2008 IEEE/RSJ International Conference on Intelligent Robots and Systems
Acropolis Convention Center
Nice, France, Sept, 22-26, 2008

Learning Nonparametric Policies by Imitation

David B. Grimes and Rajesh P.N. Rao
Department of Computer Science
University of Washington
Seattle, WA 98195
grimes,rao@cs.washington.edu

Abstract— **A long cherished goal in artificial intelligence has been the ability to endow a robot with the capacity to learn and generalize skills from watching a human teacher. Such an ability to learn by imitation has remained hard to achieve due to a number of factors, including the problem of learning in high-dimensional spaces and the problem of uncertainty. In this paper, we propose a new probabilistic approach to the problem of teaching a high degree-of-freedom robot (in particular, a humanoid robot) flexible and generalizable skills via imitation of a human teacher. The robot uses inference in a graphical model to learn sensor-based dynamics and infer a stable plan from a teacher's demonstration of an action. The novel contribution of this work is a method for learning a nonparametric policy which generalizes a fixed action plan to operate over a continuous space of task variation. A notable feature of the approach is that it does not require any knowledge of the physics of the robot or the environment. By leveraging advances in probabilistic inference and Gaussian process regression, the method produces a nonparametric policy for sensor-based feedback control in continuous state and action spaces. We present experimental and simulation results using a Fujitsu HOAP-2 humanoid robot demonstrating imitation-based learning of a task involving lifting objects of different weights from a single human demonstration.**

I. INTRODUCTION

Science fiction stories are rife with instances of "robot see, robot do": a robot watches a human perform an action and in a short amount of time, replicates and generalizes the behavior in a myriad different situations.

Learning by imitation, long recognized as a crucial means for skill and general knowledge transfer between humans, only recently has become an active research area in robotics and machine learning communities. Robotics researchers have become increasingly interested in learning by imitation (also known as "learning from demonstration") as an attractive alternative to manually programming robots.

Researchers in artificial intelligence and robotics have long sought to achieve this goal [2], [27] but have had to confront a number of problems including learning in the presence of uncertainty, learning in high-dimensional spaces, lack of processing speed for real-time computation, etc.

Early work of Kuniyoshi, Inaba and Inoue termed the approach "learning by watching" [27]. Kuniyoshi et al. identified key components necessary for any such imitation learning system. These include functional units for segmentation and recognition of human actions and motion, as well as an algorithm for constructing an imitative motor plan for a given robotic platform.

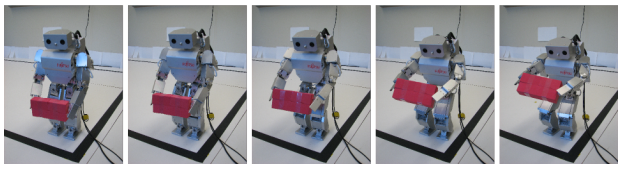

Fig. 1. **The HOAP-2 humanoid robot performing the task of lifting and offering an object.** The motion was learned via imitating by observing a human demonstrator perform the desired motion. Using an inference-based motion planning algorithm a stable imitative motion is obtained.

Researchers have since studied imitation learning in robots using a wide array of techniques from many sub-fields of electrical engineering, mechanical engineering, and computer science. Some recent examples on robotic imitation learning include, for example [11], [16], [18], [22], [28]. Some approaches rely on producing imitative behaviors using non-linear dynamical systems (e.g., [10]) while others focus on biologically motivated algorithms (e.g., [3]), or in achieving goal-directed behaviors (e.g., [4], [5], [24]).

In this paper, we propose a new probabilistic approach to learning by imitation in high degree-of-freedom robots (e.g., humanoid robots). The key contribution of this work is a method for learning a nonparametric policy which uses sensory feedback to generalize pre-computed action plans (based on fixed tasks) to a broader space of task variation. The goal is to obtain imitative behaviors which are reactive to different task and/or environmental conditions such a lifting a heavier object than originally planned. Although this paper builds upon previously published algorithms which yield fixed action plans [8], [9], the problem which we seek to solve here is distinct. For completeness this paper summarizes the planning and dimensionality reduction methods used, however for full details we refer the reader to [8], [9].

Our approach leverages recent advances in probabilistic reasoning, learning, and dimensionality reduction, and allows a humanoid robot to learn nonparametric policies directly from human demonstrations. The proposed method does not rely on any physics-based models (e.g., mass and rotational inertia properties) of the robot or environment; rather, a sensor-based dynamics model is learned during the course of imitation and used for inferring a stable plan (sequence of actions). This plan is then used to bootstrap the learning of a nonparametric policy that generalizes the imitated behavior

978-1-4244-2058-2/08/$25.00 ©2008 IEEE.

