Emergent emotion as a regulatory mechanism for a cognitive task implemented on the iCub robot

Murat Kirtay, Lorenzo Vannucci, Egidio Falotico, Cecilia Laschi
The BioRobotics Institute
Scuola Superiore Sant'Anna
Pontedera (PI), Italy
{m.kirtay, l.vannucci, e.falotico, c.laschi}@santannapisa.it

Erhan Oztop
Department of Computer Science
Ozyegin University
Istanbul, Turkey
erhan.oztop@ozyegin.edu.tr

***Abstract*—In this study, we employed an emergent emotion model, based on neuro-computational energy regulation, to carry out a cognitive task. The experiment involved visual recalling and was performed by a physically embodied agent (iCub humanoid robot). In this task, the agent operates its associative memory (Higher-Order Hopfield Network) to form a stimulus-energy association for each perceived input. Then, the agent uses these associations to derive an internal reward signal in a reinforcement learning framework to make a sequence of actions (i.e., coordinated head-eye movements) to discover the states where the minimal computational energy is required to perform the task. The results indicate that the agent successfully utilizes this model to act in an unknown environment by following the energy minimization principle. On the basis of obtained results, we suggest that exploiting this approach will give rise to rich applications for developmental robotics where emergent (that is, not reflexive) behaviors are necessary for higher cognitive functions – planning, decision making, etc.**

***Index Terms*—emotion, energy regulation, visual recalling**

I. Introduction

Emotion is one of the critical mechanisms that plays essential roles in cognitive functions and developmental processes such as recalling memories, behaviorally responding caregivers while playing peekaboo, regulating prediction of the reward and short-cutting cognitive processes [1], [2]. In humans, the effects of the emotion can be behaviorally observed throughout the lifespan – starting from early infancy to the late stage of the development. That is why emotion has been a fertile field for many diverse disciplines ranging from the behavioral economics to the humanoid robotics. However, we restrict the scope of this study to provide an overview of the emotion-related studies in the context of cognitive developmental robotics [3].

In this paper, we designed an experiment – by extending our previous works [4], [5] – to emulate the regulatory role of the emotions in a physically embodied agent (iCub Humanoid Robot). We note that with emotions we refer to high-level such as well-being and boredom, we indicate that the types of emotion which bring about computational benefits while performing cognitive functions (e.g., recalling memories) rather than reflex-like responses which are performed just in time. To be concrete, we hold that short-cutting computationally intensive processes to reduce neural energy consumption yields what we term "emotion". This functional aspect enables a biological agent to rely on suboptimal solutions rather than searching the best solution for critical – sometimes intractable– problems such as mate selection, visual attention, to mention a few examples.

Here, we provide representative approaches related to the emotion: however, the extensive review and biological support of the proposed implementation can be found in our previous studies [4], [5]. One approach is to recognize human facial expressions and label them based on basic emotions such as happiness, disgust, fear, surprise or implementing these predefined categorical emotions on an artificial agent to adjust the behaviors [6], [7]. Another approach is to use a weighted combination of the basic emotions or use arousal/valence dimensions to extract the affective state of the agent [8], [9], [10]. Note that the literature studies are mostly concerned about the two approaches mentioned above. We suggest that these approaches may simplify the complex nature of emotions and may not answer how an emotion emerges in the agent's neural system to create computational and behavioral benefits to the agent. More importantly, most of the studies were not employed on an actual hardware platform to asses whether proposed methods can be employed in a real-world scenario [4], [5].

Instead of focusing on basic emotions, our proposal in this study centered on the computational benefits (e.g., relying on "good enough" choices over searching the best options) of the emotion during the decision-making that an agent may face throughout its developmental processes. In doing so, we addressed the cognitive aspects of emotions regarding the following functionalities: short-cutting cognitive processes, storing and recalling memories and value (e.g., reward) extraction for action selection. To employ these functionalities on the iCub robot, we designed a visual recalling experiment in which the robot process perceived visual stimulus from the environment and construct stimulus-energy association via its auto-associative memories. Then, the robot employs this association to extract a reward value in order to select an action (i.e., moving its gaze towards a state where visual stimulus is

This research has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2) and the Specific Grant Agreement No. 720270 (Human Brain Project SGA1).

located) which leads to minimizing the computational cost of the recalling task. To perform this action selection sequentially, we adopt a temporal difference learning algorithm in which the reward signal is internally generated.

