Intrinsically Motivated Multimodal Structure Learning

Jay Ming Wong† and Roderic A. Grupen‡

***Abstract*— We present a long-term intrinsically motivated structure learning method for modeling transition dynamics during controlled interactions between a robot and semi-permanent structures in the world. These structures serve as the basis for a number of possible future tasks defined as Markov Decision Processes (MDPs). We apply a structure learning technique to a multimodal affordance representation that yields a population of forward models for use in planning. We evaluate the approach using experiments on a bimanual mobile manipulator (uBot-6) that show the performance of model acquisition as the number of transition actions increases.**

I. INTRODUCTION

This paper adopts an affordance representation that is lightweight (unlike *Object–Action Complexes* (OACs) [1]) and thus, better serves planners that need to roll out a number of these forward models during planning. In fact, only essential Markovian components concerning information regarding states, actions, and transition dynamics, $s, a \mapsto s'$ are encoded, allowing reusability for a large number of task formulated as MDPs. The main contribution of this paper is the presentation of an intrinsically motivated structure learning approach that builds complete action-related representations of objects using multimodal percepts. The resulted is called an *Aspect Transition Graph* (ATG) model. Previous planning architectures using hand built versions of these models have been successful, however, this paper contributes a structure learning approach to acquiring them autonomously.

Approach: We present the first autonomously learned ATG representation with continuously parametrized action edges in the literature. These representations can be used to serve as forward models in belief-space planning infrastructure on real robot systems [2]. A number of studies have integrated ATG affordance representations into the model base as a fundamental attribute in the model-referenced belief-space planning architecture but do not encode fundamental system uncertainties nor inherently encode transition dynamics learned by the robot and therefore are not robust to unexpected outcome [3–5]. ATG encode affordances in a graphical structure defined as a directed multi-graph $G = (\mathcal{S}, \mathcal{A})$ where $\mathcal{S}$ denotes a set of *aspect nodes* connected by action edges $\mathcal{A}$. Sensory information is integrated into the *aspect node*, a state representation defined as a geometric constellation of features derived from multiple sensor modalities. Each parameterized action $a \in \mathcal{A}$ uses a learned search distribution for motor references that reliably transition between aspects. References are defined to be (multivariate) Gaussian distributions $\mathcal{N}(\mu, \Sigma)$ in Cartesian space describing the areas in object frame where the robot has successfully detected a target perceptual reference from this initial state in the past. Each edge in the ATG is a closed-loop controller $\phi|^{\sigma}_{\tau}$ that combines potential functions ($\phi \in \Phi$) with sensory ($\sigma \subseteq \Sigma$) and motor resources ($\tau \subseteq \mathcal{T}$) [5, 6].

Affordances of a given object are determined by exploring actions that cause a transition in aspect space. We select actions to learn affordances either by encouraging coverage through a Latin Hypercube Space (LHS) or by intrinsic motivation. The key insight for using an intrinsic reward like [7] with value iteration is that it encourages the *consumption* of reward through actions, promoting the selection of actions that produce a high differential variance. As these distributions converge, intrinsic reward diminishes, hence encouraging other action parameters contributing to other transitions to be selected. The overall algorithm is as follows,

Algorithm 1 Multimodal Structure Learning

$f \leftarrow$ NIL
do
    $a, \rho \leftarrow$Select params by LHS or $\arg\max_a f(s, a, s')$
    Do a, ρ and obtain experience $\langle s, a, \rho, s' \rangle$
    $r(s, a(\rho), s') \leftarrow \text{abs}(||\Sigma_k||_2 - ||(\Sigma_{k-1})||_2)$
    Update value $f(s, a, s')$ with reward $r(s, a(\rho), s')$
    Update $\mathcal{N}_{s,a\to s'}$ with current action params $a(\rho)$
while $f(s, a, s') > \varepsilon : \forall s, a, s'$

