An Integrated Epigenetic Robot Architecture via Context-influenced Long-Term Memory

Ferdian Pratama
School of Information Science
Japan Advanced Institute
of Science and Technology
Ishikawa, Japan
Email: ferdian@jaist.ac.jp

Fulvio Mastrogiovanni
Department of Informatics, Bioengineering,
Robotics and Systems Engineering
University of Genoa
Genoa, Italy
Email: fulvio.mastrogiovanni@unige.it

Nak Young Chong
School of Information Science
Japan Advanced Institute
of Science and Technology
Ishikawa, Japan
Email: nakyoung@jaist.ac.jp

***Abstract*—In this paper, we present a conceptual design for a context-influenced Long-Term Memory architecture. The notion of context is used as a means to organize the information flow between the Working Memory and Long-Term Memory components. In particular, we discuss the major influence of the notion of context within the Episodic Memory on the Semantic and Procedural Memory, respectively. In other words, we address how the occurrence of specific events in time impacts on the meaning of those events and the subsequent use of objects through robot actions. The general architecture design and its implementation in a simulated scenario are described. Such issues as memory items representation, individual structures of Long-Term Memory components, as well as memory-based recognition and item retrieval processes, are discussed in detail.**

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 architecture to ground perception, representation and retrieval, as well as action-oriented behaviors. Although the most relevant characteristics of *human memory* are far from being assessed, it constitutes a useful source of inspiration to manage information flow in robot cognitive architectures. Although there is no general consensus about a general framework, memory models typically assume a multi-storage organization, roughly dividing the whole memory space in two areas, namely the Working Memory (WM), which some authors refer to also as Short-Term Memory (STM), and the Long-Term Memory space (LTM), which is divided in sub-areas, e.g., 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 have been pursued [3], as well as LTM components, e.g., EM [4]–[10] and PM [11]. Among the various approaches in the literature, 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 but 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, event nesting, and event types of complexity. Despite their claim of having designed an EM-like memory structure, it is noteworthy that they do 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 consider the explicit functional interconnection between memory modules in a principled way. On the other hand, when adopting a holistic approach to the definition of the architecture itself, discrepancies between the role of each module and its influence to other modules can be frequently found.

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 relationships between the various modules within a memory-based, robot cognitive architecture, specifically integrating the notion of *context*. In humans, context processing is believed to occur in the hippocampus [14]. Specifically, it refers to those mechanisms to differentiate a particular situation from other situations so that the correct behavior or mnemonic output can be retrieved. In order to achieve this objective also in robots, this paper presents and discusses an explicitly 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.

Developmental does not necessarily means machine learning, and from the perspective of this work, it can be considered as a continuous knowledge acquisition that allows progressive improvements of knowledge quality (from overall robot motoric skills until the advance usage of knowledge obtained from the robot's personally experienced events) that grows along with human interaction. 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 the robot behavior (as mediated by the architecture) when the notion of context is considered. It is expected that the introduction of context-based information affects memory retrieval processes, specifically as a means to mimic awareness mechanisms.

The paper is organized as follows. Section II formally introduces the envisaged architecture on the basis of available Neuroscience and Psychology literature. Section III describes the experimental setting and reports about the related discussion. The conclusion follows.

II. A Memory-based Robot Architecture

In this Section, we describe the proposed architecture detailing each component. As previously anticipated, we consider a multi-storage model where both the Working Memory and the 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 passive storages namely the Phonological Loop (PL), the Visuospatial Sketchpad (VSSP) and the Episodic Buffer (EB). The LTM consists of three basic components, i.e., the Episodic Memory (EM), which keeps track of personally experienced events localized in time, the Procedural Memory (PM), which stores the elementary and composite motor skills, and the Semantic Memory (SM), which deals with facts, their meaning and in general with common sense knowledge.

As a working hypothesis, we assume a human-robot interaction scenario where a robot is capable of perceiving its environment by means of visual information. The robot can perceive scenes, where simple objects of well-defined color and shape are placed on a table. A human can pose the robot questions related to its perceptions, 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 saw from its memory and (although such aspect is not covered in this paper), verbalize it.