The results show that the proposed method enables the robot to find states associated with a low amount of energy consumption to accomplish visual recalling task. Contrary to most of the literature studies, the displayed behavior relies only on the internal dynamics of the agent without having an explicit reward assignment by an operator. Based on the experimental results, we conclude that emotion should be considered as a regulatory mechanism for developmental robots which performs higher-order cognitive processes (such as planning and decision making) rather than reflex-like behavior.

This paper is organized as follows: we introduce the performed methods in Section II. Section III provides details about the experimental setup and iCub robot. The experimental results and discussions based on obtained results are explained in Section IV. Lastly, the conclusions of this work and research objectives for future studies are highlighted in Section V.

II. Methods

In this section, we introduce the methods to form associative memories for perceptual processing and to extract reward values by exploiting energy-stimulus associations. For the former part we employed a customized Hopfield network, for the latter part we implemented a temporal difference learning algorithm. We noted that the detailed description of the performed methods and biological background can be found in [4], [5].

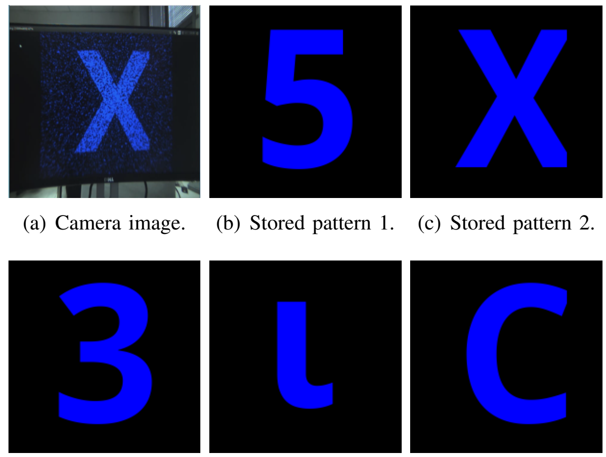
A. Higher-Order Hopfield Network: stimulus-energy association

To form an associative memory based on perceived visual stimuli from the environment, we employed the Higher-Order Hopfield network (HHOP) [11]. To do this, the network trained with the five different visual patterns which are selected to be either a number or a letter. Figure 1 shows one of the received camera images and stored patterns to train the network.

The perceptual processing begins with seeing an image from the robot's camera. Then the visual pattern processed by applying standard image processing algorithms – namely grayscaling, binarization, and downscaling image to 20×20– to obtain bipolar representation $(-1, 1)$. Activation of a unit (e.g., neuron) in the network obtained by Eq. 1 as the product of the activation of pair-units in the network.

$$S_i = sgn\left(\sum_{jk} W_{ijk} S_j S_k\right) \tag{1}$$

The $sgn(x)$ function outputs 1 if $x \geq 0$, otherwise -1. To store a set p of inputs with bipolar representation, the weights, W, computed as $W_{ijk} = \sum_p \xi_i^p \xi_j^p \xi_k^p$. In this derivation, ξ_i^p, ξ_j^p and ξ_k^p are the jth, jth and kth bipolar bits of the pth image pattern, ξ^p, where $p = 5$. After obtaining the weights, the network will automatically associate perceived pattern, ξ, to a converged pattern, as $\bar{\xi}$. This converged pattern can be one of the trained patterns, the inverse of the one of the trained pattern or the combination of the trained patterns. We note that a converged pattern has the same size (20×20) of the perceived pattern.



(a) Camera image. (b) Stored pattern 1. (c) Stored pattern 2.

(d) Stored pattern 3. (e) Stored pattern 4. (f) Stored pattern 5.

Fig. 1: Camera image and stored patterns to construct an associative memory.

The *energy*, $E(\xi)$, required in order to obtain a converged pattern (i.e., a recalled pattern) is function of the activation switches of the network units was calculated as $E(\xi) = \sum_{i=1}^{N} \frac{|\xi_i - \bar{\xi}_i|}{2}$ where N is defined as the total number of bipolarized values of the perceived pattern. Due to the asynchronous activation of the units in the network, this is a lower-bound estimate of the actual number of switched activation. Therefore, it can be considered as the minimum amount of energy required to recall the image stored in memory.