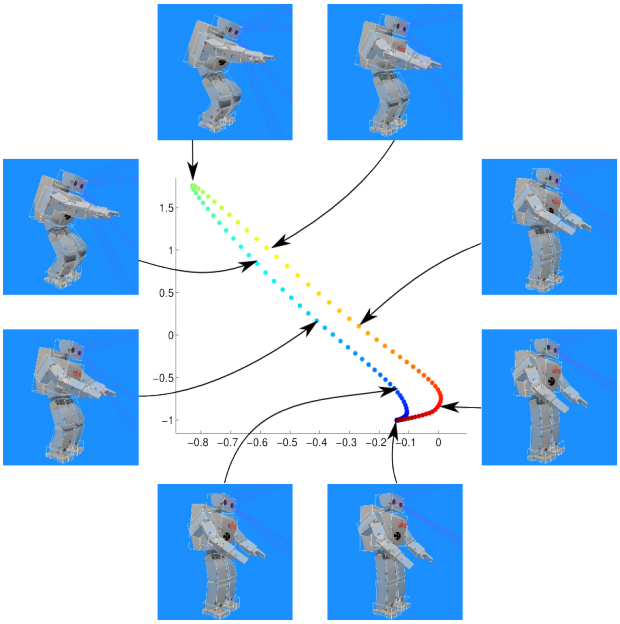

Fig. 2. **Example latent kinematic space used in policy learning.** Postures are represented in a low-dimensional 2D latent space to afford a compact representation of both robot state and control signal. The state history of a stabilized task plan is shown (clockwise with time represented by the change from blue to red). Example kinematic configurations are shown and indicate the corresponding points in the 2D space.

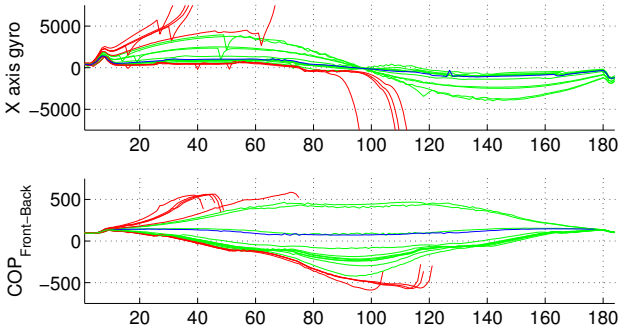

Fig. 3. **Dynamics sensor feedback during motion optimization.** Sensory signals are shown for each trial of the imitation-based planning algorithm. Dynamically unstable trials are shown in red, while stable ones are shown in green. The dynamics signals corresponding to the inferred plan are shown in blue. The upper plot demonstrates the (simulated) gyroscope signal which measures the angular rate of rotation about the X-axis (a vector emanating from mid-torso out to the right side of the robot). The lower plot shows the (simulated) center of pressure (COP) in the frontal-backward direction.

to a range of conditions not encountered in the teacher demonstration.

We illustrate our approach in the context of a task where the robot has to learn to lift and move objects of differing properties (such as mass), given only the following:

- A human teacher who demonstrates the task with a single object
- A kinematic correspondence between the human demonstrator and the humanoid robot

We show, using a simulated Fujitsu HOAP-2 humanoid robot, that the approach allows the robot to learn a non-parametric (Gaussian process) policy for lifting objects of different weights given only a single human demonstration. Unlike methods such as [9] which plan a sequence of imitative actions to be executed open-loop, this paper seeks to addresses the issue of using closed-loop sensory feedback to generalize to novel situations and environments.

II. POLICY LEARNING VIA IMITATION

A. Problem definition

Consider an agent in an environment whose state is characterized by the continuous (real-valued) multivariate vector $\mathbf{s}_t \in \mathcal{S} = \mathbb{R}^{d_\mathbf{s}}$. We assume that the state is either directly observable or can be robustly estimated from sensory observations using a temporal filtering algorithm. At each discrete point in time t the agent must select an action $\mathbf{a} \in \mathcal{A} = \mathbb{R}^{d_\mathbf{a}}$ to execute. An action in our setting is also a continuous multivariate vector. Motivated by the humanoid learning problem described in Section I we assume that the stochastic transition function based on the conditional probability distribution $P(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t)$ is unknown.

For some particular parameterized task $\mathcal{T}(\theta)$ we assume there exists a function $F_{\mathcal{T}(\cdot)} : \mathcal{S} \rightarrow \mathcal{A}$ which defines a policy mapping states to actions. Additionally we assume a known initial state $\mathbf{s}_1 \in \mathcal{S}$, and a known terminal or goal state indicator $\mathcal{G} : \mathcal{S} \rightarrow \{0, 1\}$. I.e. when the agent observes that $\mathcal{G}(\mathbf{s}_t) = 1$ execution stops. By starting in state $\mathbf{s}_1$ and selecting actions according to the policy function $F_{\mathcal{T}(\cdot)}$ the agent visits the sequence of states $(\mathbf{s}_1, \mathbf{s}_2, \cdot\cdot, \mathbf{s}_T)$ where $\mathcal{G}(\mathbf{s}_T) = 1$.