II. EXPERIMENTS

Methodology: Experiments are done on a dynamic simulation of the uBot-6 platform, a 13 DOF, toddler-sized, dynamically balancing, mobile manipulator [8] equipped with an Asus Xtion Pro Live RGB-D camera and two ATI Mini45 Force/Torque sensors one in each hand. Performing a visual observation creates a feature list consisting of maximum likelihood Cartesian features derived from a Kalman filter that summarizes the history of observations to this point in terms of a mean and spatial covariance. Primitive tactile features consist of the contact force $\hat{\boldsymbol{f}} \in \mathbb{R}^3$, from which the sum of squared contact forces between the left (L) and right (R) hands, $\sum_{i=L,R} \boldsymbol{f}_i^T \boldsymbol{f}_i$ and sum of squared contact moments $\sum_{i=L,R} (\boldsymbol{r}_i \times \boldsymbol{f}_i)^T ((\boldsymbol{r}_i \times \boldsymbol{f}_i)$ are computed at the centroid of the pair of contacts measured resulting in bimanual grasp configurations where the squared force and moment residuals

†Planning, Autonomy, and Automation Group, Draper Laboratory, Cambridge, MA, USA jmwong@draper.com

‡Laboratory for Perceptual Robotics, College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA, USA grupen@cs.umass.edu

This material is based upon work supported under Grant NASA-GCT-NNX12AR16A. Any opinions, findings, conclusions, or recommendations expressed in this material are solely those of the authors and do not necessarily reflect the views of the National Aeronautics and Space Administration. For more details, the full manuscript is available on arXiv.

$\lvert\mathcal{A}\rvert$	P-VALUE	PROPOSED	RAND+MEMORIZE
50	0.0034	1.4934 ± 0.0732	1.7551 ± 0.2097
100	0.0063	0.7598 ± 0.1477	1.0919 ± 0.2929
150	0.0068	0.4242 ± 0.1062	0.7666 ± 0.2876
200	0.0125	0.2701 ± 0.0630	0.5797 ± 0.3345
250	0.0071	0.2180 ± 0.0648	0.5474 ± 0.3435
300	0.0153	0.1574 ± 0.0409	0.4626 ± 0.3511
350	0.0217	0.1447 ± 0.0258	0.4388 ± 0.3511
400	0.0254	0.1251 ± 0.0187	0.4197 ± 0.3608
450	0.0251	0.1219 ± 0.0177	0.4071 ± 0.3511
500	0.0323	0.1112 ± 0.0060	0.3865 ± 0.3469

TABLE I: Model error comparison between the proposed structure learning approach and a base-line approach.

are minimized simultaneously. Control actions are executed by the robot to establish new sensor geometries and reveal new aspects causing probabilistic transitions to new aspect nodes. We implement both GRASP and a locomotive ORBIT control program. Since our approach makes no assumption regarding the underlying object and only concerns the aspects that are afforded, it can theoretically be applied to any object, however, we use a simple object geometry whose ATG can be evaluated. In a total of over 250 hours of robot simulation, ARcube objects, 29 cm cubes with a single ARtag on each face, were used.

Results: The first experiment (containing five trials over 150 hours of simulation) compares the proposed approach to a baseline in which the robot randomly explores control parameters, observes the scene, and memorizes its effects in terms of aspect transitions, against a ground truth ORBIT model. Such a method is guaranteed to converge to a complete affordance model given sufficient time and serves as a valid contender for comparisons. Error (in radians) is computed by the absolute difference between the learned model and the ground truth for the means of the distribution along all transition edges. Table I lists the average error for both the proposed and the random memorization approaches after a specific number of actions. In all cases, the proposed method achieves lower errors and in many of these cases, the difference is statistically significant ($p < 0.05$). It is also evident that the proposed approach is capable of acquiring more accurate affordance representations faster and more reliably (with significantly lower standard deviation). The second experiment (consisting of five trials over 100 hours) aims to inspect the result when additional sensor modalities (vision and touch) and actions (ORBIT and GRASP) are introduced, which resulted in slightly slower convergence, yet continues to discover all the transitions in the learned ATG. As the the number of transitions discovered in the model increases, the likelihood of novelty diminishes—this is captured in the decreasing values in the model. Structure learning with the extended action set requires 300–400 actions to produce a complete model, a magnitude similar to other approaches [9].