B. Temporal Difference Learning: energy-reward association

To guide robot actions, composed of coordinated movements of the eyes and head, we adopt an on-policy temporal difference (TD) learning algorithm known as SARSA [12]. In this setting, the robot has no prior information about the expected reward values while moving from one state to the another by following ε-*greedy* policy, where ε is chosen to be 0.3. It should be emphasized that there are no predefined end states to terminate the experiment. We note that this experiment is designed to address whether the robot can find the states in which less amount of the computational energy needed to perform visual recalling task.

In Eq. 2, $Q(s, a)$ represents the current value of state-action pairs. Similarly, $Q(s', a')$ indicates the value for the action a' in the next state s'.

$$Q(s, a) \leftarrow Q(s, a) + \mu(R(s, s') + \gamma Q(s', a') - Q(s, a)) \tag{2}$$

The μ variable is the step size learning parameter, γ is an adjustment factor that discounts expected future rewards. The μ and γ variables are set to 0.7 and 0.4, respectively.

$$R(s,s') = \begin{cases} -1 & if\ E(\xi^{s}) < E(\xi^{s'}) \\ 1 & if\ E(\xi^{s}) \geq E(\xi^{s'}) \end{cases} \tag{3}$$

We extracted the reward value of a s, s' pair, $R(s, s')$, as a function of the computational energy consumed to process a visual pattern perceived in a state. To be more concrete, we derive the reward value of a s, s' pair based on Eq. 3. In this equation, ξ^{s} and $\xi^{s'}$ are the image patterns received in the states s and s' respectively, then the energy values for the execution of recalling operations, annotated by $E(\xi^{s})$ and $E(\xi^{s'})$, are obtained and compared. Based on this operation, the reward value is representing whether the agent moves from an higher energy state to lower energy one or vice versa.

III. Experimental setup

The experimental setup, shown in Figure 2, consists of the iCub humanoid robot and a screen which facilitates visual perception. We note that perceived images can be one of the memory patterns as shown in Figure 1, noisy version of the memory pattern or a completely new pattern. The robot perceives a visual stimulus, with a resolution of 640×480, trough a camera located its left eye. The robot explores the

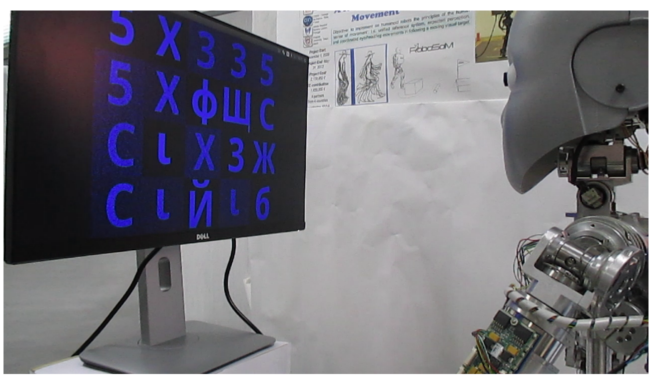
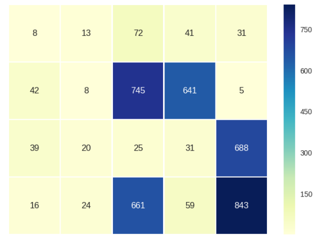


Fig. 2: Experiment setup: iCub humanoid robot and the constructed scene for perceptual processing.

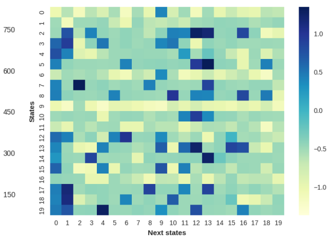
environment by visiting the states (i.e., directing its gaze from a discrete region in the scene to another) via coordinated head and eye movements [13], [14]. We highlight that the purpose of the robot experiments is to validate the proposed method in a real-world scenario in which hardware constraints (e.g., camera resolution) and environmental noise (e.g., reflections) exist.

