A Memory-based Robot Architecture based on Contextual Information

Ferdian Pratama[1] and Fulvio Mastrogiovanni[2] and Nak Young Chong[1]

***Abstract*— In this paper, we present a preliminary conceptual design for a robot long-term memory architecture based on the notion of *context*. Contextual information is used to organize the data flow between Working Memory (including Perceptual Memory) and Long-Term Memory components. We discuss the major influence of the notion of context within Episodic Memory on Semantic and Procedural Memory, respectively. We address how the occurrence of specific object-related events in time impacts on the *semantics* associated with the representation of those events and the subsequent use of those objects through robot actions. The general architecture design and its implementation are described. A preliminary validation both in simulation and in a real-world environment is discussed. Current work includes an implementation on a Baxter dual-arm manipulator.**

I. INTRODUCTION

Current research activities on robot cognitive architectures mainly aim at providing holistic conceptual models in order to enable robots to perform cognitive tasks [1], [2]. High-level robot cognitive tasks tacitly assume the presence of an underlying memory-based framework to ground perception, representation and retrieval, as well as action-oriented behaviors. Although the most relevant characteristics of *human memory* are far from being understood, the bunch of available scientific knowledge constitutes a useful source of inspiration to organize sensori-motor processes in robot cognitive architectures. Although there is no general consensus about how to design a memory-based framework, memory models typically assume a multi-storage organization, roughly dividing the whole memory space in two areas, namely the Short-Term Memory (STM), which some authors refer to as Working Memory (WM), and the Long-Term Memory space (LTM), which is divided in sub-areas, i.e., the Episodic Memory (EM), the Procedural Memory (PM) and the Semantic Memory (SM).

In the literature, attempts to characterize computational models related to individual WM components [3], as well as LTM components, e.g., EM [4], [5], [6], [7], [8], [9], [10] and PM [11], have been pursued. Among the various approaches, Stachowicz and Kruijff provide a thorough explanation of both design requirements and formal concepts needed to characterize EM and its storage structure [10]. They also provide a brief review about EM in the ISAC framework [5], the SOAR architecture [4], and EPIROME [7], just to name a few. However, the focus of their work is on the notion of *event*, its properties, and its use in such processes as event recognition. Despite their claim of having designed an EM-like memory structure, it is noteworthy that they do not exploit the notion of *context*, which is considered of the utmost importance in [12], [13].

Given an analysis of the literature, two important topics need to be addressed:

- On the one hand, no architectural model seems to explicitly consider functional interconnections among memory modules in a principled way.
- On the other hand, when adopting a holistic approach to the definition of the architecture, a discrepancy between the role of each module and its influence on other modules can be frequently observed.

In particular, if we want to design a robot able to proactively understand its environment and to engage humans in interaction tasks, the need arises to characterize the information flow among the various modules within a memory-based cognitive architecture, specifically integrating the notion of *context*. In humans, context processing is believed to occur in the hippocampus [14]. In particular, it is referred to those mechanisms used to differentiate a given situation from other situations so that the correct behavioral or mnemonic output can be retrieved. In order to achieve this capability also in robots, this paper presents and discusses an interconnected, memory-based robot architecture explicitly taking the notion of context into account. Such an architecture is to be considered the foundation for the design of more complex cognitive processes to occur in robots based on the developmental paradigm.

The main contribution of this paper is twofold: (i) to utilize the notion of *context* as the means to interconnect EM, PM and SM; (ii) to analyze the impact of events on robot behavior (as mediated by the overall architecture) when the notion of context is considered, specifically when context-based information retrieval is employed. It is expected that the introduction of context-based information affects memory retrieval processes, specifically as a means to mimic information awareness mechanisms.

II. ARCHITECTURE AND EXPERIMENTAL SETTING

A. System Architecture

As previously anticipated, we consider a multi-storage model where both Working Memory and Long-Term Memory are explicitly represented. WM is based on the Baddeley updated model [15], which includes the supervisor Central Executive (CE) component, as well as the three so-called *slave* components, namely the Phonological Loop

[1]F. Pratama and N. Y. Chong are with the School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa, Japan. {ferdian, nakyoung}@jaist.ac.jp.