A. Working Memory

WM is a theoretical framework which refers to both structures and processes used to temporarily store and reason upon information. It is believed that WM and LTM interact continuously. However, the nature of such interaction is still not completely clear. According to the Atkinson and Shiffrin model, information is transferred to LTM as long as it is attended in WM [16]. From the WM point of view, LTM is only considered a (possibly complex) memory storage, whose encoding and decoding processes are arranged and managed by CE. In the proposed architecture, we address how CE can encode memory items in LTM such that they are interconnected with each other in EM, PM, and SM. In human memory, CE is responsible for processing information originating from different sources, coordinating a number of passive slave subsystems, as well as performing selective attention and inhibition strategies. In our architecture, we model CE as a system able to perform a number of tasks, as follows.

- Explicitly managing memory encoding and decoding processes in such LTM components as EM, PM, and SM, specifically using contextual information.
- Exhibiting *familiarity*-like information retrieval, i.e., *how* to identify cues to be used, based on logical processes involving cues analysis and *problem awareness* [17], [18].
- Manifesting *recollection* behaviors, i.e., recalling LTM memory items from the results of familiarity retrieval processes if they match the desired retrieval cues.
- Supervising the Phonological Loop slave component, i.e., by analyzing verbal information related to recalled LTM memory items, and the Visuospatial Sketchpad slave component, related to visual information, i.e., object shapes, colors or locations (as perceived in a scene).

B. Long-Term Memory

LTM is considered as a virtually infinite store of information where memory items can be kept indefinitely through synaptic consolidation. When considering a robot-targeted implementation of a memory-based architecture, such issues as efficient information storage and retrieval must be carefully designed. If we consider robots operating in real-world environments, we expect their knowledge to exponentially grow as they operate. In a memory-based robot architecture, this means that the WM must be able to encode in and, above all, decode from LTM a huge amount of memory items. In order to retrieve a given memory item, in the absence of any contextual information, it would be necessary to check all the LTM content. Such process may seriously disrupt robot behavior. In this paper, we focus on designing an efficient method for information retrieval by exploiting the so-called *recognition* principle.

Recognition can be further specialized in two principles, namely familiarity and recollection, as being defined by Tulving [19] and later by Mickley [20]. Recognition assumes the availability of associative memory mechanisms. Specifically, when a memory item is to be recollected, various related cues in WM are used. Then, memory items in LTM that have associations with the cues are considered. Recollected memory items are characterized by very strong associations (i.e., a *sense* of familiarity) with the used retrieval cues. Starting from experiments with subjects, Tulving suggested that recollection is based on EM due to autonoetic consciousness properties, whereas familiarity is based on SM given its noetic characteristics [19], [21]. In other words, whilst familiarity corresponds to a subjective sense of having previously encountered a stimulus, recollection refers to specific contextual events and details obtained in the past experience.

Based on Tulving's definition, we further specialize the familiarity principle in two processes: 1) being aware of the memory item's existence in LTM, before any recall occurs, and 2) being able to recall the memory item independently of the contextual information related to it. The second process well adheres to the classical Tulving's definition. In our architecture, we introduce a *familiarity filtering index* (FFI), with the aim of modeling the first process by limiting the number of possible memory items that must be checked in LTM. In particular, given an LTM memory structure, a retrieval cue r, and a desired memory item (i.e., a *content*) i, the process can be modelled as

$$i = \texttt{recall}(\texttt{FFI}(r, LTM)).$$

The `FFI` procedure searches for memory items characterized by a high level of familiarity with the provided cues. To accurately define what familiarity is, it is necessary to describe how memory items are represented. Each memory item is characterized by two sets of descriptors, namely the *Body ID* and the *Skills ID*. The first set of descriptors is related to the robot body schema and in particular to robot parts actually involved in the memory item (e.g., left or right body part, hands or arms). The second set of descriptors refers to the specific goal-oriented behaviors employed by the robot while interacting with the environment, as well as the goals of the interaction. `FFI` first seeks to satisfy cues related to the robot body (in order to weight perceptions), then to its actual behaviors (i.e., past experiences).

Our model is based on the notion of *factor*.