We consider a particular case of policies in which the actions taken maximize the (expected) likelihood of the state sequence with respect to a (given) probabilistic constraint model $P_C(\cdot|\cdot)$:

$$\mathcal{L} = \prod_{t=1}^{T} P_C(\mathbf{c}_t|\mathbf{s}_t) \tag{1}$$

A task $\mathcal{T}(\theta)$ is parameterized by a set of environmental or task configuration variables that dictate the underlying transition function. A valid policy $F_{\mathcal{T}(\cdot)}$ is defined as one which solves the task $\mathcal{T}$ for some valid range of values for the parameter θ. This framework assumes that the task parameters are not directly observable by the agent during execution and therefore not represented in the robot's state $\mathbf{s}_t$. Note that the policy function $F_{\mathcal{T}(\cdot)}$ does not depend on a particular value of the parameter θ. Rather, information about θ is encoded by the sensory components of the state representation. Consider the problem of a humanoid robot lifting objects with varying density. The intrinsic task parameter θ would represent the object's mass, however the policy would be represented in terms of the sensory consequences of the various torques generated by picking up a heavier or lighter object. This framework also handles multivariate task parameters, for instance allowing for variation in both the mass and moments of inertia of an object.

The approach relies on the presence of an offline imitation-based planning method which given the observed demonstration by the teacher and a task parameter setting θ^i can formulate a sequence of actions to be executed open-loop which (approximately) maximizes the expected value of Equation 1. As has been shown in [9], an inference based method leveraging a sensor-based dynamic balance model can efficiently plan such a sequence of actions which will produce dynamically stable imitative motions in the HOAP-2 humanoid robot.

The goal of this paper is to learn the unknown policy function $F_{\mathcal{T}(\cdot)}$ from a small number of planning iterations using a few values of θ^i. We solve this problem by treating the policy as as an unobserved function and modeling it via the posterior distribution:

$$P\left(F_{\mathcal{T}(\cdot)}(\mathbf{s})|\mathbf{s}, \{H_i\}_{i=1}^{M}\right) \tag{2}$$

where $\{H_i\}_{i=1}^{M}$ represents a history of past data obtained by the robot via exploration (see Section III for details).

This policy will then generate actions which successfully perform the task $\mathcal{T}$ for a large range of task parameter values. Here a successful trial is defined by one that reaches the end/goal condition while maximizing the expected state sequence likelihood. An example of a parameterized task is shown being performed by a Fujitsu HOAP-2 humanoid robot in Figure 1. We refer to this example motion as a motivating example throughout the paper, and demonstrate the results of our proposed method in Section IV.

The desire to minimize re-planning is motivated by the fact that for many applications planning is only practical to be done offline, while we want the robot to be instantly reactive to many different task situations. While many fast, online planning algorithms exist (for example see [21]), typically such methods do not allow for learning a new environment without complex physics modeling. Additionally, many such methods are based on deterministic algorithms which can be very sensitive to estimation errors of environmental parameters.

B. Latent posture space and dynamics representation

The reduction of the space $\mathcal{S} \times \mathcal{A}$ is crucially important for the proposed use of nonparametric policy learning via the use of Gaussian processes. Representing and learning the policy in the full 25×25 dimensional space (ignoring dynamics) is obviously intractable. Fortunately, with respect to a certain policy, the full number of degrees of freedom (25 in the HOAP-2) is highly redundant. Additionally, dynamics relevant to a particular task can be efficiently represented by low-dimensional sensor-based quantities.

Linear principal components analysis (PCA) is used to represent a latent posture space but other non-linear embedding techniques (such as the GPLVM [15]) may allow for further reduction in dimensionality. In this paper we use a linear projection of the full posture space onto a two dimensional space as shown in Figure 2. The choice of two dimensions is based on the fact that the first two principal components contain 98% of the total kinematic variance

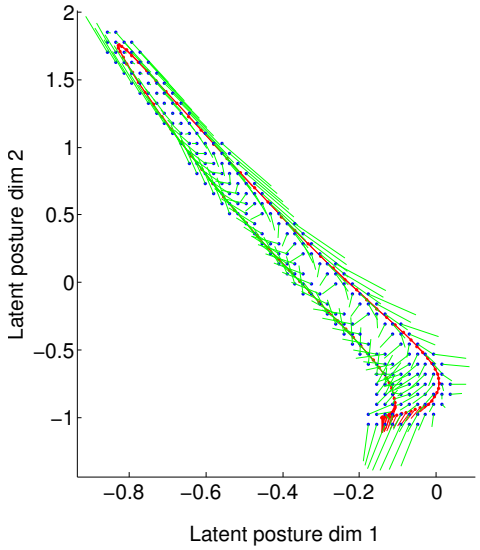

Fig. 4. **Nonparametric policy example.** A simple nonparametric policy is depicted for visualization purposes. In this case a policy is learned in a kinematic 2D space. The policy is learned from a single history of states and actions (shown in red). The resulting actions (green arrows) are shown by evaluating the learned policy on a grid of states (blue points). Leveraging the full posterior distribution over actions afforded by the Gaussian process model, actions are only shown if the entropy of the action distribution is below some threshold.

