtica Industrial (CSIC-UPC)
Llorens Artigas 4-6, 08028 Barcelona, Spain
{lrozo, jimenez, torras}@iri.upc.edu

Abstract

Researchers are becoming aware of the importance of other information sources besides visual data in robot learning by demonstration (**LbD**). Force-based perceptions are shown to convey very relevant information – missed by visual and position sensors – for learning specific tasks. In this paper, we review some recent works using forces as input data in LbD and Human-Robot interaction (**HRI**) scenarios, and propose a complete learning framework for teaching force-based manipulation skills to a robot through a haptic device. We suggest to use haptic interfaces not only as a demonstration tool but also as a communication channel between the human and the robot, getting the teacher more involved in the teaching process by experiencing the force signals sensed by the robot. Within the proposed framework, we provide solutions for treating force signals, extracting relevant information about the task, encoding the training data and generalizing to perform successfully under unknown conditions.

1 Introduction

One of the main goals of LbD is to enable non-expert human users to teach a robot the tasks to perform. Therefore, it is one of the settings where HRI is most challenged, because of the need to provide natural and non-trivial means of communication in a changing unstructured environment. A human user carries out examples of a given task while a robot *observes* these executions and extracts relevant information for learning, generalizing and executing the taught task under unknown conditions [Billard *et al.*, 2008; Argall *et al.*, 2009].

Human-robot interaction requires suitable communication channels between the human and the robot for conveying information [Goodrich and Schultz, 2007]. In LbD, most works rely on vision or on motion sensors as input channels to the robotic system. As for vision-based input, positional information about the objects in the scene is captured with cameras, which may also be used to locate and follow markers placed on the teacher's body [Dillmann, 2004]. Most state-of-the-art approaches consider vision as the best choice for extracting information from teacher examples, as human beings do in everyday tasks [Bentivegna *et al.*, 2004; Grollman and Jenkins, 2007]. However, vision-based systems must deal with typical problems as occlusion, appearance changes and complex human-robot kinematics mapping, which can be solved by using motion sensors instead. They allow to track the teacher's motion more precisely and to establish a straightforward mapping, which make them appropriate to teach tasks to humanoid robots [Dillmann, 2004; Calinon and Billard, 2007].

In contrast to these works, we are concerned with learning from force-based perceptions. Force conveys relevant information for several tasks where vision or motion sensors can not provide sufficient data to learn a motion or a set of primitives. In many daily tasks, people use force-based perception to perform successfully. Examples include assembly processes, opening doors, pulling drawers, cutting slices of bread, etc. Robots may also take advantage of force/torque information for learning this kind of tasks. Evrard et al. [Evrard *et al.*, 2009] described a learning structure similar to ours, where a humanoid robot learns to carry out collaborative manipulation tasks (object-lifting in vertical axis) using Force/Torque (**F/T**) data. As an extension [Gribovskaya *et al.*, 2011], combines LbD and adaptive control for teaching the task, which endows the robot with variable inertia and an adaptive algorithm to generate different reference kinematic profiles depending on the perceived force. [Kormushev *et al.*, 2011] proposed to use a haptic device for defining the force profile of contact-based tasks (ironing and door opening) while the robot follows a previously learned trajectory.

Recent trends in force-based control have arised in the context of the corrective/refinement phases of learning frameworks and in HRI-based experimental settings for assuring safety and natural interactions. In broad words, the main idea is to take advantage of impedance control theory for modifying the stiffness and compliance characteristics of the robot. [Calinon *et al.*, 2010b] suggested a control strategy for a robotic manipulator performing ironing tasks while interacting with a human operator. Their system takes the most relevant features of the task and the redundancy of the robot into account to determine a controller that is safe for the user. An active control strategy based on task–space control with variable stiffness is suggested, and combined with a safety strategy for tasks requiring humans to move in the vicinity