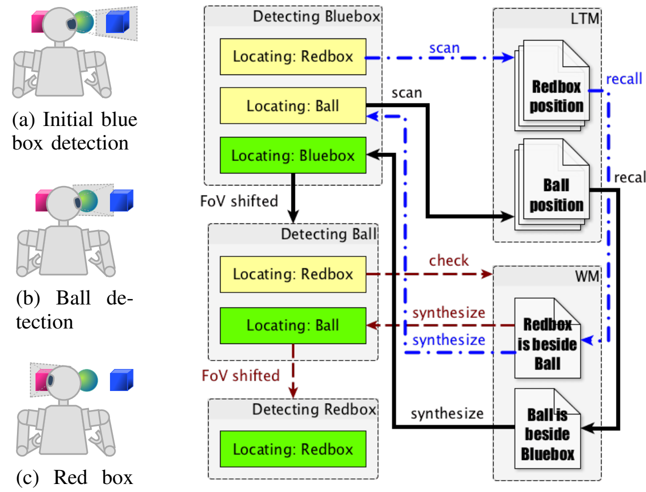
Definition 1 (Factor): A factor $f \in F$ is a single element that forms a SM and PM item, where $F = SM \cup PM$. A factor consists of n cue-value pairs, such as $f = \{(r_1, v_1), \ldots, (r_n, v_n)\}$.

1) Semantic Memory: We now formally define SM, where concepts and common sense knowledge are stored, and then PM, which stores robot motor behaviors. It is noteworthy that SM is characterized by robot-independent knowledge, provided that different robots have the same perception capabilities, whereas the content of PM is robot-dependent. In a sense, it is possible to say that PM stores information about *knowing how* to do something, whereas SM stores information about *knowing what* (knowing about something).

Definition 2 (Semantic Memory): A Semantic Memory SM consists of 5 factor types, i.e., $SM = \{N, H, L, T, W\}$, where: N represents known (or previously identified) entities; H stores information related to humans the robot may interact with; L is related to spatial information; T represents unexperienced factual, future time-reference (i.e., activities to be performed in the future), and W is related to lexical knowledge, which may be used in human-robot interaction tasks.

Of particular interest for our discussion are the notions of entities and locations, which we define as follows.

Definition 3 (Entity): An entity $n \in N$ can be one of 5 *primitive* entities (i.e., cube, plane, disc, cylinder, sphere) or



(a) Initial blue box detection

(b) Ball detection

(c) Red box detection

(d) Information flow of entities by relative position

Fig. 1: Recursive memory item callback during position recollection: the location of an entity is determined relative to previous memory information.

a *compound* entity, such as $N = N_P \cup N_C$, where $N_P = \{N_{cu}, N_{pl}, N_{di}, N_{cy}, N_{sp}\}$, and $N_C = {N_P}^c$, where N_P is the set of known primitive entities, and N_C is the set of known compound entities (i.e., complementary to N_c).

It is noteworthy that each entity is characterized by a number of multi-valued factors, such as *name*, *shape* and *color*, as well as a number of Boolean parameters, e.g., *graspable* and *manipulable*. Each factor is characterized by specific cues. For example, a *bluebox* entity is characterized by $\{(name, bluebox)$, $(shape, cube)$, $(graspable, true)$, $(manipulable, true)$, $(color, blue)\}$.

Definition 4 (Location): A spatial location $l \in L$ can be either absolute or relative, such as $L = \{L_a, L_r\}$, where L_a and L_r are the set of items related to absolute and relative locations, respectively. A relative location contains relative position and its reference. It is applicable to current robot behaviors and can be recursively used (Figure 1).

For example, further characterization of *bluebox* includes location information, which may be relative with respect to the robot perspective, such as $(location, < 0.72, 0.13, -0.29 >)$, as expressed using a robot-centered Cartesian reference system, or a previously detected object.

In the current SM implementation, Body ID descriptors correspond to factors for each memory item, whereas Skills ID descriptors correspond to specific aspects of each factor. This enables the robot to filter out cue-unrelated memory items.

Definition 5 (Context): A context c is made up of cue-value pairs corresponding to a particular factor f, and it is defined as $c = \{(r_{f_1}, v_{f_1}), \ldots, (r_{f_n}, v_{f_n})\}$, where n is the number of desired contextual elements provided by humans during the interaction with the robot.

