Imagine a human! UAivatar - Simulation Framework for Human-Robot Interaction

Mona Abdel-Keream[1] and Michael Beetz[1]

Abstract— **Designing reliable, effective, and safe robotics systems for Human-Robot Interaction (HRI) remains still a major challenge due to the non-symmetrical nature of HRI and the absence of a comprehensive human model. For a meaningful interaction is suggested that both participants must have knowledge of one another. To achieve this an adequate representation of the other's belief, intentions, goals, actions, etc. is necessary. We introduce in this paper the UAivatar simulation framework that allows for the simulation of various HRI scenario. It includes a cognitive control for a human character and a knowledge-based human model that is integrated into the AI software and control architecture of the operating robot. Our framework, therefore, extends the robots inner world model with rich knowledge about humans allowing the robot to 'imagine' and simulate humans performing everyday activities.**

I. Background and Related Work

Robots are steadily becoming a fundamental part of our daily life. They have the potential to take over everyday activities at home [1], office [2], health-care [3] or public places [4]. Since robots will work in close proximity to human beings interaction, their interaction is inevitable. Human-Robot Interaction (HRI) is an interdisciplinary field that studies the dynamic interaction between human beings and robots.

Simulated environments are great tools to design, test, and verify such agent models and rightfully it has been explored in the robotics field for more than 30 years [5]. In robotics, the typical usages of simulated environments ranges from design, development, and validation of robotic systems. Although, the applications can be more straightforward as to teach students about robots or more demanding/challenging where robots are trained to perform everyday human activities.

Naturally, there have been many publications centered around the development of robotics in which the robot ends up treated as an isolated agent and very few works have been done where a human model is provided as an active agent that can take part in the robot's world. For example, NVIDIA Isaac Sim [6] is a very powerful tool for robot development that can provide randomized human-like avatars that share the space with the robots. However,it does not support any kind of active interaction with the robots and they are merely used for testing collision avoidance algorithms. On the other hand, there is VirtualHome [7], a very advanced project with programmable humanoid avatars that can not only move around a house but also interact with the world and make changes to it. However, the main focus of this work is acquiring human knowledge and so they do not provide any infrastructure for the integration of robots. Another simulator for studying human-robot interactions is USARSim [8]. This framework, however, examine HRI from an operational level (commander-executor) rather from an collaborative level (peer-to-peer) and without even including an avatar in their simulation environment. On the other hand,[9], a simulation platform for social human-robot interaction offers control for a human agent either through VR or keyboard input. However, the human agent here is not introduced to its robotic partner as a further cognitive agent but instead as a social and physical interactive partner. Finally, there is MORSE [10], a robust simulator for academic robotics fairly used from 2012 [11] to 2014 [12] on HRI simulation, some of its applications introduce the use of knowledge base systems, however, they are not meant for robot-human awareness. Also, MORSE is an already outdated tool as they used the Blender Game Engine, which was discontinued a couple of years ago.

Simulation environments for the development of robots still miss a comprehensive human model. Existing human models are often very simple and do not provide any insights into the cognitive process of the human therefore realizing the autonomous behavior of the simulated human character is very limited. Hiatt et al. [13] review different approaches to model human behavior for human-robot collaboration tasks. According to their research on a computational level, a simple probabilistic model can capture human behavior by defining cost functions [14]. However, such models are a generalized characterization of human behavior and do not provide any deep insight into the underlying processes that influence the behavior of humans.

With machine learning, [15] more complicated mathematical models can be formulated to capture a broader array of human behavior but still without answering why humans behave the way they do. While computational models make sense of low-level signals like movement patterns to anticipate better or recognize human behavior, Petri Nets [16], for instance, analyze human behavior on a higher level. Petri Nets describe the behavior of humans as a discrete event system and analyze how system resources, e.g. people or objects, change over time. Similar to computational models, they do not provide an answer to the underlying reasons for human behavior. A more powerful approach is knowledge-

*This work was supported by the d by Deutsche Forschungsgemeinschaft (DFG) collaborative research center 1320 "Everyday Activity Science and Engineering (EASE)" and by the DFG project AI4HRI.