[2]F. Mastrogiovanni is with the Department of Informatics, Bioengineering, Robotics and Systems Engineering, University of Genoa, Genoa, Italy. fulvio.mastrogiovanni@unige.it.

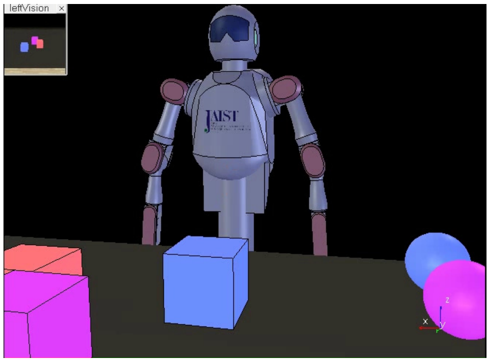


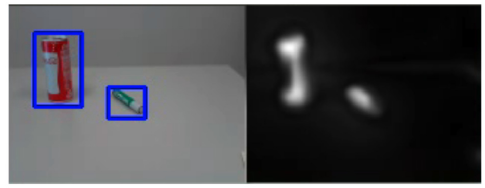
Fig. 1. The experimental scenario in simulation.

(PL), the Visuospatial Sketchpad (VSSP) and the Episodic Buffer (EB). LTM consists of three basic components, i.e., Episodic Memory (EM), which keeps track of specific events localized in time, Procedural Memory (PM), which encodes elementary and composite motor skills, and Semantic Memory (SM), which deals with facts, their meaning and with common sense, general knowledge.

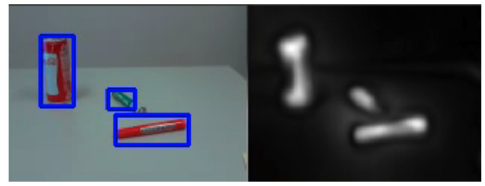
As shown in Figure 1, we assume a human-robot interaction scenario where a human changes the number and the configuration of objects located on a table in front of the robot. In the current implementation, the robot is a passive observer: it is capable of perceiving the environment by means of visual information. From time to time, the human can pose questions to the robot, which are related to what it perceived during the interaction, such as *How many red boxes have been shown?* or *What is the location of the blue box with respect to the red sphere?* In order to answer such questions, the robot must be able to recollect what it previously perceived from its memory and (although such aspect is subject to current work), provide a verbal account of the memory recall process.

B. The Target Human-Robot Interaction Scenario

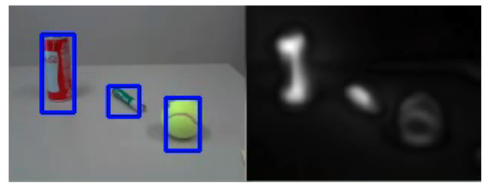
We set-up a scenario where a human operates on objects, which are located on a table, in front of the robot. Objects with different colors and shapes are inserted to and removed from the scene. In this scenario, the robot is just a passive observer. The robot perceives the scene using vision. A visual stream is continuously acquired by the robot as long as the human operates in the scene. Within the robot Field of View (FoV), actions performed by the human are visually spotted by the robot through saliency information. As the visual stream is active, the system processes the input, infers useful information about detected entities (e.g., color, shape, position, size) using an image processing module, based on gist [16] and saliency analysis [17], and consolidates them into LTM. In the experiment, when the human replaces one object with another one, a scene change occurs from the robot perspective. As a consequence, a new memory element (related to the new spotted entity) is consolidated inside LTM. Actions performed by the human in the experiment are aimed at addressing different memory modules. Specifically, SM is involved whenever a novel entity is detected, EM is releated to whole changes in the scene, whereas PM is addressed during an event occurrence.



(a) Scene 1, captured during *knowledge acquisition* step 1 and 5



(b) Scene 2, captured during *knowledge acquisition* step 3



(c) Scene 3, captured during *knowledge acquisition* step 7

Fig. 2. The sequence of scenes as perceived by the robot: the raw feed on the left hand side, the corresponding saliency map on the right hand side.