Context is given by human and it is directly related to scene and indirectly related to an event. During an event recollection, context have the effect of filtering down the number of personally experienced scenes, yielding a series of matching scenes which is an *event with respect to a certain context*.

Definition 6 (Retrieval Cues): A retrieval cue $r \in R$ is an element used to recollect a memory item, such as $R = R_g \cup R_c \cup R_s$, where R_g is a set of general cues, R_c is a set of context-dependent cues, and R_s is a set of state-dependent cues.

The three types of cues are defined as follows.

Definition 7 (General Cue): A general cue $r \in R_g$ is available regardless of the type of memory it resides in.

Definition 8 (Context-dependent Cue): Given a *scene* δ the robot is observing, a context-dependent cue $r \in R_c$ indirectly corresponds to an event occurring in δ.

It is noteworthy that $R_c \subset EM$, i.e., context-dependent cues are part of the Episodic Memory.

Definition 9 (State-dependent Cue): Given a factor f and a sequence of (possibly goal-oriented) robot behaviors Q, a state-dependent cue $r \in R_s$ corresponds to f as it is encoded in SM and PM.

From the last definition, it follows that $R_s \subset SM \cap PM$.

It is now possible to better define Procedural Memory.

2) Procedural memory: As a preliminary step, it is necessary to specify the difference of meaning between the concepts of movement and skill that are used in the paper. We refer in a generic sense to *movement* as a robot motion not targeted to any specific objective, whereas we refer to *skill* as a goal-oriented behavior, which may be *a priori* known or learned.

Definition 10 (Procedural Memory): A Procedural Memory PM consists of 3 classes of skills, namely elementary skills P_e, alias skills P_a and composite skills P_c, such that $PM = \{P_e, P_a, P_c\}$.

A Procedural memory item consists of general and state-dependent cues. It is noteworthy that in this case a state-dependent cue is a triple defined by an objective, a *significant* symbolized by the cue *denotes* and a sequence of skills, which are defined as follows.

Definition 11 (Objective): An objective o is the set of all the parameters which must be grounded to perform a given goal-oriented behavior, such that $o = \{x_1, \ldots, x_n\}$, where the generic x element is a name-value pair.

Definition 12 (Elementary Skill): An elementary skill $p \in P_e$ is a basic and atomic goal-oriented robot behavior.

With *atomic*, we refer to the fact that an elementary skill is so simple that cannot be further decomposed. Such a skill can be performed regardless of the existence of an objective [22].

Definition 13 (Alias Skill): An alias skill $p \in P_a$ constitutes a variation of any existing skill, such that some or all of the parameters characterizing the skill may be differently grounded. It is characterized by an additional state-dependent cue *denotes* that signifies the referred basic skill, although with a different *name* and *objective*.

Definition 14 (Composite Skill): A composite skill $p \in P_c$ is a complex skill consisting of already existing low-level elementary, alias or recursively composite skills.

Composite skills are made up of general and state-dependent cues.

Definition 15 (Sequence): A sequence $q \in R_s$ is a state-dependent cue that contains an ordered set of skills which will be sequentially performed, such that $q = \{p_1, \ldots, p_n\}$, with $n > 1$, where the generic p element stands for low-level skills (i.e., elementary, alias or composite).

3) Episodic memory: Visual representation of detected environment from the robot vision system are stored in the EM. Humans have the ability to "mentally travel through time" to reexperience their past during event-recollection. Even though being deducted having a relation to the EM [23]–[25] and the process of to do so is still under investigation, we relate the significant aspects to construct the formal design in robotics.

Storing visual streams of information continuously in a constant & periodic fashion is considered not an efficient implementation, as it increases storage consumption, especially when no significant changes are detected during a considerable period of time. Here, instead of recording the visual stream continuously, we store the analysis results of visual representation snapshots of the detected robot's environment obtained by vision systems. These snapshots are a representation of any significant changes detected through the robot's Field of View (FoV). It is captured in a non-linear fashion based on the occurrence of an event, meaning that if changes caused internally (from the robot's movement) or externally (otherwise) is visually detected, a scene will be encoded. This way solved the storage consumption problem, as the only information during an occurrence of an event being encoded.

