URoboSim — a simulation-based predictive modelling engine for cognition-enabled robot manipulation

Michael Neumann and Andrei Haidu and Michael Beetz
University Bremen
neumann.michael@uni-bremen.de, haidu@uni-bremen.de, beetz@cs.uni-bremen.de

Abstract

In a nutshell robot simulators are fully developed software systems that provide simulations as a substitute for real-world activity. They are primarily used for training modules of robot control programs, which are, after completing the learning process, deployed in real-world robots. In contrast, simulation in (artificial) cognitive systems is a core cognitive capability, which is assumed to provide a "small-scale model of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilise the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it." [8] This means that simulation can be considered as an embodied, online ***predictive modelling engine*** *that enables robots to contextualize vague task requests such as "bring me the milk" into a concrete body motion that achieves the implicit goal and avoids unwanted side effects. In this setting a robot can run small-scale simulation and rendering processes for different reasoning tasks all the time and can continually compare simulation results with reality — it is a promising Sim2Real2Sim setup that has the potential to create much more powerful robot simulation engines. We introduce* ***URoboSim****, a robot simulation framework that is currently designed and developed with this vision in mind.*

1. Simulation in engineering and cognition

In recent years the disruptive progress of virtual reality simulation and rendering tools has been leveraged for realizing several high-performance robot simulators and simulation environments [24, 2, 19, 6, 25, 26, 27, 22, 20, 17]. These simulators have been used, as substitutes for the real world, to train and test the skills of robot agents [17, 1, 7, 3]. The learned and tested skills are then transferred to the real robots.

The view of robot simulators as real-world substitutes is very limited compared to the role that simulation and prospection takes for cognition-enabled agency. Williams [23] has proposed a framework that views the brain as a probabilistic *prospective modelling engine* that can abstractly project plans and alternatives, concretely simulate the agent environment interactions, predict future situations and how they look. Schacter et al.[21] define prospection as the ability to represent what might happen in the future and propose a taxonomy of prospective capabilities distinguishing their function as well as their level of abstraction. Another example is Hesslow's simulation theory of cognition[12, 13], which proposes that thinking is simulated interaction with the environment. The role of building, maintaining, and using models that predict and explain what is happening and include causal and intuitive physics knowledge has also been stressed in the perspective article "Building machines that learn and think like people"[16].

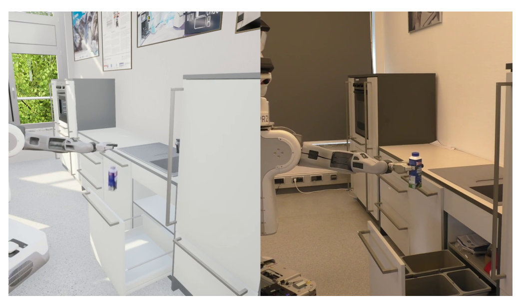

Figure 1. Belief-state (left) during action of real robot (right).

2. URoboSim

URoboSim is a simulation framework that is the work horse of the cognitive architecture CRAM (Cognitive Robot Abstract Machine) [5]. CRAM with the help of URoboSim enables robot agents to accomplish underdetermined task requests that require accomplishing fetch and place tasks and other simple manipulation tasks such as pouring, wiping. In each cycle of the perception-action loop the simulation environment is updated by physically simulating the body motions and their physical effects generated in the respective cycle and correcting the current scene in URoboSim to better match the captured camera image. Figure 1 shows how URoboSim is emulating the execution of a robot plan in order to compute the belief state of the robot. URoboSim models robots using the data structures and service libraries of the open-source robot middleware ROS, which means that is functionally equivalent to the robot simulation framework Gazebo. URoboSim includes models of different robots and a kitchen laboratory environment and a small drugstore. While the kitchen environment is handmade, the drugstore environment can be created by autonomous robot mapping given a grammar of how shelf system are configured and realistic models of all products in the store.

3. Characteristics of URoboSim

Because simulation in URoboSim is a cognitive mechanism rather than a substitute for the real world, URoboSim has several unique characteristics:

Characteristic 1: The scene graph that URoboSim is a rendering of a symbolic knowledge base. All entities and their parts that are depicted in a visualization of a URoboSim simulation have a symbolic name and are axiomatized in the ontology of the symbolic knowledge base, which makes the URoboSim scenes machine-interpretable as depicted in 2 [4, 11]. This means that one can ask open queries on scenes such as all container objects that have a volume larger than 0.5m^3 and have a horizontal handle, or highlight the handle of the container that contains the open milk carton and display the articulation model of the container door. This functionality is important for the generation of semantically segmented and annotated learning data as well as automated skill testing.