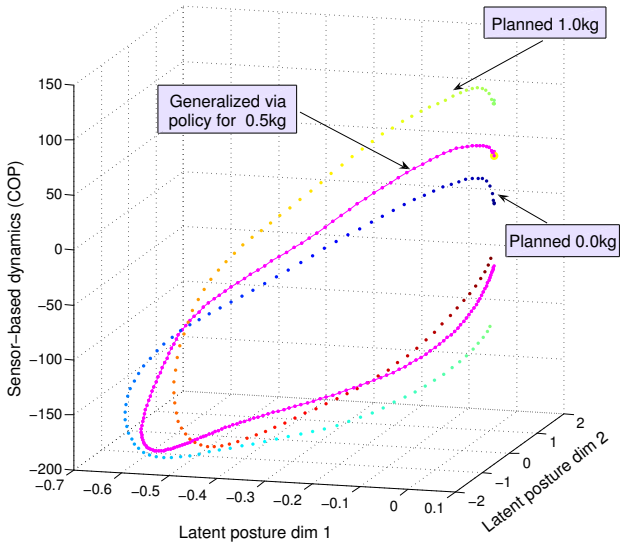

Fig. 5. **A learned policy for generalizing an object lifting motion.** For the task depicted in Fig. 1 results of a policy trained on two action sequences corresponding to 0.0 kg and 1.0 kg objects are shown. The policy can sufficiently generalize to a novel and unknown object mass of 0.5kg. In this example the learned policy maps a 3D state representation of two latent posture dimensions and a foot sensor based dynamics dimension to actions in the 2D latent posture space. The sequence of states for the 0.0kg case are shown changing from blue to green, for 1.0kg changing from green to red, and the policy result is shown in magenta.

in the original motion. Assuming positional control of the servos, the reduced posture space is also used to represent actions.

In order to respond to the particular environment dictated by the task parameter θ the latent kinematic state is augmented by a set of sensory state variable which allow for reacting properly for θ. Note that this does not mean that θ is directly observed. Rather in the lifting of different mass objects example, we utilize a foot-pressure sensor derived quantity closely related to the center of pressure (COP). The HOAP-2 humanoid robot has four pressure sensors on each corner of each foot. Thus the front to back COP (for dual contact point motions) is simply obtained by subtracting the sum of the back four sensors from the sum of the front four. In concert with the two dimensional posture space, the COP measurement can detect a different mass load on the robot. Thus for the results described in Section IV the learned policy maps a 3D state (2 kinematic + 1 dynamics dimensions) to 2D action. An example of dynamics-sensory observations is shown in Figure 3.

C. Inference based planning

We briefly describe the inference-based motion imitative planning method used in this paper. In principle many other approaches to planning could be employed. Recently other probabilistic and Bayesian approaches to sensorimotor learning have been proposed allow a robot to learn from empirical exploration. For example Bayesian locally weighted regression (BLWR) [23] and the randomly varying coefficient (RVC) [6] model tackle the problem of efficiently performing nonparameteric regression. Ko et al. proposed using a hybrid approach to combining both a physics-based model and learning based on an "unscented" approximation of the Gaussian process posterior [12]. However, none of these approaches consider the problem of planning imitative behaviors or have been demonstrated in the humanoid domain. Deterministic motion planning algorithms (for example see [13], [20]) are in widespread use, but require a complex physics-based model to specified in advance.

For details of the planning method used see [8], [9]. The method produces a sequence of actions (HOAP-2 motor commands) which yield a dynamically stable imitative motion.

First, pose estimates are obtained for the human demonstrator using an optical motion capture system. The central idea of the method is to iteratively infer and learn a dynamic Bayesian network representation of the dynamics, and to probabilistically constrain actions (via a sensor based representation of the dynamics). This approach allows for finding dynamically stable motions without requiring *a priori* knowledge of the robot's dynamic properties. Using a constrained exploration algorithm a probabilistic sensorimotor prediction model is learned directly from actuation and sensory information. The dynamics feedback used for planning the object lifting task is shown in Figure 3.