Definition 16 (Episodic Memory): The Episodic Memory EM is the set of experienced past scenes analysis results of visual representation snapshots.

Definition 17 (Scene): A scene $\delta \in EM$ is the changes of visually detected input, which indicates the occurrence of an event at a particular time. In short, a scene is an *event marker*. Anything occurs between two distinct scenes is defined as an *event*. Scenes that have been captured are stored in the EM.

A scene δ consists of $\{t, c, \nu, e_T, Q, o\}$, where t is the encoding time, c is the context for that event, ν is the event name, $e_T \in E_T$ is the event type, $Q = \begin{cases} Q & \text{if } e_T = e_A \\ \varnothing & \text{if } e_T = e_P \end{cases}$, $o = \begin{cases} o & \text{if } e_T = e_A \\ \varnothing & \text{if } e_T = e_P \end{cases}$.

Definition 18 (Event): An event e is associated with, and occurred over a period of time, which marked from two distinct scenes correspond to the beginning & end of an event. It consists of multiple scenes during the period of that event, defined as $e = \{\delta_1, \ldots, \delta_n\}$, given n is the number of scenes

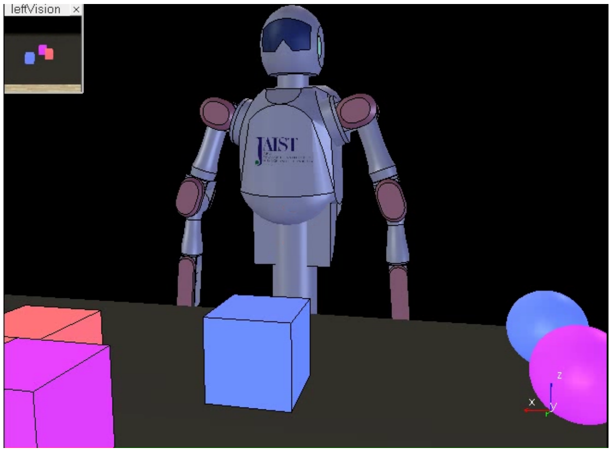


Fig. 2: The experimental scenario in simulation.

for that particular event.

Definition 19 (Type of Event): An event e can either be an Active event e_A, where the robot performs a sequence of sensorimotor skills from the procedural memory during the encoding of the event, or a Passive event e_P, where it only witness phenomena that change the state of the surroundings, without performing any motoric skills from the procedural memory. It is formally defined as $E_T = \{e_A, e_P\}$.

III. EXPERIMENTAL VALIDATION

A. Simulated Scenario and Experiment

In order to evaluate how contexts influence the Episodic Memory, scene-encoding experiments have been conducted. The experimental setting is described as follows.

- Different entities are presented to the robot, which are characterized by different *color* (e.g., purple or blue) and *shape* (e.g., cube or ball). The scene-encoding process starts as soon as the presented entities are detected.
- A number of robot basic skills are available (e.g., *grasp*, drag_left or release), which may or may not be encoded as composite skills.
- During the experiment, the scene perceived by the robot is expected to change (e.g., different objects are presented or objects are displaced).
- After the experiments, the robot is requested to describe a given scene or to explain what happened between different scenes, on the basis of a user-provided context.

In the experiment, the robot *field of view* is kept fixed. Five distinct scenes are used to demonstrate how contextual information influences robot's *understanding* about a scene, i.e., about the replaced entities with different shapes and/or color, as well as new entities. Scene analysis covers both the global and the local information about each detected entity via blob detection [26]. This algorithm is used to get the shape features to determine each entity detected, by dividing the image region into smaller blocks and analyze the shape by the amount of orientation angles considered and each of the amount detected, which the details are available in [26]. The global information yields statistical measures to determine whether changes are visually spotted, which constitute a new perceived scene. Local feature extraction analyzes each detected entity and yields entity count, blob information, as well as grounded color and shape features (i.e., the orientation histogram using a Gabor filter). These information is encoded in SM.

Table I illustrates the extracted local features to be encoded in the memories from their respective snapshots during any visual changes detected. It is noteworthy that scenes are independent from each other, meaning that a distinct *tracking* module is applied to track a particular entity across multiple scenes, by comparing the difference of output values of each entity with respect to a particular threshold value. Changes in objects configuration across scenes can be caused by either robot actions, or external influence (i.e., a human operating on entities).

