Structure and Growth: A Model of Development for Grasping with Robot Hands

R.A. Grupen J.A. Coelho, Jr.
Department of Computer Science – Laboratory for Perceptual Robotics
140 Governors Dr., Amherst, MA 01003
{grupen, coelho}@cs.umass.edu

Abstract

According to recent theories of sensorimotor development in biological systems, the dynamics of physical interaction with the world encodes control knowledge. Control is derived by reinforcing and learning to predict constructive patterns of interaction, and behavior is an artifact of coupled dynamical systems with a number of controllable degrees of freedom. For grasping and manipulation, we propose a closed-loop control process that is parametric in the number and identity of contact resources. In this paper, we will show how control decisions can be made by estimating patterns of membership in a family of prototypical dynamic models. A grasp controller can thus be tuned continuously on-line to optimize performance over a variety of object geometries. This same process can be used to estimate the haptic category in which the object resides. We will illustrate how a grasping policy that is incrementally optimal for several objects can be acquired using our Salisbury hand with tactile sensor feedback.

1 Introduction

The human hand has often been cited as an important factor in the development of the ability of the human brain to form critical categories in sensorimotor experience. Many are of the opinion that this faculty for building predictive models underlies much of what we recognize as human-level cognitive ability.

While experts disagree on cause and effect, it is clear that the mechanical dexterity and redundancy afforded in the hand requires a neural architecture capable of modeling a huge variety of interactions with the world. We postulate that the processes underlying multifaceted world models effect problem solving in general as well as the formulation of skillful manipulation strategies.

In this paper, we provide a biologically-inspired account of how sensorimotor strategies are formed. The process is constrained to employ a compact instruction set in the form of a closed-loop control basis. Each controller in the control basis must satify safety, schedulability, and developmental constraints. The decision regarding which controller to activate next is described as a Markov decision problem (MDP), which state is based on empirically-derived models describing the dynamics induced by the set of control primitives. These generative models allow the agent to consider all possible outcomes and select the control primitive with highest utility in each context. The representation proposed is applied to the multifingered grasping problem, in which the agent must grasp objects with unknown geometries, independent of the initial relative orientation between hand and object. The results obtained show that indeed the representation proposed allows the robot to adapt its control strategy to each object (or context), even in the absence of specific sensors for object identity.

2 Literature Review

This work gathers insight from developmental psychology, control theory, reinforcement learning, robotics and artificial intelligence to look for mechanisms with which to program robot hands automatically by direct haptic and visual interaction with manipulation tasks.

2.1 Developmental Psychology

Infants are born with neural and skeletomuscular systems that produce patterned and timed movements. During the first several months in an infant's life, reflexive responses begin to organize into coherent motor strategies, sensory modalities are coordinated and attentional mechanisms begin to emerge. Native reflexive responses like the primary walking reflex and the palmar grasp reflex [1] provide primitive, closed-loop sensorimotor behavior that accomplish sensory-driven work in the world. Subsequently, policies for coordinating multiple sensory and motor modalities appear as primary circular reactions [12] which are practiced until the infant finds it possible to prolong certain interactions with the world.

The *interactionist* account grounds human knowledge in *activity*. From this perspective, "...motor timing in skilled actions is discovered ... through perceptual exploration of the body's intrinsic (or autonomous) dynamics within a changing task and physical space[16]." This perspective is consistent with recent trends in robotics as well, where roboticists actively develop artificial muscles and neural oscillator models [6, 15, 17] to exploit the intrinsic dynamics of the control process. For example, series elastic actuators[13] are designed to provide an appropriate passive behavior to the limb. These approaches minimize muscular effort by looking for synergistic kinematic and dynamic coupling between the robot and the world. Moreover, such methods are generally robust to changing environments and other perturbations.

2.2 Grasping

Much progress has been made in the mathematical analysis of phenomena associated with grasping and manipulation tasks. We have relatively standard models of contact types [9, 3], including soft-point contacts, and point contacts with and without friction. Some of the most widely read literature on grasping concerns the form closure problem, in which the placement of frictionless point contacts so as to fully restrain an object is studied [8]. Force closure is arguably more appropriate for grasp planning and control, since this condition speaks to the ability of a grasp to reject disturbance forces; the analysis usually considers frictional forces[5, 11, 4].