In the examples presented in this paper, the planner is able to obtain an imitative sequence of actions $\mathbf{a}^*_{1:T}$ with high posterior likelihood given the human pose estimates $\mathbf{e}_{1:T}$ and a set of dynamics constraint targets $\mathbf{c}_{1:T}$:

$$\mathbf{a}^*_{1:T} = \text{argmax}_{\mathbf{a}_{1:T}} P(\mathbf{a}_{1:T}|\mathbf{e}_{1:T}, \mathbf{c}_{1:T}). \quad (3)$$

We set the constraint targets to regions of the dynamics state space for which the robot is empirically stable. Specifically we represent the dynamics state of the robot using two axes of the torso gyroscope, and a set of features of the eight foot pressure sensors located on the bottom of the feet. These features can be thought of as generalizing the commonly used center of pressure (COP) measure. The dynamics constraint targets $\mathbf{c}_{1:T}$ are all set to be highly peaked at zero torso rotation and zero pressure difference. Note that while these targets could be time-dependent, for this application we use a single target value across all time steps. Also we note that additional objectives can be incorporated by describing them as probabilistic constraints in the form of a likelihood function $P(\mathbf{c}'|\mathbf{s})$ which is conditioned on any candidate state.

III. NONPARAMETRIC POLICY LEARNING FROM IMITATIVE PLANS

We now propose a method for learning a policy for a large space task parameters from a small number of imitative plans (generated off-line). Our method begins by simply executing each plan $\mathbf{a}^{\theta_i}_{1:T}$ associated with a particular instance of the parameterized task. By recording the sequence of states visited by executing each plan the history $H_i = \{\mathbf{a}^{\theta_i}_{1:T}, \mathbf{s}^{\theta_i}_{1:T}\}$ is formed. The goal is to represent the unknown policy function using the information contained in the set of histories $\{H_i\}_{i=1}^M$. Gaussian process priors provide an elegant way to approach the unknown policy function learning in a Bayesian manner.

Gaussian process regression [7], [14], [17], [26] is used to model the posterior of the output of the policy function. As Gaussian process regression is generally restricted to dealing with univariate functions we simply model each dimension of the action space $\mathcal{A}$ independently. Specifically the j-th dimension of the action $a^{(j)}$ is modeled as being normally distributed given a input state vector $\mathbf{s}$.

$$P\left(F_{\mathcal{T}(\theta)}(\mathbf{s})^{(j)}|\mathbf{s}\right) = \mathcal{N}\left(a^{(j)}; \mu^{(j)}_*, \Lambda^{(j)}_*\right) \quad (4)$$

For convenience let S_H be a column oriented matrix of all states in the set of histories. For instance if the state space is reduced to three dimensions, then S_H is a $3 \times MT$ matrix. Likewise A_H is a matrix of all actions where the action in the j-th column was taken while in the state corresponding to the j-th column of S_H. The mean and variance of the action distribution are functions of the $n \times n$ covariance matrix $K(S_H, S_H)$ where $n = MT$ denotes the number of values in the history set. The covariance matrix is defined by the covariance (or kernel) function $k(\cdot, \cdot)$. The squared exponential kernel is also referred to as the radial basis function (RBF) kernel:

$$k(\mathbf{x}_p, \mathbf{x}_q) = \sigma_f^2 \exp\left(-\frac{1}{2}(\mathbf{x}_p - \mathbf{x}_q)'\Sigma_l^{-1}(\mathbf{x}_p - \mathbf{x}_q)\right) + \sigma_n^2 \delta_{pq} \quad (5)$$

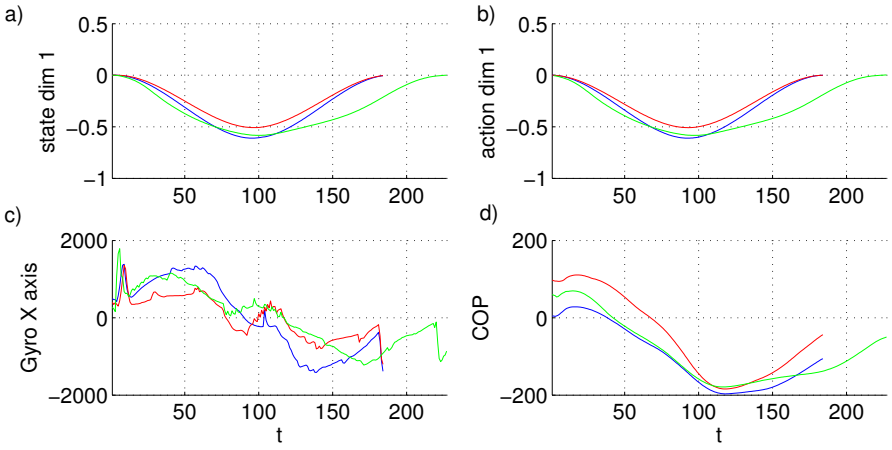

Fig. 6. **Online policy vs planned actions and sensor readings.** All plots compare the results of the 0.0kg plan (red) and the offline planned actions for 0.5kg (blue) with the result of executing the learned policy (green). (a) The state in the first latent dimension, (b) the actions executed (for the first latent dimension, (c) the X-axis gyroscope sensor readings, (d) the front to back center of pressure (COP) difference.