Conclusion: This manuscript presents an intrinsically motivated structure learning approach to learn semi-permanent Markovian state representations of structures that are reusable in future tasks. The affordance representations learned here serve as forward models in belief-space object identification architectures [2] by predicting how state distributions change in response to interaction. Despite success in the past using hand-crafted models of this type, the methods presented in this paper allows us to acquire them autonomously and encodes parameters robust to robot uncertainties derived from the properties of the system and its interaction with the word that would otherwise be difficult to precisely hand define. Structure learning allows robots to build models themselves without supervision and promotes informed action selection, exploiting known structure and promoting a sense of discovery. Results demonstrate the acquisition of models that are significantly better than approaches that solely select random actions to learn from. We believe that autonomously learning affordance representations as forward models with more complex actions and modalities allows for a richer set of future solvable tasks and perhaps reduces the complexity in model-referenced planning, thus reducing planning time and rollouts necessary.

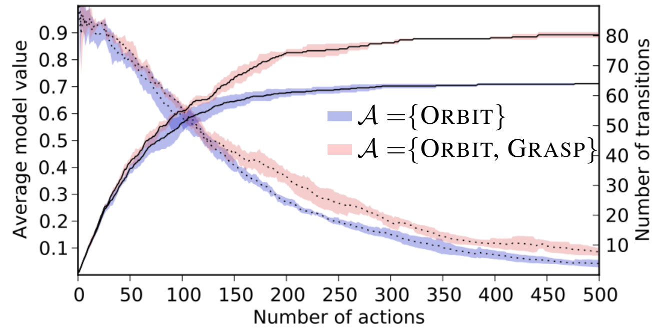

Fig. 1: Average Q value and the number of transitions discovered in the model using the proposed approach. The dotted lines correspond to average model value and solid lines describe the number of transitions in the affordance model (Best viewed in color).

REFERENCES

[1] N. Kruger, C. Geib, J. Piater, R. Petrick, M. Steedman, F. Worgotter, A. Ude, T. Asfour, D. Kraft, D. Omrcen, A. Agostini, and R. Dillmann, "Objectaction complexes: Grounded abstractions of sensorymotor processes," *Robotics and Autonomous Systems*, vol. 59, no. 10, pp. 740 – 757, 2011.

[2] R. A. Grupen, M. Hebert, M. W. Lanighan, T. Liu, D. Ruiken, T. Takahashi, and J. M. Wong, "Affordance-based active belief recognition using visual and manual actions," in *IEEE/RSJ the International Conference on Intelligent Robots and Systems (IROS)*, Oct 2016.

[3] L. Y. Ku, S. Sen, E. Learned-Miller, and R. Grupen, "Action-based models for belief-space planning," in *Workshop on Information-Based Grasp and Manipulation Planning, at Robotics: Science and Systems*, Berkeley, California, July 2014.

[4] ——, "The aspect transition graph: An affordance-based model," in *Second Workshop on Affordances: Visual Perception of Affordances and Functional Visual Primitives for Scene Analysis, at the European Conference on Computer Vision*, Zurich, Switzerland, Sept 2014.

[5] S. Sen and R. Grupen, "Integrating task level planning with stochastic control," University of Massachusetts Amherst, Tech. Rep. UM-CS-2014-005, 2014.

[6] M. Huber, W. S. MacDonald, and R. A. Grupen, "A control basis for multilegged walking," in *IEEE International Conference on Robotics and Automation*, Apr 1996.

[7] S. Hart, "An intrinsic reward for affordance exploration," in *IEEE International Conference on Development and Learning*, June 2009.

[8] D. Ruiken, M. W. Lanighan, and R. A. Grupen, "Postural modes and control for dexterous mobile manipulation: the umass ubot concept," in *IEEE-RAS International Conference on Humanoid Robots*, Oct 2013.

[9] E. Ugur and J. Piater, "Emergent structuring of interdependent affordance learning tasks," in *Joint IEEE International Conferences on Development and Learning and Epigenetic Robotics*, Oct 2014.