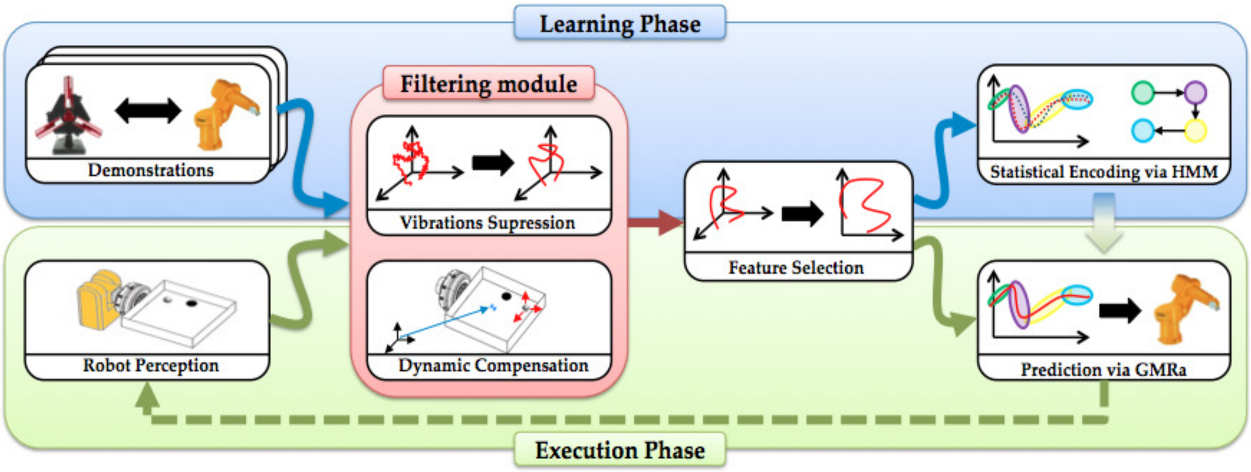


Figure 1: Entire learning framework: (*Top*) Task learning stage, (*Bottom*) Robot execution stage.

of robots. On the other hand, [Lee and Ott, 2010] aimed at extending imitation learning for HRI by incorporating physical contacts, which are treated by real-time motion reshaping and impedance control. Lastly, [Buchli *et al.*, 2011] proposed a model-free, sampling-based learning method named *Policy Improvement with Path Integrals* (PI^2) for achieving variable impedance control via reinforcement learning. No one of these works exploits haptic feedback as a means to enrich the teacher's experience during the learning phase.

The choice of force feedback as a human-robot communication channel has obvious advantages, but it also requirements when clean and realistic signals are desired to be displayed to the human during the demonstration phase. We contribute with a complete F/T data-based learning framework that includes filtering processes and high-fidelity haptic feedback (Figure 1). This allows to create a force-based bidirectional communication channel, which has been very little exploited as a human-robot interaction tool in LbD in contrast to kinesthetic-based teaching. We constructed an experimental setup where a human user holding the end-effector of a 6-DOF haptic interface (Delta device from Force Dimension) teleoperates a robotic arm (RX60 from Stäubli) which has a force/torque sensor (Shunk FTC-050) placed on its wrist. The robot holds a plastic container with a steel sphere inside it, as shown in Figure 2. At the demonstration phase, the human teacher repeatedly carries out the task to be learned, which consists of taking the ball out of the box through the hole, following a specific motion strategy (see Figure 2). During the demonstrations, the teacher feels at the end-effector of the haptic device the F/T sensed at the robotic wrist. Note that the teacher has an additional information source by watching the scene directly. No visual data are provided to the robot.

2 Processing force-based data

The force signals obtained during the demonstration or interaction phases are composed of the ideal signal, external forces and noise. It is necessary and appropriate to process the training data in order to: i) obtain a high-fidelity communication channel between the teacher and the robot, ii) make the task easier to be learned and iii) remove undesirable signals that may be very distracting for the teacher. Most works do not mention how they tackle these issues, but it is important to highlight that this is a necessary step to achieve a reliable learning framework. In our experimental setting, the entire process implies to solve several technical and research issues. Regarding the acquisition of suitable training data from teacher demonstrations, first it is necessary to take into account that the box is not a rigid structure, it vibrates when the robot moves. Here, the solution is to implement a digital low-pass filter to reduce the effects on the sensor readings caused by vibrations. Second, it is important that the teacher can feel F/T generated by the motion of the ball in the box as faithfully as possible without distracting him/her with other signals. For solving this, we propose to feedback only the ball's dynamics inside the structure without reflecting F/T generated by the box's mass, which is achieved by compensating this mass dynamically while executions are carried out (please see [Rozo *et al.*, 2010] for more details).