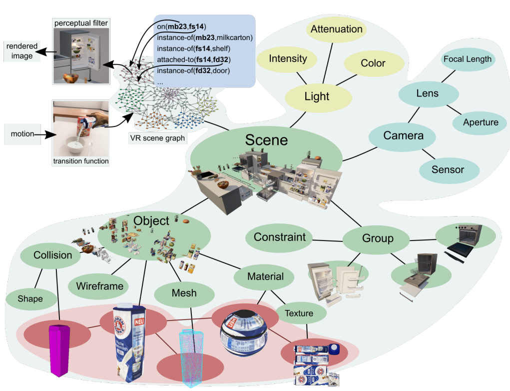

Figure 2. Casting of the URoboSim scene graph as a virtual symbolic knowledge base.

Characteristic 2: URoboSim can emulate the ongoing activities of a robot agent. URoboSim cannot only be operated as a stand-alone robot simulator but also as an integrated component of the robot control system that can emulate ongoing activities [18, 15]. In this mode the robot maintains a scene graph as a belief state with respect to the current state of the robot operating environment. To this end, the robot can render the expected camera image, as depicted in 3, given its belief state and then adapting its belief state in order to minimize the rendered and the real camera image. In addition, the robot simulates its body motion in its belief state. This way it can infer that a bowl falls to the floor when it opens its gripper when holding it. Note that action emulation forms a Sim2Real2Sim loop where deviations of predicted and perceived states can be used to improve the simulation capabilities.[1]

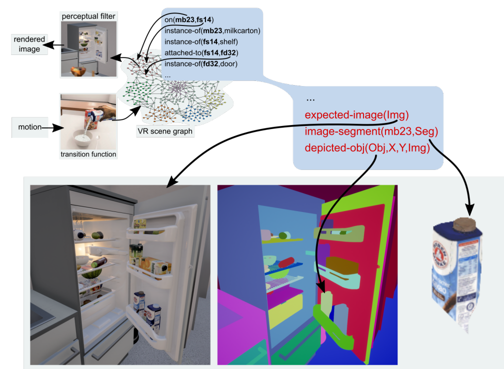

Figure 3. Rendering of the expected image based on the current belief state.

Characteristic 3: URoboSim automatically perceives events in simulated episodes and represents robot manipulation episodes in first-order time interval logic according to the Flanagan action model. URoboSim (1) detects force-dynamic events such as hand making contact to the object to be picked up and object breaking contact with supporting surface, (2) parses the events into hierarchical action models with motion phases such as reaching, grasping, and lifting, and (3) recognizes, structures, and semantically annotates actions such as picking up and placing objects. Event perception and event cognitionare key steps for open question answering about simulated manipulation episodes [10, 9].

4. Pointers to more information

A video showcasing the capabilities of URoboSim can be found at http://ease-crc.org/link/video-urobosim. URoboSim can simulate real robot experiments described in [14]. Semantically structured and annotated log files of URoboSim simulations of autonomous robot table setting episodes can be retrieved through the open web-based knowledge service openEASE https://data.open-ease.org/QA?neem_id=603127322113d53026863697.

[1] This is future research.

References

[1] Target-driven visual navigation in indoor scenes using deep reinforcement learning. *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2016. 1

[2] AI2-THOR: An Interactive 3D Environment for Visual AI. *ArXiv*, abs/1712.05474, 2019. 1

[3] Iro Armeni, Sasha Sax, Amir R Zamir, and Silvio Savarese. Joint 2d-3d-semantic data for indoor scene understanding. *arXiv preprint arXiv:1702.01105*, 2017. 1

[4] Michael Beetz, Daniel Beßler, Andrei Haidu, Mihai Pomarlan, Asil Kaan Bozcuoglu, and Georg Bartels. Knowrob 2.0 – a 2nd generation knowledge processing framework for cognition-enabled robotic agents. In *International Conference on Robotics and Automation (ICRA)*, Brisbane, Australia, 2018. 2