The squared exponential kernel, due to its modeling flexibility is considered an appropriate choice when one doesn't have domain specific knowledge about how the output of the process (co)varies as a function of its inputs. The exact form of the kernel is selected by setting the kernel hyperparameters $\Sigma_l^{-1}, \sigma_f^2, \sigma_n^2$ representing the length scales of each of the inputs (which determine how sensitive an output dimension is to changes in each input dimension), the noise-free function variance, and the additive process noise variance respectively. Hyperparameters were optimized using a standard scaled conjugate gradient (SCG) approach, maximizing the log likelihood of the data $\{H_i\}_{i=1}^M$ over these hyperparameters. In principle local minima can be a problem with SCG on non-convex functions, yet empirically we found that multiple runs with different initial parameters all converged to very similar hyper-parameters.

By evaluating the kernel function for each pair of data points in S_H, we can compute the kernel matrix $K(S_H, S_H)$ and its inverse necessary for the predicted posterior mean:

$$\mu_*^{(j)} = K(\mathbf{s}, S_H)\Psi A_H^{(j)} \tag{6}$$

and variance:

$$\Lambda_*^{(j)} = K(\mathbf{s}, \mathbf{s}) - K(\mathbf{s}, S_H)\Psi K(S_H, \mathbf{s}) \tag{7}$$

where:

$$\Psi = \left[K(S_H, S_H) + \sigma_n^2 I\right]^{-1}. \tag{8}$$

The Gaussian process equations above require $\mathcal{O}(n^3)$ operations which can prohibitive when either the number of histories or the motion length grows large. Fortunately, important recent advances have been made in approximate or "sparse" GPs which can accommodate much larger quantities of data [1], [14], [19].

An illustrative example of a nonparametric policy obtained from a single history is shown in Figure 4.

IV. RESULTS

We present the results of policy learning for the task of a humanoid robot lifting a variety of different objects. The task parameter θ is simply the mass of the object in kilograms. The object grasping problem is simplified by not considering objects of different sizes or with different grasp requirements. The results presented here utilize the Webots dynamic simulator [25].

The first step was to record motion capture data of the human demonstrator performing the desired lifting and offering action on an object (a basketball with approximate mass 0.5kg). From this data a latent posture space is induced as described in Section II-B. Time is discretized and actions are planned for the robot at regular 32 millisecond intervals for the approximately 7 second duration motion. Using the technique mentioned in Section II-C we obtain a sequence of actions optimized for the case where $\theta = 0$. After execution (shown in Figure 7(a)) the history $\{\mathbf{a}_{1:T}^0, \mathbf{s}_{1:T}^0\}$ is obtained for the successful performance of the task.

We then experimented with different mass objects, and observed that with objects of mass greater than 0.4 kg the open-loop plan $\mathbf{a}_{1:T}^0$ is no longer stable (as shown in Figure7(b)). We then re-plan for the $\theta = 1.0$ kg case and learn a nonparametric policy based on the method described in Section II-C. Empirical results show that the proposed policy learning method is able to generalize information from the open-loop plans to provide feedback based control for novel task parameters. For example the learned policy can be used to control the robot in the 0.5 kg case as shown in Figure 7(c). In fact we found that the learned policy was stable and successfully completed motions for θ between 0 and 1.2 kg. Above 1.2 kg the initial pose selected is no longer statically stable, and the torque necessary to simply hold an object of this mass with arms fully extended exceeds the specifications for the ankle servo motors. This suggests that our method was able to generalize the task to the full range

