Cognitive architecture for robot perception and learning based on human-robot interaction

Uriel Martinez-Hernandez[1], Luke W. Boorman[1], Andreas Damianou[2], and Tony J. Prescott[1]

***Abstract*— In this paper, we present a cognitive architecture for human-robot interaction to allow robots to perceive, learn and understand the changing state of their surrounding environment. The cognitive architecture is composed of modules for processing of multiple sensory inputs, perception, memory and control of actions. The human-robot interaction is based on the recognition and learning of human faces and actions. Also, a verbal communication between the human and robot is established to control the interaction. Our cognitive architecture has been integrated and tested with the iCub humanoid robot allowing it to recognise and interact with humans, and to understand and learn from their actions.**

I. Introduction

Interaction and observation are essential capabilities in robotics to endow with intelligent systems able to perceive and learn from their changing surrounding environment. Several systems have used the learning by observation approach in passive mode, such as learning assembly tasks and games [1],[2]. This approach does not take advantage of the interaction with humans to improve the perception and learning processes in robots. Research on human-robot interaction, perception and learning has shown significant progress, however, development of integrated approaches to understanding user and robot expectations, intentions and actions are still in their infancy [3]. This is related to the increasing complexity in the design and capabilities of today's robots.

We designed and developed a cognitive architecture for robust perception and learning based on observation and human-robot interaction. Our cognitive architecture, composed of various modules, allows robots to learn with a biologically inspired memory module, which receives inputs from multiple sensory modalities and provides outputs for behaviour control of the robot platform. Thus, our cognitive architecture provides a suitable platform for robot learning based on observation and human-robot interaction. Figure 1 shows an example of the human-robot interaction process where the iCub humanoid robot observes and learns from actions performed by the human.

*This work was supported by EU Framework project WYSIWYD (FP7-ICT-2013-10).

[1]Uriel Martinez-Hernandez, Luke W. Boorman and Tony J. Prescott are with the Sheffield Robotics Laboratory and the Department of Psychology, The University of Sheffield, Sheffield, U.K. uriel.martinez, l.boorman, t.j.prescott@sheffield.ac.uk

[2]Andreas Damianou is with the Department of Computer Sciences, The University of Sheffield, Sheffield, U.K. andreas.damianou@sheffield.ac.uk

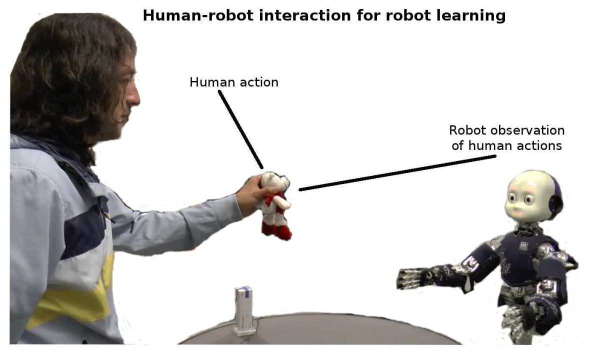

Fig. 1. Robot learning of actions performed by the human. The iCub humanoid detects, recognises and tracks the human face whilst observing his actions. Human faces and actions are learned by our cognitive architecture.

II. Methods

A. Robotic platform

We used the iCub humanoid robot for integration of our cognitive architecture and interaction with humans. The iCub humanoid is a highly complex and open source robot platform composed of 53 DoF, multiple sensory modalities and designed for research in cognitive development [4]. Multiple sensory input and control of gaze and neck are used for the human-robot interaction. Visual input allows the robot to detect, recognise and track human face and arms. Audio input allows to establish a verbal communication between the human and robot. Robust control of gaze and neck of the robot platform provides a more natural human-robot interaction.

B. Cognitive architecture

The core of our cognitive architecture is based on a Synthetic Autobiographical Memory model (SAM) inspired by hippocampal memory models for navigations previoulsy investigated [5]. The SAM model receives inputs from multiple sensory modalities, i.e., vision, audio, touch, proprioception, in an appropriate format. The data is processed with a Gaussian Process approach using the GPy library [6]. Then, the most probable output is provided for the current situation, for instance, face prediction, action recognition and prediction of human movements. Figure 2 shows the different modules that compose the SAM model.

1) Human face detection: To start a human-robot interaction, the robot needs to detect the human in front of it. For this reason, integration of face recognition and tracking modules in the robot is required. We have implemented this process using the SAM model as the core for learning and

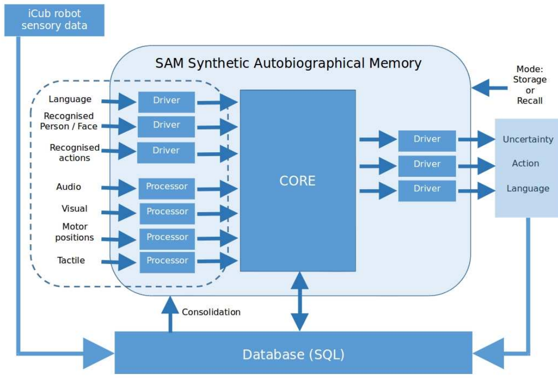

Fig. 2. Cognitive architecture for learning by human-robot interaction. The core of the architecture is based on the SAM model, which receives inputs from multiple sensory modalities and provides the most probable output for an specific event.