In the first experiment, we test the recognition of entity properties, i.e., the assessed relationships between EM and SM. Given the whole experiment represented by the five scenes, we first query the retrieval cue with a desired context, then we observe the result (Table II). Currently, only a single cue can be processed at a time with the provided context. Note that the position is currently set as verbal cues (e.g. *leftMost* and *rightMost*) as the representation of minimum and maximum value of *posX+sizeX* respectively.

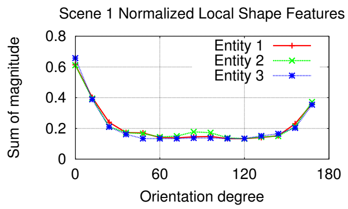
In the second experiment, in order to evaluate the inter-connection of EM and PM, several robot skills within PM are performed during the scene-encoding process. Performed robot skills are associated with a particular scene, i.e., the scene showing the outcome of robot action. In particular, we build two composite skills, namely *left_skill1* and *left_skill2*, which are made up of basic skills as follows: *left_skill1* = {*confirm_given_entity, grasp, reach_table, release, standby*}, (the blue box is inserted in the scene) and *left_skill2* = {*grasp, drag_left, release, left_skill1*} (first, the blue box is removed from the robot field of view, then the blue ball is inserted into the scene). These two composite skills are assumed to be executed, respectively, in between scenes $3-4$, and $1-2$.

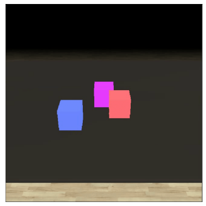
B. Discussion

As argued in Section II, one of the characteristics of EM is that its content does not have to be deliberately encoded. As a matter of fact, experiments show that scene-encoding is a seamless process. Any change within the environment that is visually detected by the robot is processed and any past experienced event is only bounded by the *strictness* of the provided context (i.e., the actual number of cues that are used), instead of being *atomic* or *complex* as defined in [10].

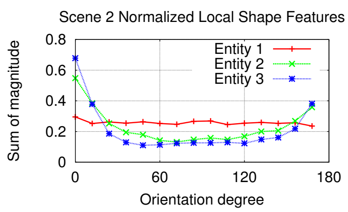
The EM-SM interconnection experiment demonstrates the ability of the robot to be able (in principle) to answer questions like: *How many colored cubes/boxes have you seen so far? Did you see a blue ball at the right hand side when there were five entities being presented?* This can be done since once a scene has been perceived and analyzed, the non

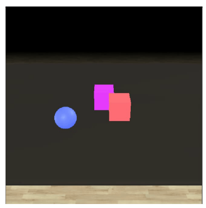
	Blob information		Color (hue)	
	posX, posY	sizeX, sizeY	mean	var
Scene 1				
Entity 1	172, 317	83, 97	0.637	0.0000
Entity 2	294, 257	64, 82	0.8444	0.0044
Entity 3	342, 285	71, 89	0.9905	0.0000



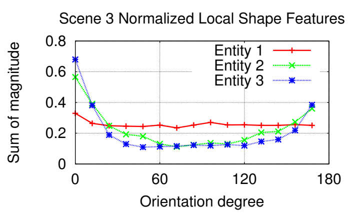


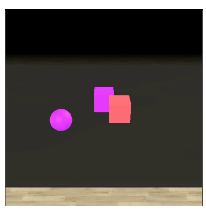
Scene 2				
Entity 1	162, 331	73, 71	0.636	0.0000
Entity 2	294, 259	66, 82	0.847	0.0047
Entity 3	342, 287	71, 89	0.9906	0.0000



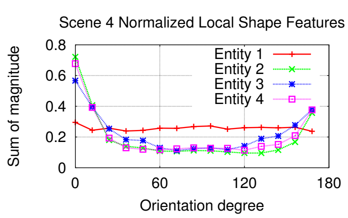


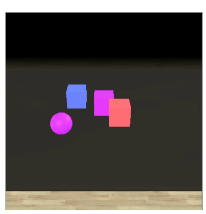
Scene 3				
Entity 1	148, 330	73, 71	0.8151	0.0001
Entity 2	294, 258	66, 82	0.8475	0.0048
Entity 3	342, 286	71, 89	0.9909	0.0000