3 Feature Selection

The *what to Imitate?* problem means to determine which features of the demonstrations are relevant for learning the task successfully [Billard *et al.*, 2008; 2004]. Most works tackle this problem by analyzing the variability across demonstrations of the task at trajectory level. Those parts with large variances do not have to be learned precisely, whereas low variances suggest that the corresponding motion segments are significant and deserve to be learned [Kormushev *et al.*, 2011; Calinon *et al.*, 2007]. This approach exploits variance for constructing task constraints [Calinon and Billard, 2008; Calinon *et al.*, 2010b] as well as for determining secure interaction zones in a robot coaching framework [Lee and Ott, 2010]. However, these works do not focus on the relative relevance of each individual input dimension for the task to be learned. But irrelevant or redundant information may actually be present across input dimensions, which can increase

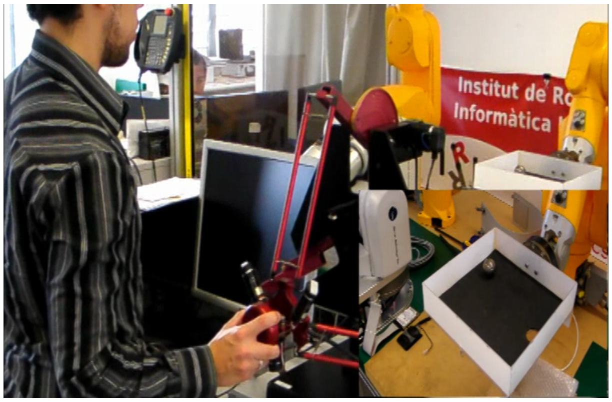


Figure 2: Experimental setup: the teacher holds the end-effector of the haptic device and teleoperates the robotic arm for teaching the task, while feeling forces/torques sensed at the robotic wrist. *Bottom right:* the box held by the robot.

the computational cost of the learning stage and make the task harder to learn. The point is to select the most relevant subset of input variables. The benefits in computational cost and noise reduction during the learning stage do outperform a hypothetical and marginal loss of information. Furthermore, this approach is compatible with the previously described variance-based analysis criterion.

Here we use the Mutual Information (**MI**) criterion, that allows to establish which input variables give more information with respect to their effects on the outputs (i.e. how F/T perceptions affect the teacher actions). Depending on how the uncertainty of the output data is reduced, an input gives more or less information about the output, or in other words, it is highly or lowly correlated with the output [Shannon, 2001]. Formally, the MI value is computed as:

$$I(\mathbf{x}, \mathbf{y}) = \sum_{x} \sum_{y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \tag{1}$$

In our case, inputs are the forces/torques $\mathbf{W} = \{F_x F_y F_z T_x T_y T_z\}$ sensed and transformed to the robot's frame, and output is the position of the robot in its joint space defined by $\mathbf{q} = \{q_1 \ldots q_6\}$. Using equation 1, we computed the MI value for each input/output pair using entire trajectories from the training data. Simple average MI values were calculated for these input/output pairs, which shows a good estimation of how relevant each input variable is with respect to all outputs. In general terms, the input variables F_y and T_z show less relevance whereas T_x and T_y are the most correlated variables with the outputs. This does make sense as they are the variables that give the most useful information for knowing where the ball is inside the box (see Figure 3).

