

Artificial Curiosity Emerging Human-Like Behavior: Toward Fully Autonomous Cognitive Robots

Kurosh Madani, Christophe Sabourin and Dominik M. Ramík

Abstract This chapter is devoted to autonomous cognitive machines by mean of the design of an artificial curiosity based cognitive system for autonomous high-level knowledge acquisition from visual information. Playing a chief role as well in visual attention as in interactive high-level knowledge construction, the artificial curiosity is realized through combining visual saliency detection and Machine-Learning based approaches. Experimental results validating the deployment of the investigated system have been obtained using as well simulation facilities as a real humanoid robot acquiring visually knowledge about its surrounding environment interacting with a human tutor. As show the reported results and experiments, the proposed cognitive system allows the machine to discover autonomously the surrounding world in which it may evolve, to learn new knowledge about it and to describe it using human-like natural utterances.

1 Introduction and Problem Stating

If nowadays machines and robotic bodies are fully automated outperforming human capacities, nonetheless, none of them can be called truly intelligent or pretend defeating human's cognitive skills. The fact that human-like machine-cognition is still beyond the reach of contemporary science only proves how difficult the problem is. Somewhat, it is due to the fact that the science is still far from fully understanding the human cognitive system. On the other hand, it is so because if contemporary machines are often fully automatic, they linger rarely fully autonomous in their

K. Madani (✉) · C. Sabourin · D.M. Ramík
Images, Signals and Intelligence Systems Laboratory (LISSI/EA 3956),
University PARIS-EST Creteil (UPEC), Senart-FB Institute of Technology,
Bât.A, Av. Pierre Point, 77127 Lieusaint, France
e-mail: madani@univ-paris12.fr

C. Sabourin
e-mail: sabourin@univ-paris12.fr

D.M. Ramík
e-mail: dominik.ramik@univ-paris12.fr

© Springer International Publishing Switzerland 2016
K. Madani et al. (eds.), *Computational Intelligence*,
Studies in Computational Intelligence 613,
DOI 10.1007/978-3-319-23392-5_28

knowledge acquisition. Nevertheless, the concepts of bio-inspired or human-like machine-cognition remain foremost sources of inspiration for achieving intelligent systems (intelligent machines, intelligent robots, etc...).

Emergence of cognitive phenomena in machines has been and remains active part of research efforts since the rise of Artificial Intelligence (AI) in the middle of the last century. Among others, [1] provides a survey on cognitive systems. It accounts on different paradigms of cognition in artificial agents markedly on the contrast of emergent versus cognitivist paradigms and on their hybrid combinations. It is also worth of mentioning the work of [2] which brings an in-depth review on a number of existing cognitive architectures such those which adheres to the symbolic theory and reposes on the assumption that human knowledge can be divided to two kinds: declarative and procedural. Another discussed architecture belongs to class of those using “If-Then” deductive rules dividing knowledge again on two kinds: concepts and skills. In contrast to above-mentioned works, the work of [3] focuses the area of research on cognition and cognitive robots discussing purposes linking knowledge representation, sensing and reasoning in cognitive robots. However, there is no cognition without perception (a cognitive system without the capacity to perceive would miss the link to the real world and so it would be impaired) and thus autonomous acquisition of knowledge from perception is a problem that should not be skipped when dealing with cognitive systems.