IV. Results and Discussions

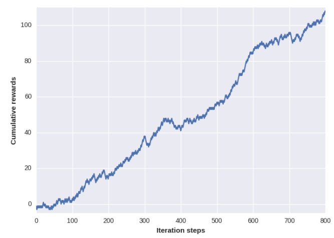
In this section, we report and discuss the obtained results by employing emotion mechanism on the iCub robot. In the conducted experiment, the temporal difference learning algorithm run for 800 iterations and in each iteration the robot receives a visual stimulus from the environment (i.e., a screen that located in front of the robot). In that, the robot forms the stimulus-energy association via performing HHOP, then employs this association to facilitate energy-reward derivation in order to learn how to act in the environment while minimizing the consumed computational energy for visual recalling. We depict subfigures in Figure 3 to interpret the robot's



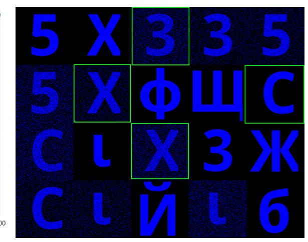
(a) Average energy per state



(b) Q matrix heat map



(c) Cumulative rewards



(d) Final states

Fig. 3: Experiment results for 800 iterations.

behavior during the experiment. The average computational energy consumed by the robot for a specific state shown as a heat map in Figure 3(a). In this figure, the constructed heat map can be considered as a zero-indexed matrix where each element corresponds to a state given in Figure 3(d) and the energy values depicted by a color scale (i.e., darker gradient refers to high energy state). In this way, it can be seen that the average computational cost of recalling a memory pattern will mostly require less amount of energy than recalling a noisy pattern or an unseen pattern. This stimulus-energy association can be observed by comparing the energy values of the first, sixth, and seventh states.

The values in this figure will be used to asses whether the behavior of the robot operates by performing energy minimization principle. We use the values in this heat map to determine how many times the robot chooses the correct actions. For instance, if the robot directs its gaze from a high energy state to the lower one, we assign the action as correct. In the inverse situation, we considered the action as wrong. Here, we report that the percentage of choosing the correct action by the robot is 85%.

To confirm that the robot learns to direct its gaze from the state that requires high energy consumption for visual recalling to the low one, we plot the cumulative reward curve in Figure 3(c). Based on the increment trend in this figure, we interpret that the robot learns the environmental dynamics by increasing the cumulative reward over iterations. Some fluctuations can be observed in this figure due to environmental noise and exploration rate of the SARSA. Lastly, Figure 3(d) shows the discovered final states are demonstrated with green rectangles. These states are extracted from the Q matrix. In that, the Q matrix elements are populated to determine the most valuable state-action pairs. The green-rectangled states indicate that regardless of an initial state in which robot directs its gaze, the robot will end up displaying a cyclic behavior that sequentially moves among the discovered states. To illustrate that discovered states are the states in which the robot consumes less amount of energy to perform visual recalling, we depict state transition diagram, in Figure 4, based on average energy values on Figure 3(a). Figure 4 presents

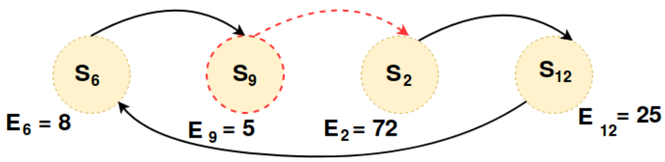


Fig. 4: State transition diagram of the final policy extracted from the Q matrix. S_9 contains a stored pattern while S_2, S_6, and S_{12} contain the noisy version of the stored patterns.

the state-action transition of the final policy. In this figure, each yellow circle indicates a state and the transition to move from one state to the another shown with the directed arrows. For each circle, the average energy that requires to perform visual recalling shown by E. To be more concrete, if the robot discovered the 2nd state (S_2), it would spend an average E_2 energy to process perceived stimuli then it will move to the 12th state, as S_{12}.

An onlooker might expect the robot should continuously gaze towards the S_9 as a final state. However, the external factors (such as operating in a real environment and being exposed to noisy sensory information) and internal factors (e.g., asynchronous update rule of the HHOP and exploration rate of the SARSA) of the experiment prevent the robot to perform this behavior. Our ongoing studies will address to eliminate this transition from the final state by analyzing external and internal factors.

On the basis of the presented results, we draw the following conclusions. Firstly, we show that the proposed methods enable the robot to perform the cognitive task by minimizing required computational energy. Secondly, the demonstrated behavior by the robot is non-trivial regarding stimulus-energy-reward associations. Lastly and more importantly, this non-deterministic behavior emerges from the robots internal mechanisms (e.g., associative memories and internal reward) while operating in an unknown environment.