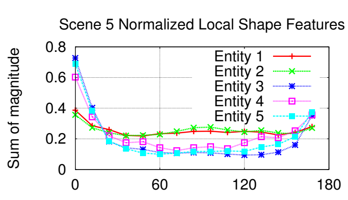


Scene 4				
Entity 1	150, 330	71, 71	0.815	0.0001
Entity 2	200, 238	69, 77	0.6374	0.0000
Entity 3	294, 258	66, 82	0.8478	0.0048
Entity 4	342, 286	71, 89	0.9911	0.0000





Scene 5				
Entity 1	212, 361	77, 75	0.6437	0.0012
Entity 2	150, 331	72, 72	0.8056	0.0017
Entity 3	200, 239	69, 77	0.6374	0.0000
Entity 4	294, 259	66, 82	0.8474	0.0047
Entity 5	342, 287	71, 89	0.9909	0.0000



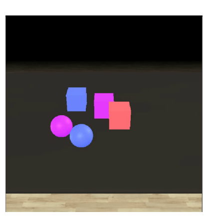


TABLE I: Extracted scene local features.

No.	Cue	Context				Matched Results		
		color	shape	count	pos	scene	ent id	value
1		-	ball	4	rightMost	4	1	$\{0.815, 0.0001\}$
2	Color	-	ball	3	rightMost	2,3	1,1	$\{0.636, 0\}$, $\{0.8151, 0.0001\}$
3		-	-	3	rightMost	1,2,3	3,3,3	$\{0.990, 0\}$
4		-	-	3	rightMost	1,2,3	3,3,3	cube
5	Shape	blue	-	3	rightMost	1,2	1,1	cube,ball
6		blue	-	5	rightMost	5	1	ball
7		blue	ball	5	rightMost	5	1	$true$
8		blue	ball	-	-	2,5	1,1	3,5
9	Count	purple	ball	-	-	3,4,5	1,1,2	3,4,5
10		blue	cube	-	-	1,4,5	1,2,3	3,4,5
11		purple	ball	4	-	4	2	$\{200, 238\}$
12	Pos	blue	ball	5	leftMost	5	1	$false$
13		blue	ball	3	leftMost	2	1	$true$

TABLE II: EM-SM interconnection experimental results.

No.	Cue	Context				Matched Results	
		color	shape	count	pos	scene	value
1		-	ball	3	-	2,3	left_skill2,-
2		-	ball	5	-	5	-
3	Skill name	-	cube	3	-	1	left_skill2
4		-	cube	-	-	1,4	left_skill2, left_skill1
5		blue	cube	-	-	1,4	left_skill2, left_skill1

TABLE III: EM-PM interconnection experimental results.

event-specific features (i.e., shape and color) of a new entity are encoded in SM as *general* knowledge. Table I shows, for all five scenes, both blob and color information for each entity. Blob information is computed as feature vectors using the *sum of magnitudes* [26]. On the one hand, the robot can still recognize the shape of an arbitrary *cube* or what the *blue* color means in hue space (in terms of mean and variance), regardless the connection with any particular event. On the other hand, the robot can recognize the *blue* color according to its knowledge in SM, relate it to a specific event with the given context, and compare it with a given numerical value.

Let us consider Table II in detail. Case 1 shows the result of a blue color represented by mean and variance in hue space, based on the color cue using the context $\{(shape, ball), (count, 4), (position, rightmost)\}$. It corresponds to a hypothetical question like *What is the color of the rightmost*[1] *ball when 4 entities were presented?* In case 2, the context is changed so that $(count, 3)$. Based on that, it yields two different results which show that entity 1 are actually different across two scenes (scene 2 and scene 3). If we generalize the context as in case 3, several scenes matched, but one result is returned due to floating point comparison, which indicates that the three results refer to the same entity. Hence, it is possible to say that during scenes $1-3$, the color based on that particular context is *purple* (see Table I).