Prominently to the machine-cognition’s issue is the question: “what is the compel or the motivation for a cognitive system to acquire new knowledge?” For human cognitive system Berlyne states, that it is the curiosity that is the motor of seeking for new knowledge [4]. Consequently a few works have been since there dedicated to incorporation of curiosity into a number of artificial systems including embodied agents or robots. However the number of works using some kind of curiosity motivated knowledge acquisition with implementation to real agents (robots) is still relatively small. Often authors view curiosity only as an auxiliary mechanism in robot’s exploration behavior. One of early implementations of artificial curiosity may be found in [5]. Accordingly to the author, the introduction of curiosity further helps the system to actively seek similar situations in order to learn more. On the field of developmental and cognitive robotics a similar approach may be found in [6] where authors present an approach including a mechanism called “Intelligent Adaptive Curiosity”. Two experiments with AIBO robot are presented showing that the curiosity mechanism successfully stimulates the learning progress. In a recent publication, authors of [7] implement the psychological notion of surprise-curiosity into the decision making process of an agent exploring an unknown environment. Authors conclude that the surprise-curiosity driven strategy outperformed classical exploration strategy regarding the time-energy consumed in exploring the delved environment. On the other hand, the concept of surprise, relating closely the notion of curiosity, has been exploited in [8] by a robot using the surprise in order to discover new objects and acquire their visual representations. Finally, the concept of curiosity has been successfully used in [9] for learning affordances of a mobile robot in navigation task. The mentioned works are attempting to respond the question: “how

an autonomous cognitive system should be designed in order to exhibit the behavior and functionality close to its human users”.

That is why even though *curiosity killed a cat*,¹ taking into consideration the enticing benefits of curiosity, we have made it our principle foundation in investigated concept. The present paper is devoted to the description of a cognitive system based on artificial curiosity for high-level human-like knowledge acquisition from visual information. The goal of the investigated system is to allow the machine (such as a humanoid robot) to observe, to learn and to interpret the world in which it evolves, using appropriate terms from human language, while not making use of a priori knowledge. This is done by word-meaning anchoring based on learning by observation stimulated (steered) by artificial curiosity and by interaction with the human. Our model is closely inspired by juvenile learning behavior of human infants [10, 11].

In Sect. 2, we detail our approach by outlining its architecture and principles. We explain how the machine generates its beliefs about the world from observing the surrounding environment and the role of human-robot interaction in the learning process. Section 3 focuses the validation of the proposed approach using as well simulation facilities as a real robot evolving in real environment. Finally Sect. 4 discusses the achieved results and outlines the future work.

2 Brief Overview of Multi-level Cognitive Concept

Accordingly to Berlyne’s theory of human curiosity [4], two cognitive levels contribute to human’s desire of acquiring new knowledge. The first is so-called “perceptual curiosity”, which leads to increased perception of stimuli. It is a lower level cognitive function, more related to perception of new, surprising or unusual sensory input. It contrasts to repetitive or monotonous perceptual experience. The other one is called “epistemic curiosity”, which is more related to the “desire for knowledge that motivates individuals to learn new ideas, eliminate information-gaps, and solve intellectual problems” [12]. It also seems that it acts to stimulate long-term memory in remembering new or surprising (e.g. what may be contrasting with already learned) information [13]. By observing the state of the art (including the referenced ones), it may be concluded that the curiosity is usually used as an auxiliary mechanism instead of being the fundamental basis of the knowledge acquisition. To our best knowledge there is no work to date which considers curiosity in context of machine cognition as a drive for knowledge acquisition on both low (perceptual) level and high (“semantic”) level of the system. Without striving for biological plausibility whilst by analogy with natural curiosity, we founded our system on two cognitive levels ([14, 15]). Depicted in Fig. 1, the first ahead of reflexive visual attention plays the role

¹In ‘Different’, Eugene O’Neill, 1920: BENNY—(with a wink): “*Curiosity killed a cat! Ask me no questions and I’ll tell you no lies.*”

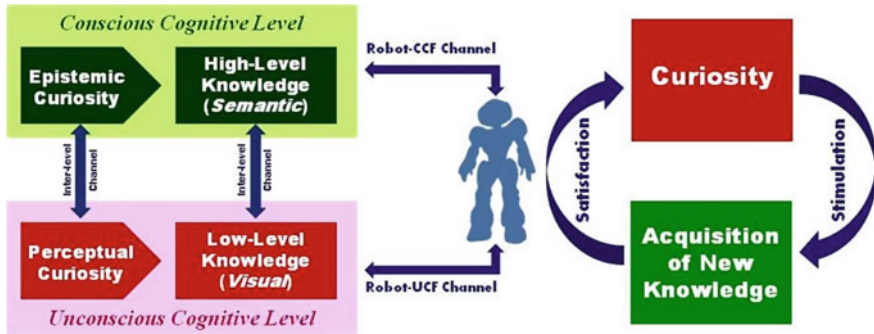


Fig. 1 General bloc-diagram of the proposed curiosity driven architecture (*left*) and principle of curiosity-based Stimulation-Satisfaction mechanism for knowledge acquisition (*right*)