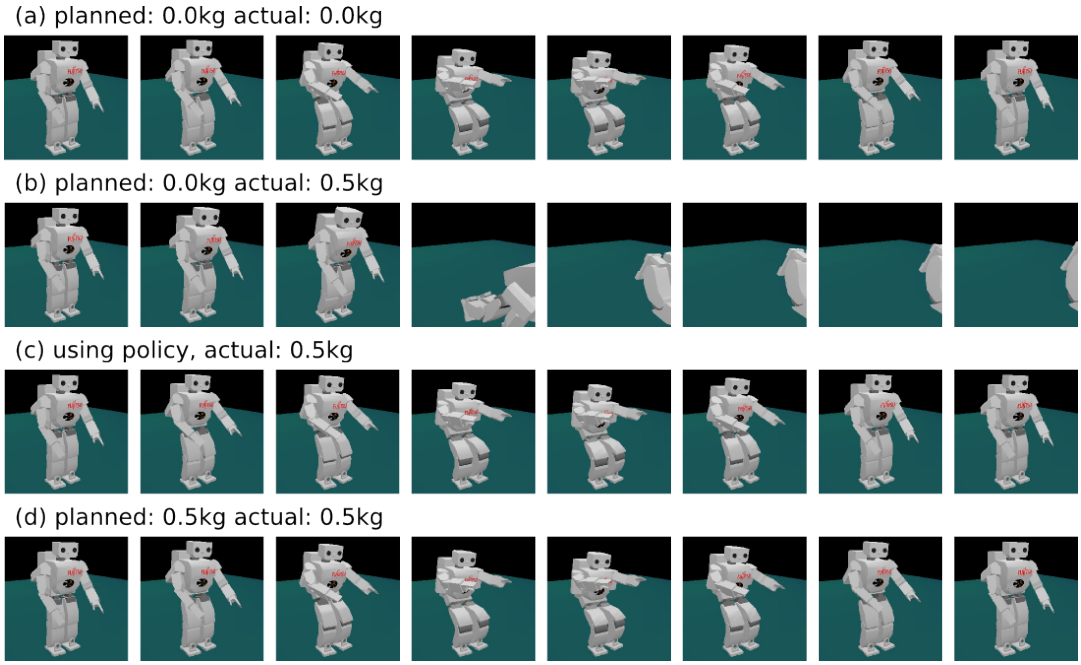

Fig. 7. **Policy results for task generalization.** (a) Webots simulation of the HOAP-2 robot executing the actions planned in the case of no extra mass. (b) The unstable result of using the actions planned for no extra mass in the case of a 0.5kg mass. Note that for technical reasons the extra mass is not visibly rendered. (c) The stable result of using the learned policy to generalize to the novel 0.5kg extra mass condition. Note that this result is visually indistinguishable from the first row, see text for a quantitative comparison. (d) For comparison the execution of the action plan for 0.5 is also shown to demonstrate the similarity between the on-line policy and the off-line planning result.

of feasible variability within this task. We also found that the dynamics feedback signals obtained during policy use matched closely with those from the offline motion planner shown in Figure 6. Analytic analysis of our method is beyond the scope of this paper. This is due in part to the fact that many properties of Gaussian Processes cannot be currently described analytically. We refer the interested reader to [1] for a survey of the issues in providing bounds on Gaussian Process errors.

To illustrate the need for a non-linear regression model which can fit data locally we also attempted to represent the policy using a linear-least squares regression model. On a set of hold out data, a linear fit to the policy data shown in Figure 5 yielded a mean squared error (MSE) of 0.339 whereas the GP regression fits the data with an MSE of 0.023.

Note that our reliance on the simulator is based only on practical concerns such as wear on the robot during learning and the amount of human interaction required. Importantly, note that we do not access the internal dynamic state of the simulator or perform expensive computation between simulation steps. Additionally, we note that in our previous work we have found the Webots simulator to be very accurate in simulating real-world dynamics. For instance we were able to learn complex motion plans [8] (such as balancing on one leg) via the simulator and directly transfer the result to the HOAP-2 robot. Thus we strongly believe that the results we have obtained here have significance in terms or real-world performance. We are in the process of developing the evaluation of a learned policy in real-time on the HOAP-2 humanoid robot. This is mainly a programming challenge as we have determined that the evaluation of the GP based policy can be indeed be performed in realtime.

V. CONCLUSION

We have proposed a new probabilistic technique for imitation-based learning in high degree-of-freedom robots. The approach combines dimensionality reduction, inference-based planning, and Gaussian process regression to produce a nonparametric policy that significantly generalizes the plan produced from the teacher demonstration. The technique allows sensory feedback-based control to be learned in high-dimensional continuous state and action spaces given a few teacher demonstrations. We demonstrated the viability of the approach using a simulated Fujitsu HOAP-2 humanoid robot in the context of a task involving lifting objects of different weights. Given a single demonstration by a human, the robot was able to learn a policy that generalized across different weights.

A number of issues remain to be addressed. For example, how effective and stable are the policies learned in simulation when executed by the robot? Earlier results with inference-based planning seem to suggest that some results carry over but others require some re-learning on the robot. How does the approach scale to more complex behaviors, for example, those involving locomotion? To what extent is the general-

ization provided by the Gaussian process policy appropriate for various behaviors? Can the approach be extended to hierarchical policies for solving more complex tasks, such as picking an object, manipulating it, and transporting it to a different location? Our current efforts are focused on addressing some of these issues.

VI. ACKNOWLEDGMENTS

This research was supported by NSF, the ONR Adaptive Neural Systems program, the Sloan Foundation, and the Packard Foundation.

REFERENCES

[1] *Bayesian Gaussian Process Models: PAC-Bayesian Generalisation Error Bounds and Sparse Approximations*. PhD thesis, University of Edinburgh, 2003.

[2] C. Atkeson and S. Schaal. Robot learning from demonstration. In *Proceedings of the Fourteenth International Conference on Machine Learning (ICML'97)*, pages 12–20, 1997.

[3] A. Billard and M. Mataric. Learning human arm movements by imitation: Evaluation of a biologically-inspired connectionist architecture. *Robotics and Autonomous Systems*, 37(941):145–160, 2001.

[4] S. Calinon, F. Guenter, and A. Billard. On learning, representing and generalizing a task in a humanoid robot. *IEEE Transactions on Systems, Man and Cybernetics, Part B. Special issue on robot learning by observation, demonstration and imitation*, 37(2):286–298, 2007.