[1] The author is with the Institute for Artificial Intelligence (IAI), Universitaet Bremen, 28359 Bremen, Germany {abdelker,beetz}@cs.uni-bremen.de

based human models that provide a richer and hierarchical-based understanding of human behavior.

We suggest that for a robot to have an understanding of a human and its complexity, humans must be introduced as further cognitive agents to its robotic partner. Thus,our proposed framework **UAivatar**[1] includes, therefore, a fully controllable and programmable human model. The human model is integrated into the AI software and control architecture of the operating robot. The robot software architecture consists of the high-level planning framework CRAM, and the knowledge processing system KNOWROB2 [17] with **UAivatar** human modeling and simulation tools,allowing the reader to design and test various operational aspects in HRI. In addition, it provides a digital twin model of a human partner.

Our human agent model is described in the same way as our robot PR2 using the Semantic Robot Description Language (SRDL)[18] which is tightly integrated into KnowRob. KNOWROB2 [17] is a knowledge representation and reasoning system that equips robots with a query answering machine that runs within the perception-action loop. With the introduction of SRDL, early efforts were made to model self-aware robots. SRDL promotes self-awareness by equipping robots with knowledge about themselves. It links the components such as sensors, actuators, and control programs via capabilities to actions in the form of an ontology. This approach allows for bridging the gap between high-level actions and the low-level description of the robot such as the structure and kinematics of its manipulator. In this way a robot can infer the required components for performing a specific action.

For modeling human behavior we choose the high-level planning framework CRAM(Cognitive Robot Abstract Machine) [19] which is the same planning framework running on the operating robot. CRAM is flexible and abstract enough that it can be used across many different kinds of robotic agents; including human models. The human model is therefore integrated into the system as a robotic agent with its own CRAM high-level plans, action designators, and process modules. Since CRAM provides tools that enable the agents to reason about their actions as well as other agent's actions integrating a robotic human model provides the robotic agent insight into the human's intention.

Our framework equips not only robots with symbolic knowledge about humans but makes use of a Digital Twin World that provides mechanisms for accessing the full joint configuration of our agent models which allows us to simulate HRI on a highly sophisticated level. With URoboSim [20] the robot is now capable to utilize its mental simulation to anticipate action and action effects. The framework offers many features to aid the robot in making informed decisions about its actions. In addition to a symbolic knowledge base, it utilizes modern information processing technologies such as physics simulation and rendering mechanisms of game engines to generate an inner world knowledge base. This inner world model is a detailed, and photo-realistic reconstruction of the robot's environment. The robot can, therefore, geometrically reason about a scene by virtually looking at it using the vision capability provided by the game engine, and predict the effects of actions through semantic annotations of force dynamic events monitored in its physics simulation. Being in a virtual environment we have access to ground truth data during the whole simulation.

[1]https://github.com/code-iai/UAivatar

II. Framework Architecture

The framework includes a fully controllable human model integrated in the AI software and control architecture of the operating robot. Figure 1 depicts different parts of the architecture which are explained in the following paragraphs.

The **UAivatar** is developed as a plugin for the Unreal Engine (UE). It includes a digital 3D UE MetaHuman that we refer to as *Avatar*. Each Avatar is a photorealistic rigged model with a movable and posable digital skeleton mesh. The plugin provides a ROS low-level control interface to command the human agent in the virtual environment. Additional ROS interfaces are implemented to communicate with other components of the system via ROS topics, services and actions.