4 Encoding the force-based task

Training data encoding depends on the level at which learning takes place (i.e. trajectory or symbolic level) and on what the input and output data are (e.g. positions, velocities and/or accelerations in joint or operational spaces, forces and/or torques, etc.). Basically, when treating force-based

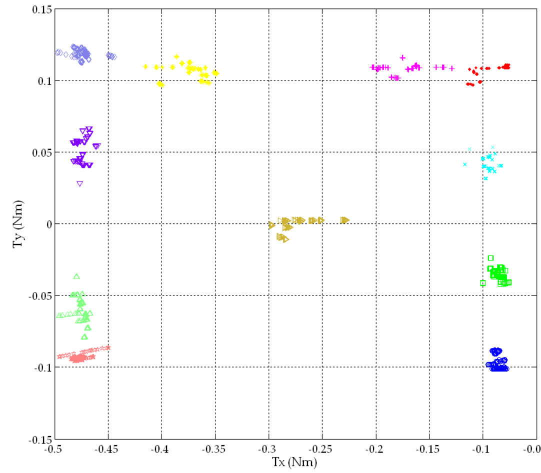


Figure 3: Torques map representing clusters for each initial position of the ball inside the container. Displaying samples of the most relevant variables to the current task T_x vs. T_y, shows that they do describe where the ball is in the box.

tasks, input data are composed of forces and torques, and possible kinesthetic information as joint robot position. Output data often correspond to velocity and maybe acceleration commands. Recent works have proposed to encode this kind of tasks through Dynamic Motion Primitives (**DMP**) [Kormushev *et al.*, 2011; Gams *et al.*, 2010], Gaussian Mixture Models (**GMM**) [Rozo *et al.*, 2010], Hidden Markov Models (**HMM**) [Calinon *et al.*, 2010a] or via mixed versions of some former methods [Gribovskaya *et al.*, 2011]. Others have addressed this problem based on control theory [Ganesh *et al.*, 2010].

In our experimental setup, the teacher's demonstrations given as training data start with the ball placed at different positions inside the box, which means that the goal can be reached from several initial conditions relying on teacher executions[1]. This implies that the learning framework's goal is not to learn merely a trajectory [Evrard *et al.*, 2009] or a task with predefined states as in assembly processes [Dong and Naghdy, 2007] that can be represented at a symbolic level. For endowing the robot with a suitable learning structure for this kind of tasks and avoiding to assume some aspects about the task to be learned, we propose to use a HMM to encode the teacher demonstrations using an ergodic topology, similar to the approach followed in [Calinon *et al.*, 2010a].

Given our experimental setting described above and following the notation of [Rabiner, 1989], let us to denote training datapoints as $\mathbf{d}_p^m \in \Re^D$, with $m = 1, 2, \ldots, M$ and $p = 1, 2, \ldots, P$, where M is the number of demonstrations, P is the number of datapoints collected along demonstration m, and D is the total number of input and output variables. In our current task, inputs correspond to F/T sensed at the robotic wrist and outputs are the velocity commands ω_l at each robot joint q_l with $l = 1, \ldots, 6$. However, thanks to the

[1] Demonstrations were carried out executing a predefined motion strategy that consisted in taking the ball to the wall adjacent to the hole, and then rolling the ball along this wall to the hole.

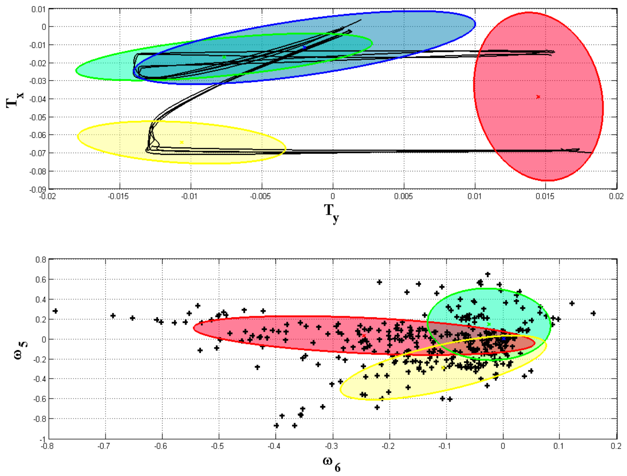


Figure 4: Resulting 4-states HMM trained with demonstrations starting at every position inside the box. *Top*: Input space composed of the most relevant inputs $\{T_x,T_y\}$. *Bottom*: Output space composed of robot joint velocities playing the most important role for the given task. The small, hard-to-see blue state is centred in (0,0).