The experimental procedure consists of two phases, namely knowledge acquisition and memory retrieval.

Knowledge acquisition. Inizially, two entities (namely a red can and a green marker pen) are presented in the visible part of the robot workspace (Figure 2 on top). The robot acquires and consolidates the scene within LTM. Then, the human presents a novel entity (i.e., a red marker pen, see Figure 2 in the mid) to the robot. Afterwards, the human replaces one entity (i.e., the red pen) with a novel one (i.e., a tennis ball, see Figure 3 on the bottom).

During each scene assessment step, the robot remembers position, color and shape features for each entity.

Memory retrieval. Using a Graphical User Interface, the human inserts cues, their value and several contexts, and the system retrieves any available data based on both cues and contextual information. In the performed experiment, the following questions are posed to the robot.

1) Which entities do you know, which are *red*?
2) Which entities do you know, which are *red*, when *three* entities were present in the scene?
3) Which *green* entity was the *leftmost* one, when *three* entities were present in the scene?
4) Was the *rightmost* entity a *ball*, when *three* entities were present in the scene?
5) Was the *rightmost* entity a *ball*, when a *box* was present in the scene?

The questions can be formally translated into a set of $(cue, value)$ pairs – for instance, $(color, red)$ – and contexts – for instance $(green, 3)$ – as shown in Table I. It is

TABLE I: Input sets for the experiment

No.	Cue	Value	Context			
			Pos	Shape	Color	Count
1	Color	Red	-	-	-	-
2	Color	Red	-	-	-	3
3	Pos	LeftMost	-	-	Green	3
4	Shape	Ball	RightMost	Ball	-	3
5	Shape	Ball	RightMost	Box	-	-

TABLE II: Results for each of the input set

No.	Result
1	Red can, red marker pen
2	Scene 2: red can, red marker pen Scene 3: red can
3	Scene 2: green marker pen Scene 3: green marker pen
4	Yes, in scene 3
5	No

noteworthy that we artificiously distinguish between the *Knowledge acquisition* and the *Memory retrieval* process. In principle, questions can be posed at *any time* during the experiment, which reflects the constructivistic nature of the developed robot knowledge.

Since no *a priori* knowledge is considered, LTM is initially empty. As the robot experiences and consolidates new scenes, the persistent nature of LTM allows it to progressively gather knowledge from its personal experience. It is noteworthy that no forgetting mechanism is employed, which means that the knowledge acquired by the robot develops *monotonically*.

III. Discussion

As a result of the experiment, four SM and three EM episodes are maintained. EM and SM data correspond to each captured scene and detected entity, respectively. Instead of four (like in SM), only three EM memory elements are consolidated. This is due to the fact that during *step 5* in Knowledge acquisition (Figure 2 on the top), the system concludes that there are no differences if compared with the scene consolidated in *step 1*. It is noteworthy that this case is very unlikely to happen in real-world settings.

The first question is related to red entities. Red entities that are known to the robot are two, namely a red can and a red marker pen. The can is always present in the robot FoV, whereas the marker pen has been detected in *scene 2* (Figure 2 in the mid). The second is more specific, in that it requires the robot to recall red objects detected only when three different objects were present in the scene. Therefore, *scene 1* is not considered. Consistently, two objects are recalled from *scene 2* and one object from *scene 3*. The third question is related to the qualitative position of the green entity, but only when two entities were present. Again, *scene 1* is not considered, whereas *scene 2* and *scene 3* are used to recall a green marker pen. It is noteworthy that in *scene 3*, two green entities are present, namely the marker and the ball, but the marker is the leftmost one. As a result of the fourth question, *scene 3* is used to recall that a ball was the rightmost object. Finally, a ball is never the rightmost entity, when scenes include a box.

Table II shows a summary of these results. It is noteworthy that some information is omitted, such as all the features of the retrieved objects, e.g., color, shape, etc.