We model four different aspects of our **UAivatar** human agent:

- Kinematics
  - Human URDF Model
- Semantics
  - KnowRob Agent Model
- Control
  - CRAM High-Level Plans
- Memory
  - KnowRob Narrative Episodic-Enabled Memories (NEEMs)

Kinematics: To model the kinematics of humans, a basic human model is first described through the Unified Robot Description Format (URDF). The URDF model specifies the kinematic structure of agents through a series of links that are connected by kinematic joints. For the visual appearance, the 3D digital human mesh was segmented and cut into different parts to obtain the individual mesh for each link. The UAivatar framework includes a tf publisher to keep track of all avatar link frames and their transformations between them and the environment during simulation. It includes also a joint state publisher as well as a subscriber to command the avatar with other ROS compatible softwares.

Semantics: Since however, the URDF does not provide any semantics to explain what the kinematic structure actually means, it is transformed to a more powerful and descriptive SRDL model. On the semantic level the kinematics links of an agent are grouped to more high-level components that can be associated to a dedicated purpose. The URDF model of the human agent describes, for instance, hand solely as a base link for several other links without any concrete meaning, context or purpose. In the SRDL model, however,

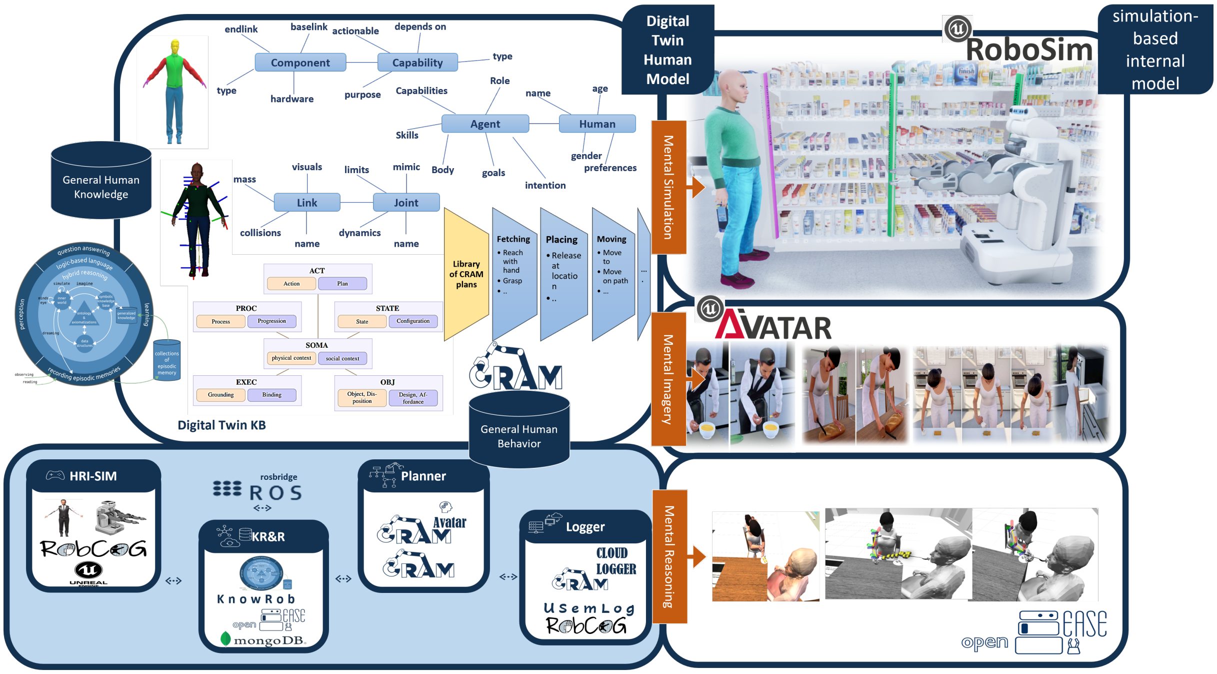

Fig. 1. An architecture for cognition-enabled robotics using a simulation-based inner world model of a human avatar. The human model is integrated in the AI software and control architecture of the operating robot.

hand is modeled as an entity with relations to other entities. Hands are now recognized as part of the human body with wrists as their kinematic root. The purpose of hands and the remaining high-level components is determined through modeling capabilities.