MI process we concluded that just torques along x and y axes (i.e. T_x and T_y) are necessary as inputs to learn the task successfully because these describe the position of the ball inside the box entirely. Thus, each training datapoint is defined as $\mathbf{d}_p^m = (T_x, T_y, \omega_1, \omega_2, \ldots, \omega_6)$.

With all the demonstrations we can encode the joint distribution $P(\mathbf{T},\omega)$ through an ergodic HMM defined as $\lambda = (\mathbf{A},\mathbf{B},\pi)$. The main idea is to adjust the model λ to maximize $P(O|\lambda)$ where O is an observation sequence $O = O_1O_2\ldots O_T$ with each O_t corresponding to a training datapoint $\mathbf{d}_p^m$. To achieve this objective, an iterative procedure such as the *Baum-Welch* method is used (more details in [Rabiner, 1989]). This permits obtaining a suitable trained HMM that represents the teacher demonstrations statistically through a states model capturing the velocity commands for given sensed torques and taking temporal coherence into account from the resulting matrix $\mathbf{A}$. This may be better understood by observing Figure 4, where the red state in input space covers the beginning of all demonstrations whose initial positions are placed on the wall opposite to where the hole is. At these starting positions, a larger velocity command is required to draw the ball out of its resting configuration (Figure 4, bottom). On the other hand, the blue state covers the trajectory segments corresponding to the end of the task (i.e. when the ball is getting out of the box) in input space. For these torque data, the robot should not carry out any movement (small blue ellipse at (0,0) in output space).

5 Generalization of the task

With respect to how the robot can perform the task autonomously under new conditions, it is crucial to establish a suitable method for computing correct actions relying on given perceptions. In this context, it is also relevant to keep in mind the kind of task the robot has to execute, that is, whether it is a trajectory-level approach or a symbolic-level one. At a trajectory-level, most of the works suggest to use regression-based techniques where the main idea is just to reproduce a generalized trajectory from given demonstrations, mostly using time as the main (sometimes unique) input data [Calinon and Billard, 2007]. On the other hand, some symbolic-level approaches find the robot's action by searching the most probable action to be executed depending on the perceptions [Bentivegna *et al.*, 2004; Dong and Naghdy, 2007].

In this work, we encode the task using HMM, which are mostly used in symbolic approaches, but we compute the robot's action via a modified version of the well-know Gaussian Mixture Regression (**GMR**) technique. This version uses the temporal information captured by the HMM [Calinon *et al.*, 2010a] for obtaining the velocity commands to be sent to each robot joint. The idea of using this temporal information (from the HMM variable α) is to predict the desired velocity command as a function of the given perception (i.e. torques sensed at the robotic wrist) and sequential information probabilistically encapsulated in the HMM, without including time as an input variable. HMM/GMRa provide an estimation by using a weight that takes into consideration both F/T and sequential information. Formally, the new definition of GMR based on temporal information for obtaining the velocity command estimation is given by:

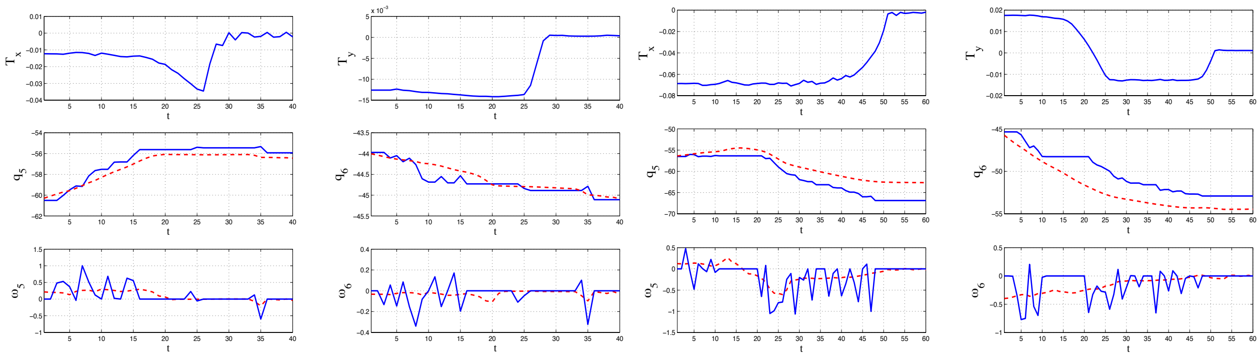
$$\hat{\omega} = \sum_{i=1}^{N} \alpha(i) \left[\mu_{\omega,i} + \Sigma_{\omega T,i}(\Sigma_{TT,i})^{-1}(T - \mu_{T,i})\right] \quad (2)$$