IV. Conclusions

In this paper, we present a conceptual design for a memory-based architecture based on contextual information. The framework is inspired by current studies in developmental psychology, and adopts a biologically-inspired image processing algorithm. Current work is aimed at implementing the architecture on a Baxter dual-arm manipulator (with a particular focus on sensori-motor processes, which are not considered in this paper), as well as integrating a speech-based human-robot interface.

References

[1] A. F. Morse, J. de Greeff, T. Belpeame, and A. Cangelosi, "Epigenetic Robotics architecture (ERA)," *IEEE Trans. Auton. Mental Develop.*, vol. 2, no. 4, pp. 325–339, 2010.

[2] F. Bellas, A. Faina, G. Varela, and R. J. Duro, "A cognitive developmental Robotics architecture for lifelong learning by evolution in real robots," in *Proc. Int. Joint Conf. Neural Networks*, 2010, pp. 1–8.

[3] J. L. Phillips and D. C. Noelle, "A biologically inspired working memory framework for robots," in *Proc. IEEE Int. Workshop Robot Hum. Commun. (ROMAN)*, 2005, pp. 599–604.

[4] A. Nuxoll and J. E. Laird, "A cognitive model of episodic memory integrated with a general cognitive architecture." in *ICCM*. Citeseer, 2004, pp. 220–225.

[5] W. Dodd and R. Gutierrez, "The role of episodic memory and emotion in a cognitive robot," in *Proc. IEEE Int. Workshop Robot Hum. Commun. (ROMAN)*, 2005, pp. 692–697.

[6] N. S. Kuppuswamy, S. H. Cho, and J. H. Kim, "A cognitive control architecture for an artificial creature using episodic memory," in *Proc. Int. Joint Conf. SICE-ICASE*, 2006, pp. 3104–3110.

[7] S. Jockel, D. Westhoff, and J. Zhang, "Epirome-a novel framework to investigate high-level episodic robot memory," in *Proc. IEEE Int. Conf. Robot. Biomim. (ROBIO)*, 2007, pp. 1075–1080.

[8] A. M. Nuxoll, "Enhancing intelligent agents with episodic memory," Ph.D. dissertation, Citeseer, 2007.

[9] Z. Kasap and N. Magnenat-Thalmann, "Towards episodic memory-based long-term affective interaction with a human-like robot," in *Proc. IEEE Int. Symp. Robot Hum. Interact. Commun. (Ro-Man)*, 2010, pp. 452–457.

[10] D. Stachowicz and G. Kruijff, "Episodic-like memory for cognitive robots," *IEEE Trans. Auton. Mental Develop.*, vol. 4, no. 1, pp. 1–16, 2012.

[11] R. Salgado, F. Bellas, P. Caamano, B. Santos-Diez, and R. Duro, "A procedural long term memory for cognitive Robotics," in *Proc. IEEE Workshop Evol. Adapt. Intell. Sys. (EAIS)*, 2012, pp. 57–62.

[12] E. E. Smith and S. M. Kosslyn, *Cognitive psychology: Mind and brain*. Pearson Prentice Hall, 2006.

[13] D. R. Godden and A. D. Baddeley, "Context-dependent memory in two natural environments: on land and underwater," *Brit. J. Psychol.*, vol. 66, no. 3, pp. 325–331, 1975.

[14] D. M. Smith and S. J. Mizumori, "Hippocampal place cells, context, and episodic memory," *Hippocampus*, vol. 16, no. 9, pp. 716–729, 2006.

[15] A. Baddeley, "The episodic buffer: a new component of working memory?" *Trends Cogn. Sci.*, vol. 4, no. 11, pp. 417 – 423, 2000.

[16] A. Oliva and A. Torralba, "Building the gist of a scene: the role of global image features in recognition," *Prog. Brain Res.*, vol. 155, pp. 23–36, 2006.

[17] S. Jeong, S. Ban and M. Lee, "Stereo saliency map considering affective factors and selective motion analysis in a dynamic environment," *Neural Networks*, vol. 21, pp. 1420–1430, 2008.