[5] J. Demiris and G. Hayes. A robot controller using learning by imitation. In *Proceedings of the 2nd International Symposium on Intelligent Robotic Systems (IROS'94)*. IEEE Press, July 1994.

[6] N. U. Edakunni, S. Schaal, and S. Vijayakumar. Kernel carpentry for online regression using randomly varying coefficient model. In M. M. Veloso, editor, *IJCAI*, pages 762–767, 2007.

[7] M. N. Gibbs. *Bayesian Gaussian Processes for Regression and Classification*. PhD thesis, University of Cambridge, 1997.

[8] D. B. Grimes, R. Chalodhorn, and R. P. N. Rao. Dynamic imitation in a humanoid robot through nonparametric probabilistic inference. In *Proceedings of Robotics: Science and Systems (RSS'06)*, Cambridge, MA, 2006. MIT Press.

[9] D. B. Grimes, D. R. Rashid, and R. P. N. Rao. Learning nonparametric models for probabilistic imitation. In *Advances in Neural Information Processing Systems 19 (NIPS'06)*. MIT Press, Cambridge, MA, 2007.

[10] A. J. Ijspeert, J. Nakanishi, and S. Schaal. Trajectory formation for imitation with nonlinear dynamical systems. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'01)*, pages 752–757. IEEE Press, 2001.

[11] T. Inamura, I. Toshima, and Y. Nakamura. Acquiring motion elements for bi-directional computation of motion recognition and generation. In *Experimental Robotics VIII*, pages 372–381. Springer, 2003.

[12] J. Ko, D. Klein, D. Fox, and D. Hahnel. GP-UKF: Unscented Kalman filters with gaussian process prediction and observation models. In *Proceedings of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS'07)*. IEEE Press, 2007.

[13] J. J. Kuffner, K. Nishiwaki, S. Kagami, M. Inaba, and H. Inoue. Motion planning for humanoid robots under obstacle and dynamic balance constraints. In *Proceedings of the IEEE International Conf. Robotics and Automation (ICRA'01)*, pages 692–698. IEEE Press, 2001.

[14] N. Lawrence, M. Seeger, and R. Herbrich. Fast sparse gaussian process methods: The informative vector machine. In S. T. S. Becker and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15 (NIPS'02)*, pages 609–616. MIT Press, Cambridge, MA, 2003.

[15] N. D. Lawrence. Gaussian process latent variable models for visualization of high dimensional data. In *Advances in Neural Information Processing Systems 15 (NIPS'02)*. MIT Press, Cambridge, MA, 2003.

[16] Y. N. M. Okada, K. Tatani. Polynomial design of the nonlinear dynamics for the brain-like information processing of the whole body motion. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'02)*, pages 1410–1415. IEEE Press, 2002.

[17] C. E. Rasmusen and C. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[18] S. Schaal, A. Ijspeert, and A. Billard. Computational approaches to motor learning by imitation. *The Neuroscience of Social Interaction*, 1(1431):199–218, 2004.

[19] E. Snelson and Z. Ghahramani. Sparse gaussian processes using pseudo-inputs. In Y. Weiss, B. Scholkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18 (NIPS'05)*, pages 1259–1266. MIT Press, Cambridge, MA, 2006.

[20] M. Stilman, C. Atkeson, J. Kuffner, and G. Zeglin. Dynamic programming in reduced dimensional spaces: Dynamic planning for robust biped locomotion. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'05)*, pages 2399 – 2404, April 2005.

[21] M. Stilman and J. Kuffner. Navigation among movable obstacles: Real-time reasoning in complex environments. In *Proceedings of the 2004 IEEE International Conference on Humanoid Robotics (Humanoids'04)*, volume 1, pages 322 – 341, December 2004.

[22] Y. Takahashi, K. Hikita, and M. Asada. Incremental purposive behavior acquisition based on self-interpretation of instructions by coach. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'03)*, pages 686–693. IEEE Press, 2003.

[23] J. Ting, A. D'Souza, S. Vijayakumar, and S. Schaal. A bayesian approach to empirical local linearization for robotics. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'04)*. IEEE Press, 2008.

[24] D. Verma and R. P. N. Rao. Planning and acting in uncertain environments using probabilistic inference. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'06)*. IEEE Press, 2006.

[25] Webots. http://www.cyberbotics.com. Commercial Mobile Robot Simulation Software.

[26] C. K. I. Williams and C. E. Rasmussen. Gaussian processes for regression. In D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems 8*, pages 514–520. MIT Press, 1996.

[27] M. Y. Kuniyoshi and H. Inoue. Learning by watching: Extracting reusable task knowledge from visual observation of human performance. *Transaction on Robotics and Automation*, 10(6):799–822, December 1994.

[28] K. Yamane and Y. Nakamura. Dynamics filter - concept and implementation of on-line motion generator for human figures. *IEEE Transactions on Robotics and Automation*, 19(3):421–432, 2003.