where $\alpha(i)$ is the forward variable for the *i*-th Gaussian in the HMM, $\mu_i = \{\mu_{T,i}, \mu_{\omega,i}\}$, $\Sigma_i = \begin{pmatrix} \Sigma_{TT,i} & \Sigma_{T\omega,i} \\ \Sigma_{\omega T,i} & \Sigma_{\omega\omega,i} \end{pmatrix}$ and N is the number of states in the HMM.

Figures 5(a) and 5(b) show the robot joint trajectories for q_5 and q_6. These trajectories are obtained from the velocity commands predicted through GMRa using the resulting HMM shown in Figure 4. Here, it is possible to observe how the robot's execution is similar to the teacher's one for a given input pattern. Moreover, the learning framework performs successfully when input data lie simultaneously on two HMM states, because it uses the temporal information encapsulated in variable α_k for deciding which state the system is in. Furthermore, robot's executions are smoother than the teacher's ones, which is good and important for the maintenance of joint motors.

6 Measuring robot performance

Many LbD works miss a very crucial issue in robotics, namely performance metrics. Most works evaluate the robot performance based on success or failure to carry out the task, or on the number of trials for accomplishing the goal satisfactorily in reinforcement learning applications. However, it is a relevant issue to determine if the robot learns the task successfully through specific methods, allowing to compare the human and robot executions, how similar the robot motion is with respect to the one shown by the teacher, how long the robot takes to achieve the specific goal, etc. In an initial



(a) Starting from position 1 (b) Starting from position 7

Figure 5: *Top*: Input torques $\{T_x, T_y\}$ during demonstrations. *Middle*: Robot joint trajectories corresponding to the teacher's demonstration (*blue solid line*) and the robot's execution using the HMM/GMRa approach (*red dashed line*). *Bottom*: Robot joint velocity profiles obtained from teacher and robot executions, the latter computed through GMRa.

work, we addressed the same task as a trajectory-level learning problem where the GMM/GMR approach was used and the robot performance measured through the Mean Squared Errors (**MSE**) between teacher and robot executions for the same input patterns [Rozo *et al.*, 2010]. Although this measure can provide a kind of similarity degree, it is not enough to evaluate the robot performance, because low MSEs at trajectory level do not guarantee that the robot accomplish a goal-driven task.

We propose to use success/failure and time-based measures to assess a degree of accomplishment. Therefore, in a first set of experiments, the ball was placed at the same initial positions where the teacher started from. For all different positions of the ball, the robot was able to take the ball out of the box. Afterwards we placed the ball at random positions (untrained ones), and the robot executions achieved the final goal as well. In these cases, the robot executes the motions learned for the closest initial position, by identifying the corresponding HMM state. It was observed that in some executions the ball reached and surpassed the hole, without falling through it. However, the robot was always able to take the ball out of the box after some more executions, as it correctly identified the HMM state corresponding to the current and past input patterns (taking into account the temporal information). This means that the robot predicts its actions as a function of its current and past perceptions, following the teached motion strategy.