Control: In order to define the high-Level plans, we prefer to use CRAM as it is a well-proven toolbox for designing, implementing and deploying software on autonomous robots. The CRAM Plan library of the avatar includes various high-level plans that have been executed in the digital twin world. The plans are defined through CRAM's action designators which are resolved into series of more specific sub-Actions or into several motion designators, creating a hierarchical plan structure. Different agents can have different kinds of actuators and the process modules allow for this flexibility, providing an agent-independent interface to the high-level planning. This last interface relies on ROS to call for the services provided by the Avatar in UE. Both the Avatar and the Robot can work with CRAM and ROS. The difference is that the atomic action controllers of the Avatar are implemented within UE while the motion controllers of the Robot are implemented with Giskard [2]. Controlling the human model with the functionality provided by UE also allows for commanding the Avatar from simpler ROS-Python scripts or even from a console command prompt within the UE's ViewPort, which results in efficient debugging. The advantage of also using CRAM with the Avatar lies in having clear action descriptions that can be easily understood and reason about by the Robot. The current UAivatar can be commanded to:

- grasp objects
- place objects
- cut objects
- spoon liquids
- pour liquids
- fork food
- feed food
- open/close doors
- follow paths
- interact with microwave/fridge
- turn book/newspaper pages
- sit on chair

Memory: During the execution of the CRAM high-level plan, event-related knowledge is automatically queried and logged. This knowledge include the action hierarchy of the performed task, parameters such as the success status for all actions and their sub-actions, and as well as the status of all task-related objects. Using our digital twin we can also extract knowledge directly from UE and log it into our knowledge base KnowRob as Narrative Episodic-Enabled Memories (NEEMs) . This knowledge comprises two parts: symbolic and sub-symbolic knowledge. Symbolic knowledge represents all the events occurring during the environment and is described using the Web Ontology Language (OWL). The sub-symbolic knowledge includes the position and orientation of all task-related objects and human body parts and the trajectories of all motions, which were performed during the Task Sequence.

III. CONCLUSIONS AND FUTURE WORK

This work presents a fully controllable human model integrated into the AI software and control architecture of the op-

[2]https://giskard.de/wiki:tutorials:introduction

erating robot. The robot's knowledge base was extended with a cognitive human model that describes different aspects of human characteristics, such as kinematics, semantics, control and memory. Moreover, NEEMs were generated and logged into the knowledge base using CRAM for commanding an Avatar within a digital twin world. These NEEMs can be queried by any robotic system, with access to the knowledge base and with CRAM as its cognitive reasoner, to construct models of human behaviors; gaining common ground access to valuable information about human activities and so having better means for exploiting its human awareness capabilities. To improve robot-human awareness, we plan to enrich our cognitive human model with the knowledge that goes beyond the kinematics and the capabilities of humans. We believe the information state of humans should also capture preferences regarding HRI as well as human emotions that might require an emotional response from a robotic agent. Further research will be conducted in the context of the theory of mind and perspective taking where **UAivatar** can elevate the process of estimating human intent.

APPENDIX

ACKNOWLEDGMENT

REFERENCES

[1] S. Miller, J. van den Berg, M. Fritz, T. Darrell, K. Goldberg, and P. Abbeel, "A geometric approach to robotic laundry folding," *International Journal of Robotic Research - IJRR*, vol. 31, pp. 249–267, 02 2012.

[2] M. Veloso, J. Biswas, B. Coltin, and S. Rosenthal, "Cobots: Robust symbiotic autonomous mobile service robots," in *Proceedings of the 24th International Conference on Artificial Intelligence*, ser. IJCAI'15. AAAI Press, 2015, p. 4423–4429.

[3] T. Dahl and M. Kamel Boulos, "Robots in health and social care: A complementary technology to home care and telehealthcare?" *Robotics (MDPI, ISSN 2218-6581)*, vol. 3, pp. 1–21, 03 2014.