of perceptual curiosity and the second coping with intentional learning-by-interaction undertakes the role of epistemic curiosity.

2.1 From Observation to Interpretation

The problem of autonomous learning conveys the inbuilt problem of distinguishing the pertinent sensory information from the impertinent one. The solution to this task is natural for human, it remain very far from being obvious for a robot. In fact, when a human points to one object among many others giving a description of that pointed object using his human natural language, the robot still has to distinguish, which of the detected features and perceived characteristics of the object the human is referring to. To achieve correct anchoring, the proposed architecture adopts the following strategy. By using its perceptual curiosity, realized thanks to artificial salient vision and adaptive visual attention (described in [16–18]), the robot extracts features from important objects found in the scene along with the words the human used to describe the objects. Then, the robot generates its beliefs about which words could describe which features. Using the generated beliefs as organisms in a genetic algorithm, the robot determines its “most coherent belief”. To calculate the fitness, a classifier is trained and used to interpret the objects the robot has already seen. The utterances pronounced by the human for each object are compared with those the robot would use to describe it based on its current belief. The closer the robot’s description is to that given by the human, the higher the fitness is. Once the evolution has been finished, the belief with the highest fitness is adopted by the robot and is used to interpret occurrences of new (unseen) objects. Figure 2 depicts through an example important parts and operations of the proposed system.

Let us suppose a robot equipped by a sensor observing the surrounding world and interacting with the human. The world is represented as a set of features $I = \{i_1, i_2, \dots, i_k\}$, which can be acquired by robot’s sensor. Each time the robot makes

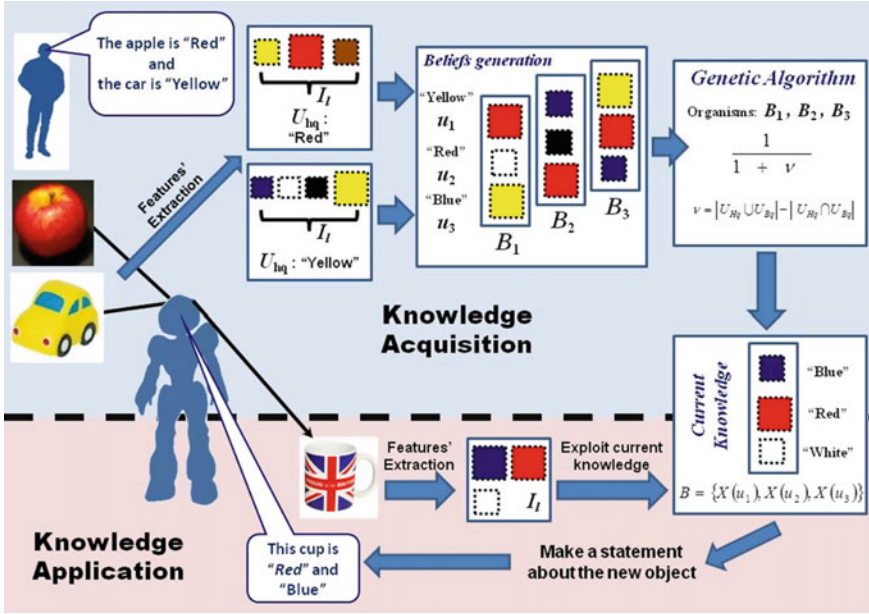


Fig. 2 Example showing main parts of the system’s operation in the case of autonomous learning of colors