Then, we evaluated the performance of the robot executions using a time-based criterion. Here, the idea is to determine how much time the robot takes to complete the task successfully compared with the time needed by the teacher, starting at each pre-defined initial position. For comparison purposes, we did also measure the time needed by the robot to reach the goal simply by chance, performing a series of random motions (Figure 6). As expected, the teacher's executions correspond to the lowest times. A relevant aspect to discuss is the fact that the robot execution times are much larger than the teacher's ones for most positions. Higher times are due to the fact that the robot starts the task by moving the joint q_6 as expected, however it also moves q_5 slightly which sometimes causes the ball go to the bottom of the box. This causes that, when the ball reaches the wall adjacent to the hole, the robot has to carry out more movements in order to take the metalic sphere towards the hole. It would be possible to formulate an equation as a function of time, the MSEs and the success/failure criteria for obtaining a kind of measure of the robot performance (videos available in http://www.iri.upc.edu/people/lrozo/index.html).

7 Conclusions

We briefly reviewed some LbD works using force-based perceptions as input data for their learning processes, where it was possible to derive how force information can be exploited in the LbD context. Furthermore, we described a learning framework that uses force/torque data in order to learn and reproduce the taught task. The paper highlighted that there are some critical issues to be addressed: filtering processes for developing a good bi-directional channel between human and robot, selection of the most relevant perceptions in order to learn the task faster and easier, encoding of force-based tasks and generalization for performing successfully under know and unknown conditions. We also demonstrated how a robot can learn a task successfully only using force data, from which we can state that F/T information can provide relevant data about a given task and complement other kind or data sources.

As future work, we plan to take the learning framework to more realistic settings where force/torque feedback is relevant: opening doors, pulling drawers or emptying deformable bags. Moreover, we would like to apply force-based skill learning to compliant robots in an active learning environment as a refinement or correction phase. In addition, this type of robots would allow us to extend our approach to human-robot collaborative tasks by taking advantage of their compliance features, from an impedance control-based perspective.

Acknowledgements

This research is partially sponsored by the European projects GARNICS (FP7-247947) and IntellAct (FP7-269959), the

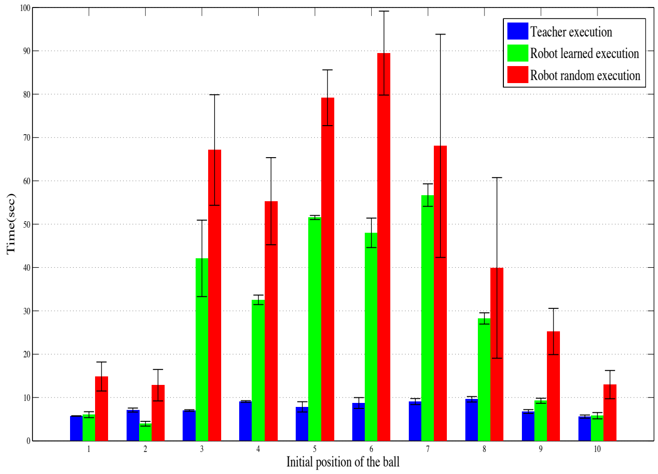


Figure 6: Mean times for teacher executions, robot learned and random executions starting at each pre-defined initial position of the ball inside the box.

Spanish projects MIPRCV (Consolider Ingenio 2010 project CSD2007-00018) and PAU (DPI2008-06022). L. Rozo was supported by the CSIC under a JAE-PREDOC scholarship.

References

[Argall *et al.*, 2009] B. Argall, S. Chernova, M. Veloso, and B. Browning. A survey of robot learning by demonstration. *RAS*, (57):469–483, 2009.

[Bentivegna *et al.*, 2004] D. Bentivegna, C. Atkeson, and G. Cheng. Learning tasks from observation and practice. *RAS*, 47(2-3):163–169, 2004.

[Billard *et al.*, 2004] A. Billard, Y. Epars, S. Calinon, S. Schaal, and G. Cheng. Discovering optimal imitation strategies. *RAS*, 47(2-3):69–77, 2004.

[Billard *et al.*, 2008] A. Billard, S. Calinon, R. Dillmann, and S. Schaal. *Springer Handbook of Robotics*, chapter 59. Robot Programming by Demonstration, pages 1371–1394. 2008.