[4] M. Cirillo, L. Karlsson, and A. Saffiotti, "Human-aware task planning: An application to mobile robots," *ACM Trans. Intell. Syst. Technol.*, vol. 1, no. 2, Dec. 2010. [Online]. Available: https://doi.org/10.1145/1869397.1869404

[5] W. Nelson and J. Chang, "Simulation of a Cartesian Robot Arm," in *IEEE International Conference on Robotics and Automation*, Atlanta, GA, USA, Mar. 1984.

[6] N. Corporation. (2020) Nvidia isaac sim. [Online]. Available: https://developer.nvidia.com/isaac-sim

[7] X. Puig, K. Ra, M. Boben, J. Li, T. Wang, S. Fidler, and A. Torralba, "Virtualhome: Simulating household activities via programs," in *CVPR*, 2018.

[8] M. Lewis, J. Wang, and S. Hughes, "Usarsim: Simulation for the study of human-robot interaction," *Journal of Cognitive Engineering and Decision Making*, vol. 1, pp. 98–120, 03 2007.

[9] T. Inamura and Y. Mizuchi, "Sigverse: A cloud-based vr platform for research on multimodal human-robot interaction," *Frontiers in Robotics and AI*, vol. 8, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frobt.2021.549360

[10] S. Lemaignan, M. Hanheide, M. Karg, H. Khambhaita, L. Kunze, F. Lier, I. Lütkebohle, and G. Milliez, "Simulation and hri recent perspectives with the morse simulator," in *Simulation, Modeling, and Programming for Autonomous Robots*, D. Brugali, J. F. Broenink, T. Kroeger, and B. A. MacDonald, Eds. Cham: Springer International Publishing, 2014, pp. 13–24.

[11] S. Lemaignan, G. Echeverría, M. Karg, J. Mainprice, A. Kirsch, and R. Alami, "Human-robot interaction in the morse simulator," in *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 2012, pp. 181–182.

[12] G. Milliez, E. Ferreira, M. Fiore, R. Alami, and F. Lefè vre, "Simulating human-robot interactions for dialogue strategy learning," in *Simulation, Modeling, and Programming for Autonomous Robots*, D. Brugali, J. F. Broenink, T. Kroeger, and B. A. MacDonald, Eds. Cham: Springer International Publishing, 2014, pp. 62–73.

[13] L. M. Hiatt, C. Narber, E. Bekele, S. S. Khemlani, and J. G. Trafton, "Human modeling for human–robot collaboration," *The International Journal of Robotics Research*, vol. 36, no. 5-7, pp. 580–596, 2017. [Online]. Available: https://doi.org/10.1177/0278364917690592

[14] A. Dragan and S. Srinivasa, "Formalizing assistive teleoperation," 07 2012.

[15] C. Breazeal, J. Gray, and M. Berlin, "An embodied cognition approach to mindreading skills for socially intelligent robots," *The International Journal of Robotics Research*, vol. 28, pp. 656–, 05 2009.

[16] C. Dondrup, I. Papaioannou, and O. Lemon, "Petri net machines for human-agent interaction," 09 2019.

[17] M. Beetz, D. Beßler, A. Haidu, M. Pomarlan, A. K. Bozcuoglu, and G. 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.

[18] L. Kunze, T. Roehm, and M. Beetz, "Towards semantic robot description languages," *2011 IEEE International Conference on Robotics and Automation*, pp. 5589–5595, 2011.

[19] M. Beetz, L. Mösenlechner, M. Tenorth, and T. Rühr, "Cram – a cognitive robot abstract machine," in *5th International Conference on Cognitive Systems (CogSys 2012)*, 2012.

[20] M. Neumann, A. Haidu, and M. Beetz, "Urobosim — a simulation-based predictive modelling engine for cognition-enabled robot manipulation," 2021.