Despite the significant theoretical impact of this literature, we have not yet developed an adequate model of the sensory and motor process of grasping and manipulation. This process moves fluidly through multiple contact regimes and can trade a margin of stability early in a manipulation strategy for constructive interactions, e.g. pick-and-place constraints, late in the strategy. Moreover, nearly all the work on multifingered grasping considers a complete geometrical model of the object and most depend on geometrical reasoning to compute a grasp - this despite the fact that grasping is inherently a force domain task. Finally, we feel that the real opportunity afforded by multifingered hands is the technology underlying the automatic modeling of complex and non-stationary modes of interaction between a robot and the world. Modeling end-to-end manipulation sequences leads immediately to issues of representation and learning - issues that have been largely ignored to date.

3 Multifingered Grasp Synthesis

Grasp synthesis requires the determination of the grasp configuration parameters (contact normals and relative contact positions with respect to the object's center of mass) that satisfy the minimum requirements set by the users in terms of a given grasp metric. While most researchers describe grasp synthesis as an optimization problem, we proposed that it is best characterized as a robust control problem. In this framework, the robot uses tactile feedback to compute incremental contact displacements, performed in order to optimize the grasp metric of interest.

For the experiments reported here, contact displacements are determined by the grasp controller π_c described by [2]. Contact displacements are computed based on local models of the interaction between the contacts and the object surface, and are aimed at reducing the squared wrench residual ϵ measured at the object's center of mass. Given the wrench residual vector for n contacts

$$\rho = \sum_{i=1}^{n} \left[f_x^i \;\; f_y^i \;\; f_z^i \;\; \tau_x^i \;\; \tau_y^i \;\; \tau_z^i \right]^T ,$$

then the squared wrench residual is defined by

$$\epsilon = \rho^T \rho. \tag{1}$$

The controller π_c displaces the subset c of contacts until a local minimum for ϵ is reached. The minimization of ϵ is closely connected to grasp stability: null wrench residual ($\epsilon = 0$) is a necessary condition for grasp stability [2], as it implies the existence of a null space of rank 1 or higher in the grasp matrix [14].

The subset of contacts c specifies which fingers and surfaces are enlisted in the grasp task. A hand with 3 fingers labeled $\{T, 1, 2\}$ allows for 4 distinct contact subsets, assuming that two or more fingers are required to grasp the object:

$$\mathcal{C} = \{(T,1), \;\; (T,2), \;\; (1,2), \;\; (T,1,2)\}. \tag{2}$$

Each instance of $c \in \mathcal{C}$ defines a new control law, affecting the outcome of the grasp process directly.

The controller π_c is better characterized as an element of a family of grasp controllers, referred to as $\Pi = \{\pi_c | \ c \in \mathcal{C}\}$; the control basis Π represents the agent's native control structure.

The control actions of π_c are independent of the global object geometry, depending solely on instantaneous, local tactile feedback. The convergent configurations for π_c correspond to local minima of ϵ, and for a given contact configuration, each choice of control law $\pi_c \in \Pi$ will lead to distinct convergent grasp configurations. Therefore, there exist an optimal choice in terms of which grasp resources c to use; this choice yields the convergent configuration with the minimum ϵ. The idea can be extended to controller sequencing: given a certain initial configuration, there exists an optimal sequence of controllers that lead the system state to the solution with the smallest possible ϵ. More importantly, controller sequencing expands the capabilities offered by individual controllers and allows one to build a system capable of adapting its strategy to the many operational contexts encountered.

4 Context-Dependent Grasp Policies

The problem to be addressed consists in identifying which sequence of grasp controllers minimizes the coefficient of friction between fingers and object surface required for stability. The object's identity, geometry, and pose with respect to the hand are unknown.

The determination of the optimal switching policy is a sequential decision problem, which can be solved within the Markov Decision Process (MDP) framework, provided one can describe the control switching as transitions between discrete states, within a Markovian state space. The next subsection describes a representation in which state information is derived from a mixture of generative models describing the grasp dynamics in the phase space defined by $(\epsilon, \dot{\epsilon})$. It is assumed that the resulting representation satisfies the Markovian property; this assumption allows us to use the Q-learning algorithm to approximate the optimal switching policy. A description of the experimental setup and parameters follows.