The remainder of the Table shows different cues with different contexts. Case 7 shows an interesting result, i.e.,

[1]We assume the availability of a procedure semantically mapping the word *rightmost* with the proper object in the scene.

providing the robot with a shape cue with a particular context, including the shape itself yields a *false* result. This shows the robot's ability to compare the given context and the desired cue.

In the EM-PM interconnection experiment, the composite skills *left_skill1* and *left_skill2* are associated with scenes 4 and 2, respectively, which means that *left_skill1* and *left_skill2* causing scene 4 and 2 to occur. Table III lists noteworthy results. Case 1 can be interpreted as the result of question *What did you do with the ball when three entities were presented?*, which yields only one result as opposed to the two matched scenes, due to the direct association of the skill *left_skill2* to scene 2 and the fact that no skills were applied during scene 3. In case 2, it yields no results even though scene 5 is identified as a matched result when queried with that particular context. Although the robot is *aware* that five entities have been presented, it did not execute any skill, which leads it to conclude that an *external* influence has changed the scene. Providing a more general or specific context might change the result for the cue or maintain the same result, as in case 5. These results corroborate the interconnection between the flexibility of context adjustment and the recollection of internally performed motor skills.

IV. Conclusion

In this paper, we present a developmental robot architecture specifically aimed at addressing the interconnections between modules of Long-Term Memory, namely Episodic, Procedural, and Semantic Memory. We formally introduce the notion of context as a means to constraint memory-based information retrieval in robots. The developed notion of context can accommodate general as well as specific queries posed to the robot memory system. In the paper, we address how the developed notion of context influences the EM-SM and EM-SM information flow. Results in simulation show how this can be achieved using a scene-encoding process. We discuss the interpretation of results and relate them to current psychological studies. On the basis of these premises, current work is aimed at, on the one hand, extending the basic developed principles and, on the other hand, to work on an implementation on a real platform.

Acknowledgment

The authors would like to thank Prof. Sungmoon Jeong for his assistance in developing the image processing module.

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 *Proc. Int. Conf. Cog. Model. (ICCM)*, Jul 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*, Oct 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)*, Dec 2007, pp. 1075–1080.

[8] A. M. Nuxoll, "Enhancing intelligent agents with episodic memory," Ph.D. dissertation, University of Michigan, 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)*, May 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] R. Atkinson and R. M. R.M. Shiffrin, "Human memory: A proposed system and its control processes," in *The Psychology of Learning and Motivation (vol. 2)*, K. Spence and J. Spence, Eds. New York: Academic Press, 1968, pp. 89–195.

[17] F. Mastrogiovanni, A. Scalmato, A. Sgorbissa, and R. Zaccaria, "Problem awareness for skilled humanoid robots," *International Journal of Machine Consciousness*, vol. 3, no. 1, pp. 91–114, 2011.

[18] F. Mastrogiovanni and A. Sgorbissa, "A biologically plausible, neural-inspired planning approach which does not solve 'the gourd, the monkey, and the rice' puzzle," *Biologically Inspired Cognitive Architectures*, vol. 2, pp. 77–87, 2012.

[19] E. Tulving, *Elements of episodic memory*. OUP Oxford, 1985.

[20] K. R. Mickley and E. A. Kensinger, "Emotional valence influences the neural correlates associated with remembering and knowing," *Cogn. Affect. Behav. Ne.*, vol. 8, no. 2, pp. 143–152, 2008.

[21] J. Gardiner, "Episodic memory and autonoetic consciousness: a first-person approach," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 356, no. 1413, pp. 1351–1361, 2001.

[22] F. Mastrogiovanni and A. Sgorbissa, "A behavior sequencing and composition architecture based on ontologies for entertainment humanoid robots," *Robotics and Autonomous Systems*, vol. 61, no. 2, pp. 170–183, 2013.

[23] H. Eichenbaum and N. J. Cohen, *From conditioning to conscious recollection: Memory systems of the brain*. Oxford University Press, 2001.

[24] E. Tulving, "Episodic memory and common sense: how far apart?" *Philos. Trans. R. Soc. Lond.*, vol. 356, no. 1413, pp. 1505–1515, 2001.

[25] ——, "Episodic memory: from mind to brain," *Annu. Rev. Psychol.*, vol. 53, no. 1, pp. 1–25, 2002.

[26] 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.