V. Conclusions and Future Works

In this study, we present experimental results that test the proposed emotion model which employs the (neural) computational cost of visual recalling to extract an internally generated reward signal to guide actions (decisions) of the iCub robot. By doing so, the robot displays non-trivial behavior – that is, finding a set of available states in an unknown environment where less amount of computational energy needed to sustain the robot's life cycle. We hold that this regulatory role of the emergent emotion can be exploited further for the higher-order cognitive functions such as planning and decision making in the context of developmental robotics.

For the future studies, we will exploit the proposed model in a simple cognitive architecture that the robot performs complex cognitive tasks which require multimodal sensory processing for planning and decision making. Another research direction is to investigate the role of "emotional affinity" from an onlooker perspective. In detail, we would like to conduct the same experiment to inquire whether the robot shows affective response towards specific visual patterns.

References

[1] M. Ogino, T. Ooide, A. Watanabe, and M. Asada, "Acquiring peekaboo communication: Early communication model based on reward prediction," in *2007 IEEE 6th International Conference on Development and Learning*, July 2007, pp. 116–121.

[2] M. A. Arbib and J.-M. Fellous, "Emotions: from brain to robot," *Trends in cognitive sciences*, vol. 8, no. 12, pp. 554–561, 2004.

[3] M. Asada, K. Hosoda, Y. Kuniyoshi, H. Ishiguro, T. Inui, Y. Yoshikawa, M. Ogino, and C. Yoshida, "Cognitive developmental robotics: A survey," *IEEE Transactions on Autonomous Mental Development*, vol. 1, no. 1, pp. 12–34, May 2009.

[4] M. Kirtay and E. Oztop, "Emergent emotion via neural computational energy conservation on a humanoid robot," in *2013 13th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, Oct 2013, pp. 450–455.

[5] M. Kirtay, L. Vannucci, E. Falotico, E. Oztop, and C. Laschi, "Sequential decision making based on emergent emotion for a humanoid robot," in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, Nov 2016, pp. 1101–1106.

[6] T. Hashimoto, M. Senda, and H. Kobayashi, "Realization of realistic and rich facial expressions by face robot," in *IEEE Conference on Robotics and Automation, 2004. TExCRA Technical Exhibition Based.*, Nov 2004, pp. 37–38.

[7] J. Gratch, "Émile: Marshalling passions in training and education," in *Proceedings of the Fourth International Conference on Autonomous Agents*, ser. AGENTS '00. New York, NY, USA: ACM, 2000, pp. 325–332. [Online]. Available: http://doi.acm.org/10.1145/336595.337516

[8] M. Saerbeck and C. Bartneck, "Perception of affect elicited by robot motion," in *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*. IEEE Press, 2010, pp. 53–60.

[9] J. Lin, M. Spraragen, J. Blythe, and M. Zyda, "Emocog: Computational integration of emotion and cognitive architecture," in *Twenty-Fourth International FLAIRS Conference*, 2011.

[10] T. M. Moerland, J. Broekens, and C. M. Jonker, "Emotion in reinforcement learning agents and robots: a survey," *Machine Learning*, vol. 107, no. 2, pp. 443–480, Feb 2018. [Online]. Available: https://doi.org/10.1007/s10994-017-5666-0

[11] T. Chaminade, E. Oztop, G. Cheng, and M. Kawato, "From self-observation to imitation: Visuomotor association on a robotic hand," *Brain research bulletin*, vol. 75, no. 6, pp. 775–784, 2008.

[12] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press Cambridge, 1998, vol. 1, no. 1.

[13] L. Vannucci, N. Cauli, E. Falotico, A. Bernardino, and C. Laschi, "Adaptive visual pursuit involving eye-head coordination and prediction of the target motion," in *IEEE-RAS International Conference on Humanoid Robots*, 2014, pp. 541–546.

[14] L. Vannucci, E. Falotico, N. Di Lecce, P. Dario, and C. Laschi, "Integrating feedback and predictive control in a bio-inspired model of visual pursuit implemented on a humanoid robot," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9222, pp. 256–267, 2015.