4.1 Experimental Setup

The kinematics of the Stanford/JPL hand were used in the simulated grasp trials. In each trial, the object to be grasped was selected from three possible object types: cylinders, cubes, and triangular prisms. The geometric parameters of the objects were drawn from a probability distribution, in an attempt to mimic measurement noise; as a result, no two grasp trials were performed on identical objects.

Initially, the hand was positioned next to the object top, in a configuration such that a patch of the object surface was within the workspace of each finger. Figure 1 depicts the setup being used in pilot studies involving physical manipulators.

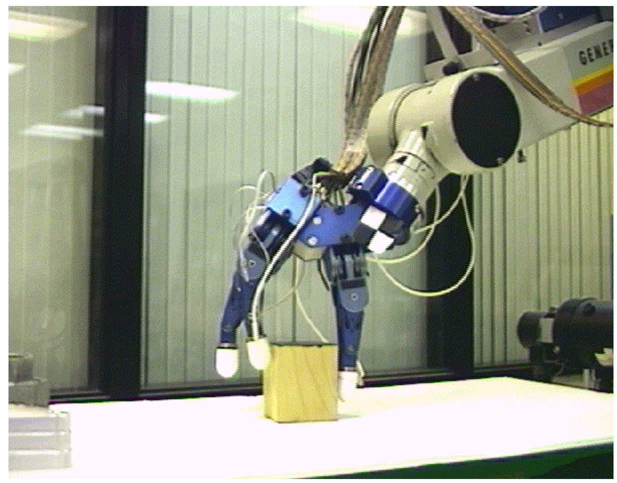

Figure 1: The setup for the grasp task: the hand is positioned so that all fingers can reach the object.

4.2 Sequential Decision Problem

4.2.1 State Representation

The state representation proposed is based on two discrete sets: the control basis $\Pi = \{\pi_1, \pi_2, \ldots, \pi_n\}$, and the membership pattern vector $\mathbf{q} = [p_1 \ \ p_2 \ \ \ldots \ \ p_m]^T$, that indicates which models out a set of m generative models are consistent ($p_k = 1$) with the last observation and which are not ($p_k = 0$).

Given the observation vector $\mathbf{o}$ collected as the agent executes the primitive π_i, the corresponding system state is denoted by the tuple $(\pi_i, \mathbf{q})$. The vector $\mathbf{q}$ expresses which models (or operational contexts) are compatible with the observation $\mathbf{o}$. It conveys more information than individual observations, as it correlates $\mathbf{o}$ with the information derived from the agent's past experiences. The resulting state representation is intrinsically situated, because it is grounded on autonomous interaction with the environment.

Generative Models A generative model describes how observation sequences can be "generated" within a particular operational context. As the agent executes the controller π_i, it can find itself operating under many distinct conditions, characterized by distinct dynamics. If an exhaustive set of generative models is available, the agent may infer which models are compatible with a given sequence of observations by computing and updating the likelihood that each model has generated the data observed. If $P(M_k)$ is the likelihood that model k generates the sequence of observations, and $P(\mathbf{o}|M_k)$ the likelihood of observing $\mathbf{o}$ given that it is drawn from the distribution expressed by model k, then $P(M_k|\mathbf{o})$ (the likelihood that model k explains the data observed given observation $\mathbf{o}$) can be computed as

$$P(M_k|\mathbf{o}) = \frac{P(\mathbf{o}|M_k)P(M_k)}{\sum_{i=1}^{m} P(\mathbf{o}|M_i)P(M_i)},$$

assuming the existence of m generative models.