[5] M. Beetz, L. Mösenlechner, and M. Tenorth. Cram — a cognitive robot abstract machine for everyday manipulation in human environments. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1012–1017, 2010. 1

[6] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym. *arXiv preprint arXiv:1606.01540*, 2016. 1

[7] Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. Matterport3d: Learning from rgb-d data in indoor environments. *arXiv preprint arXiv:1709.06158*, 2017. 1

[8] Kenneth James Williams Craik. *The Nature of Explanation*. Cambridge: Cambridge University Press, 1943. 1

[9] Andrei Haidu and Michael Beetz. Automated models of human everyday activity based on game and virtual reality technology. In *International Conference on Robotics and Automation (ICRA)*, Montreal, Canada, 2019. 2

[10] Andrei Haidu and Michael Beetz. Automated acquisition of structured, semantic models of manipulation activities from human vr demonstration. In *IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China, 2021. Accepted for IROS 2021. 2

[11] Andrei Haidu, Daniel Beßler, Asil Kaan Bozcuoglu, and Michael Beetz. Knowrob-sim — game engine-enabled knowledge processing for cognition-enabled robot control. In *International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain, 2018. IEEE. 2

[12] Germund Hesslow. Conscious thought as simulation of behaviour and perception. *Trends in Cognitive Sciences*, 6(6):242–247, 2002. 1

[13] Germund Hesslow. The current status of the simulation theory of cognition. *Brain research*, 1428:71–9, 06 2011. 1

[14] Gayane Kazhoyan, Simon Stelter, Franklin Kenghagho Kenfack, Sebastian Koralewski, and Michael Beetz. The robot household marathon experiment. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2021. Submitted for review. 2

[15] Franklin Kenghagho Kenfack, F. A. Siddiky, Ferenc Bálint-Benczédi, and M. Beetz. Robotvqa — a scene-graph- and deep-learning-based visual question answering system for robot manipulation. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9667–9674, 2020. 2

[16] Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum, and Samuel J Gershman. Building machines that learn and think like people. *Behavioral and brain sciences*, 40, 2017. 1

[17] Chengshu Li, Fei Xia, Roberto Martín-Martín, and Silvio Savarese. Hrl4in: Hierarchical reinforcement learning for interactive navigation with mobile manipulators. In *Conference on Robot Learning*, pages 603–616. PMLR, 2020. 1

[18] Patrick Mania, Franklin Kenghagho Kenfack, Michael Neumann, and Michael Beetz. Imagination-enabled robot perception, 2020. 2

[19] NVIDIA. NVIDIA Isaac. `https://developer.nvidia.com/isaac-sdk`, 2021. [Online; accessed 12-May-2021]. 1

[20] Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, et al. Habitat: A platform for embodied ai research. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9339–9347, 2019. 1

[21] Karl K Szpunar, R Nathan Spreng, and Daniel L Schacter. A taxonomy of prospection: Introducing an organizational framework for future-oriented cognition. *Proceedings of the National Academy of Sciences*, 111(52):18414–18421, 2014. 1

[22] Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5026–5033. IEEE, 2012. 1

[23] Daniel George Williams. *The Mind as a Predictive Modelling Engine: Generative Models, Structural Similarity, and Mental Representation*. PhD thesis, University of Cambridge, 2018. 1

[24] Fei Xia, William B. Shen, Chengshu Li, Priya Kasimbeg, Micael Edmond Tchapmi, Alexander Toshev, Roberto Martín-Martín, and Silvio Savarese. Interactive gibson benchmark: A benchmark for interactive navigation in cluttered environments. *IEEE Robotics and Automation Letters*, 5(2):713–720, 2020. 1

[25] Fei Xia, Amir R Zamir, Zhiyang He, Alexander Sax, Jitendra Malik, and Silvio Savarese. Gibson env: Real-world perception for embodied agents. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1

[26] Fanbo Xiang, Yuzhe Qin, Kaichun Mo, Yikuan Xia, Hao Zhu, Fangchen Liu, Minghua Liu, Hanxiao Jiang, Yifu Yuan, He Wang, et al. Sapien: A simulated part-based interactive environment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11097–11107, 2020. 1

[27] Xu Xie et al. Vrgym: A virtual testbed for physical and interactive ai. *arXiv preprint arXiv:1904.01698*, 2019. 1