[Buchli *et al.*, 2011] J. Buchli, F. Stulp, E. Theodorou, and S. Schaal. Learning variable impedance control. *Intl. Journal Of Robotics Research*, 2011.

[Calinon and Billard, 2007] S. Calinon and A. Billard. Incremental learning of gestures by imitation in a humanoid robot. In *Humanoids*, pages 255–262, 2007.

[Calinon and Billard, 2008] S. Calinon and A. Billard. A probabilistic programming by demonstration framework handling constraints in joint space and task space. In *IROS*, pages 367–372, 2008.

[Calinon *et al.*, 2007] S. Calinon, F. Guenter, and A. Billard. On learning, representing, and generalizing a task in a humanoid robot. *IEEE Trans. on Systems, Man and Cybernetics, Part B*, 37(2):286–298, 2007.

[Calinon *et al.*, 2010a] S. Calinon, F. D'halluin, E. Sauser, D. Caldwell, and A. Billard. Learning and reproduction of gestures by imitation. *IEEE Robotics and Automation Magazine*, 17(2):44–54, 2010.

[Calinon *et al.*, 2010b] S. Calinon, I. Sardellitti, and D. Caldwell. Learning-based control strategy for safe human-robot interaction exploiting task and robot redundancies. In *IROS*, pages 249–254, 2010.

[Dillmann, 2004] R. Dillmann. Teaching and learning of robot tasks via observation of human performance. *RAS*, 47(2-3):109–116, 2004.

[Dong and Naghdy, 2007] S. Dong and F. Naghdy. Application of hidden markov model to acquisition of manipulation skills from haptic rendered virtual environment. *Robotics and Computer-Integrated Manufacturing*, pages 351–360, 2007.

[Evrard *et al.*, 2009] P. Evrard, E. Gribovskaya, S. Calinon, A. Billard, and A. Khedda. Teaching physical collaborative tasks: Object-lifting case study with a humanoid. In *Humanoids*, pages 399–404, 2009.

[Gams *et al.*, 2010] A. Gams, M. Do, A. Ude, T. Asfour, and R. Dillman. On-line periodic movement and force-profile learning for adaptation to new surfaces. In *Humanoids*, pages 560–565, 2010.

[Ganesh *et al.*, 2010] G. Ganesh, A. Albu-Schäffer, M. Haruno, M. Kawato, and E. Burdet. Biomimetic motor behavior for simultaneous adaptation of force, impedance and trajectory in interaction tasks. In *ICRA*, pages 2705–2711, 2010.

[Goodrich and Schultz, 2007] M. Goodrich and A. Schultz. Human-robot interaction: A survey. *Foundations and Trends in Human-Computer Interaction*, 1(3):203–275, 2007.

[Gribovskaya *et al.*, 2011] E. Gribovskaya, A. Kheddar, and A. Billard. Motion learning and adaptive impedance for robot control during physical interaction with humans. In *ICRA*, 2011.

[Grollman and Jenkins, 2007] D. Grollman and O. Jenkins. Dogged learning for robots. In *ICRA*, pages 2483–2488, 2007.

[Kormushev *et al.*, 2011] P. Kormushev, S. Calinon, and D. Caldwell. Imitation learning of positional and force skills demonstrated via kinesthetic teaching and haptic input. *Advanced Robotics*, 25(5):581–603, 2011.

[Lee and Ott, 2010] D. Lee and C. Ott. Incremental motion primitive learning by physical coaching using impedance control. In *IROS*, pages 4133–4140, 2010.

[Rabiner, 1989] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, pages 257–286, 1989.

[Rozo *et al.*, 2010] L. Rozo, P. Jiménez, and C. Torras. Sharpening haptic inputs for teaching a manipulation skill to a robot. In *IEEE Intl. Conf. on Applied Bionics and Biomechanics*, 2010.

[Shannon, 2001] C. E. Shannon. A mathematical theory of communication. *SIGMOBILE Mobile Computing and Communications Review*, 5:3–55, 2001.