In this work, the probability $P(\mathbf{o}|M_k)$ is computed by the generative models, and the corresponding conditional probability distributions are represented by parametric models. The cost of acquiring data in real robots favors the choice of parametric models, which derivation typically require less data than for non-parametric ones. The following parameterization was used to represent the generative models:

$$\begin{aligned} \bar{\mathbf{o}} &= [\epsilon \quad -K\epsilon + \epsilon_0]^T \\ M_{\pi_i}(\theta, \mathbf{o}) &= \frac{1}{L}\frac{1}{\sqrt{2\pi\sigma^2}}\exp(\frac{-(\mathbf{o}-\bar{\mathbf{o}})^T(\mathbf{o}-\bar{\mathbf{o}})}{2\sigma^2}) \end{aligned} \quad (3)$$

The observation vector $\mathbf{o} = [\epsilon \quad \dot{\epsilon}]^T$ consists of the squared wrench residual ϵ (defined by Equation 1) and its temporal rate of change $\dot{\epsilon}$. The parameter vector $\theta = [\sigma^2 \quad K \quad \epsilon_0]^T$ contains the parameters of the normal distribution, characterized by the variance σ^2, and the parameters K and ϵ_0 used to compute an estimate of $\dot{\epsilon}$. The constant L normalizes the distribution.

Construction of Generative Models The representation of the dynamics elicited by policy π_i requires a set of m generative models, expressed as $\mathcal{M}(\pi_i) = \bigcup_{k=1}^{m} M_{\pi_i}(\theta_k, \mathbf{o})$. The parameter vector θ_k completely specifies the distribution $M_{\pi_i}(\theta_k, \mathbf{o}) = P(\mathbf{o}|M_k)$.

Each model $M_{\pi_i}(\theta_k, \mathbf{o}) \in \mathcal{M}(\pi_i)$ is constructed using data obtained empirically; as the agents executes the controller π_i, it records a sequence of n observations $\mathcal{O} = \{\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_n\}$ gathered until convergence of π_i. The derivation of model $M_{\pi_i}(\theta_k, \mathbf{o})$ consists in determining the parameter vector θ_k^* that maximizes the likelihood that $M_{\pi_i}(\theta_k, \mathbf{o})$ "generates" $\mathcal{O}$.

The log-likelihood associated with sequence $\mathcal{O}$ is defined as $L(\mathcal{O}, \theta_k) = \sum_{j=1}^{n} \log M_{\pi_i}(\theta_k, \mathbf{o}_j)$, where n is the length of the sequence of observations $\mathcal{O}$, and $\mathbf{o}_j$ denotes the j^{th} element of $\mathcal{O}$. Many optimization procedures can be used to derive the parameter vector θ_k; in the special case in which $M_{\pi_i}(\theta_k, \mathbf{o})$ is a Gaussian distribution with an average that is a linear function of the parameter vector θ_k (as in Equation 3), there exists a closed-form solution for θ_k^*.

In this paper, the models were constructed based on data collected over the course of 35 grasp trials for each family of objects. Considering the four possible choices for π_i (see Equation 2), a total of $4*3*35 = 420$ models resulted; after the elimination of the redundant models, 61 models were isolated.

4.2.2 Reward Structure

The convergent grasp configurations received a score of $1 - \mu_0$, where μ_0 is the minimum friction coefficient required for grasp stability; a score of 1 ($\mu_0 = 0$) is best; no other costs or rewards are present.

4.2.3 Transition Semantics

The system evaluates the utility of pursuing a different control law whenever a change is detected in the membership pattern $\mathbf{q}$. Changes in $\mathbf{q}$ are special events, as they signal the fact that extra information has been acquired by the system.

5 Results

A total of 1600 trials were performed; in each trial an object type was randomly chosen, and an instance of it randomly generated. The initial control law was also chosen at random, determining which set of fingers probed the object. Up to 50 probes are employed in each grasp trial; after 50 probes, the trial is interrupted, and the current grasp configuration scored as if it were a convergent configuration. Exploration was regulated by a Boltzmann distribution.

Figure 2 depicts a typical learning curve (curve labeled △□○, top left curve). The curve was smoothed with a sliding window 10 data points wide. Each point corresponds to the grasp score associated with the convergent grasp configuration. The data corresponding to each object can be separated in three classes, corresponding to the three families of objects. The resulting learning curves are labeled △, □, ○. Because the objects are not perfectly symmetrical, and the control execution is halted before $\dot{\epsilon} = 0$, it is unrealistic to expect an average score of 1. The curves for the individual objects are close to the optimal, within the limitations of the Q-learning algorithm.