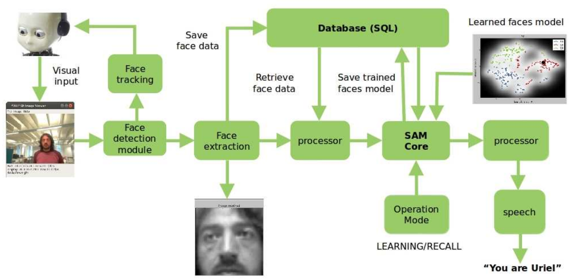

Fig. 3. Face detection. Module for human-robot interaction based on the detection, tracking and recognition of human faces. The recognition process is based on the SAM model.

recognition of human faces. Figure 3 shows the modules for face detection and extraction from visual input of the iCub humanoid robot. For the learning process, a faces model is provided by the SAM model and stored in a database. For the recall process, the observed and recognised face is sent to a speech module for interaction with the human.

2) Human action recognition: For learning by observation of actions performed by humans, an action recognition module was integrated in the iCub humanoid robot. Figure 4 shows the different modules for action recognition with our cognitive architecture. First, the human in front of the robot is detected to start the human-robot interaction. Second, the human arms are detected and tracked by the robot, allowing the robot to observe the actions performed by the human. The actions are based on hand gestures such as waving and manipulating objects such as pushing, pulling, lifting up and putting down. The actions observed by the robot are learned by the SAM model and stored in a database during the learning process. Thus, the iCub humanoid robot is able to observe and recognise various human actions by setting the SAM model in recall mode. Figure 5 shows some examples of the output from the action recognition module integrated in our cognitive architecture.

3) Control based on verbal communication: To provide a robust control and more natural human-robot interaction, a verbal communication is established. We defined a set of vocabs that engages the iCub humanoid to interact, observe

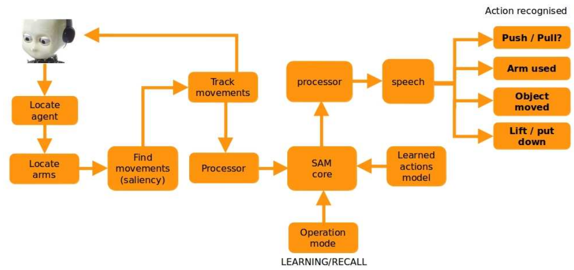

Fig. 4. Action recognition. Detection of the human arms for observation, learning and recognition of human actions based on the SAM model.

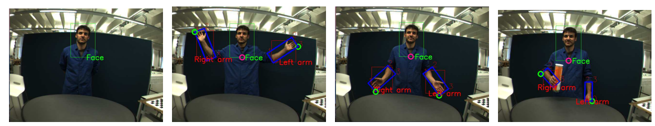

Fig. 5. Output example from the action recognition module. The human and his arms are detection and tracked by the robot to observe and learn the actions being performed. The actions can be based on hand gestures or object manipulation.

and learn from different events, i.e., 'hello iCub', 'let's start', 'who are you?', 'who am I?', 'what is this?', 'what did I do?', 'what happened before?', 'what happened next?', 'how many object are there?', 'let's finish'. Thus, once the iCub humanoid recognises the question asked by the human, it uses the implemented cognitive architecture in recalling mode to answer the specific question.

III. Conclusion

We have presented a cognitive architecture for robot learning by observation and interaction with humans. The core of the architecture, based on a biologically inspired memory model, learns from multiple sensory modalities available in the robotic platform, and provides the most probable output for an specific event. We implemented and tested our cognitive architecture with the iCub humanoid robot, which demonstrated to be a suitable approach to recognise human faces, observe and learn from actions performed by humans.

References

[1] A. L. Thomaz and M. Cakmak, "Learning about objects with human teachers," in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. ACM, 2009, pp. 15–22.

[2] Y. Kuniyoshi, M. Inaba, and H. Inoue, "Learning by watching: Extracting reusable task knowledge from visual observation of human performance," *Robotics and Automation, IEEE Transactions on*, vol. 10, no. 6, pp. 799–822, 1994.

[3] M. Hanheide, M. Lohse, and H. Zender, "Expectations, intentions, and actions in human-robot interaction," *International Journal of Social Robotics*, vol. 4, no. 2, pp. 107–108, 4 2012.

[4] G. Metta, L. Natale, F. Nori, G. Sandini, D. Vernon, L. Fadiga, C. Von Hofsten, K. Rosander, M. Lopes, J. Santos-Victor, *et al.*, "The icub humanoid robot: An open-systems platform for research in cognitive development," *Neural Networks*, vol. 23, no. 8, pp. 1125–1134, 2010.

[5] L. W. Boorman, A. Damianou, U. Martinez-Hernandez, and P. T.J., "Extending a hippocampal model for navigation around a maze generated from real-world data." in *Lecture Notes in Artificial Intelligence (Living Machines), Biomimetic and Biohybrid Systems*, 2015.

[6] The GPy authors, "GPy: A Gaussian process framework in python," http://github.com/SheffieldML/GPy, 2012–2014.