Figure 3 presents the performance histograms for the controllers in the control basis ("native" controllers), and for the composite controller after learning. Each histogram depicts data collected over 1600 trials, involving all objects. For the top histogram, a con-

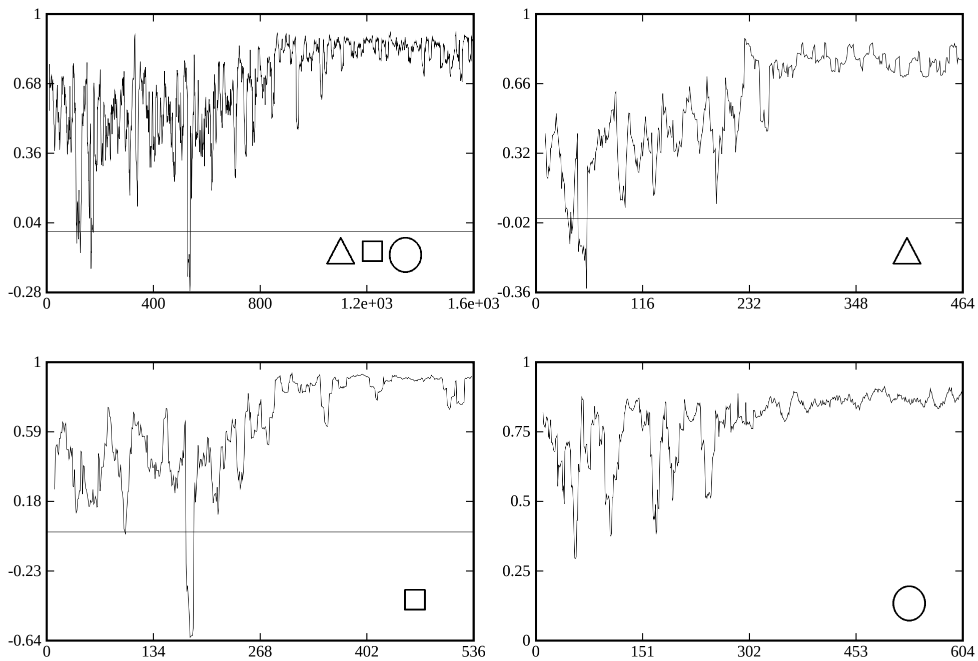

Figure 2: △□○ Typical learning curve for all objects. Other curves (labeled △, □, ○) are the corresponding learning curves for the individual objects. Vertical axes represent grasp scores $(1 - \mu_0)$, and horizontal axes are the trial number.

troller was selected from the control basis at random, and executed until convergence. The end grasp configuration was scored and included in the histogram.

As one can see, the composite controller avoids the majority of low quality solutions: 88% of the solutions have scores higher than 0.8, compared to 59% for the native controllers. The variance associated with solution quality is also substantially smaller.

6 Conclusion and Future Work

All successful organisms exploit some form of native structure (neurological, muscular, skeletal) to settle into a stable dynamic relationship with their environment; by exploiting the intrinsic dynamics of bodies and tasks. Humans learn, in addition, to exploit favorable dynamic relationships to the world by using acquired control knowledge.

In this work, we have described how context-dependent grasp strategies can be constructed through the activaction of the native control primitive most adequate to the perceived operational context. The definition of context or state is itself based on empirically derived dynamic models. The performance gains achieved are the result of a better match between the task requirements and the capabilities of the system.

The method proposed was successfully applied to the multifingered grasp task, in which a robotic hand must deploy its grasp resources (fingers) according to the object identity and orientation, both unknown to the system. In solving this task, the system develops its own notion of "object identity", and uses it effectively. Notice that object classification is not an explicit goal of the system; object categorization is only important insofar as it influences grasp performance.

We plan to implement the algorithm described in the LPR's arm and hand system, consisting of a General Electric P50 robot and a Salisbury hand. The number of trials is an issue, due to the cost of the probing operation. Like other practitioners [7, 10], we plan to store the data acquired in memory and re-use it to speed up learning.

Acknowledgments

This work was supported in part by the National Science Foundation under grants CISE/CDA-9703217, IRI-9704530 and IRI-9503687.

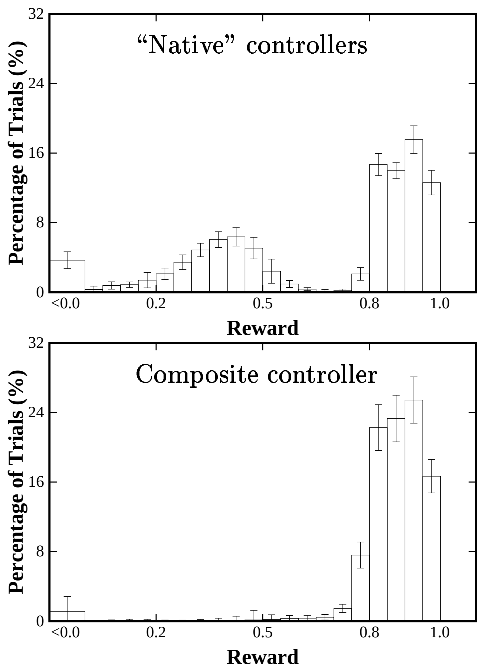

Figure 3: Grasp score distributions for the control basis controllers (“native controllers”) and the composite controller after learning; higher scores are better. Histograms are based on data collected over 100 trials.

References

[1] Aronson, A., and et al. *Clinical Examinations in Neurology.* W.B. Saunders Co., Philadelphia, PA, 1981.

[2] Coelho Jr., J., and Grupen, R. A control basis for learning multifingered grasps. *Journal of Robotic Systems 14*, 7 (1997), 545–557.

[3] Cutkosky, M., and Wright, P. Friction, stability and the design of robotic fingers. *Int. Journal of Robotics Research 5*, 4 (Winter 1986).

[4] Faverjon, B., and Ponce, J. On computing two-finger force-closure graps of curved 2d objects. In *Proc. 1991 IEEE Int. Conf. Robotics Automat.* (Sacramento, CA, May 1991), vol. 1, pp. 424–429.

[5] Ferrari, C., and Canny, J. Planning optimal grasps. In *Proc. 1992 IEEE Int. Conf. Robotics Automat.* (Nice, FRANCE, May 1992), vol. 3, pp. 2290–2295.

[6] Haugsjaa, K., Souccar, K., Connolly, C., and Grupen, R. A computational model for repetitive motion. In *Timing of Behavior: Neural, Computational, and Psychological Perspectives*, C. Collyer and D. Rosebaum, Eds. The MIT Press, Cambridge, MA, 1996.

[7] Lin, L.-J. *Reinforcement Learning for Robots using Neural Networks.* PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, January 1993.

[8] Markenscoff, X., Ni, L., and Papadimitriou, C. The geometry of grasping. *Int. Journal of Robotics Research 9*, 1 (1990), 61–74.

[9] Mason, M., and Salisbury, J. K., Eds. *Robot Hands and the Mechanics of Manipulation.* The MIT Press, Cambridge, MA, 1985.

[10] McCallum, A. K. *Reinforcement Learning with Selective Perception and Hidden State.* PhD thesis, Department of Computer Science, University of Rochester, Rochester, NY, 1996.

[11] Nguyen, V. The synthesis of stable grasps in the plane. In *Proc. 1986 IEEE Int. Conf. Robotics Automat.* (San Francisco, CA, April 1986), vol. 2, IEEE, pp. 884–889.

[12] Piaget, J. *The Origins of Intelligence in Children.* Norton, New York, NY, 1952.

[13] Pratt, J., and Pratt, G. Exploiting natural dynamics in the control of a planar bipedal walking robot. In *Proceedings of the* 36^{th} *Annual Allerton Conference on Communication, Control, and Computing* (1998).

[14] Salisbury, J. K. *Kinematic and Force Analysis of Articulated Hands.* PhD thesis, Department of Mechanical Engineering, Stanford University, Stanford, CA, May 1982.

[15] Taga, G. A model of the neuro-musculo-skeletal system for anticipatory adjustment of human locomotion during obstacle avoidance. *Biological Cybernetics 78*, 2 (1998), 9–17.

[16] Thelen, E. Timing in motor development as emergent process end product. In *The Development of Timing Control and Temporal Organization in Coordinated Action*, J. Fagard and P. Wolff, Eds. Elsevier Science, 1991.

[17] Williamson, M. Neural control of rhythmic arm movements. *Neural Networks 11*, 7-8 (1999), 1379–1394.