an observation o , its epistemic curiosity stimulates it to interact with the human asking him to give a set of utterances U_H describing the found salient objects. Let us denote the set of all utterances ever given about the world as U . The observation o is defined as an ordered pair $o = \{I_l, U_H\}$, where $I_l \subseteq I$, expressed by (1), stands for the set of features obtained from observation and $U_H \subseteq U$ is the set of utterances (describing O) given by human in the context of that observation. i_p denotes the pertinent information for a given u (i.e. features that can be described semantically as u in the language used for communication between the human and the robot), i_i the impertinent information (i.e. features that are not described by the given u , but might be described by another $u_i \in U$) and sensor noise ε . The goal is to distinguish the pertinent information from the impertinent one and to correctly map the utterances to appropriate perceived stimuli (features). Let us define an interpretation $X(u) = \{u, I_j\}$ of an utterance u as an ordered pair where $I_j \subseteq I$ is a set of features from I . So, the belief B is defined accordingly to (2) as an ordered set of interpreting utterances u from U .

$$I_l = \bigcup_{U_H} i_p(u) + \bigcup_{U_H} i_i(u) + \varepsilon \tag{1}$$

$$B = \{X(u_1), \dots, X(u_n)\} \text{ with } n = |U| \tag{2}$$

Accordingly to the criterion expressed by (3), one can calculate the belief B which interprets coherently the observations made so far: in other words, by looking for such a belief, which minimizes across all the observations $o_q \in O$ the difference between the utterances U_{Hq} made by human, and those utterances U_{Bq} , made by the system by using the belief B . Thus, B is a mapping from the set U to I : all members of U map to one or more members of I and no two members of U map to the same member of I .

$$\arg \min_B \left(\sum_{q=1}^{|O|} |U_{Hq} - U_{Bq}| \right) \quad (3)$$

2.2 The Most Coherent Interpretation Search

Although the interpretation's coherence is worked out by computing the belief B accordingly to Eq. (3), the system has to look for a belief B , which would make the robot describing a particular scene with utterances as close and as coherent as possible to those that a human would made on the same (or similar) scene. For this purpose, instead performing the exhaustive search over all possible beliefs, we propose to search for a suboptimal belief by means of a Genetic Algorithm (GA). For doing that, we assume that each organism within it has its genome constituted by a belief, which, results into genomes of equal size $|U|$ containing interpretations $X(u)$ of all utterances from U .

In our genetic algorithm, the genomes' generation is a belief generation process generating genomes (e.g. beliefs) as follows. For each interpretation $X(u)$ the process explores whole the set O . For each observation $o_q \in O$, if $u \in U_{Hq}$ then features $i_q \in I_q$ (with $I_q \subseteq I$) are extracted. As described in (1), the extracted set of features contains as well pertinent as impertinent features. The coherent belief generation is done by deciding, which features $i_q \in I_q$ may possibly be the pertinent ones. The decision is driven by two principles. The first one is the principle of "proximity", stating that any feature i is more likely to be selected as pertinent in the context of u , if its distance to other already selected features is comparatively small. The second principle is the "coherence" with all the observations in O . This means, that any observation $o_q \in O$, corresponding to $u \in U_{Hq}$, has to have at least one feature assigned into I_q of the current $X(u) = \{u, I_q\}$.

To evaluate a given organism, a classifier is trained, whose classes are the utterances from U and the training data for each class $u \in U$ are those corresponding to $X(u) = \{u, I_q\}$, i.e. the features associated with the given u in the genome. This classifier is used through whole set O of observations, classifying utterances $u \in U$ describing each $o_q \in O$ accordingly to its extracted features. Such a classification results in the set of utterances U_{Bq} (meaning that a belief B is tested regarding the q^{th} observation). The fitness function evaluating the fitness of each above-mentioned organism is defined as "disparity" between U_{Bq} and U_{Hq} (defined in previous

subsection) which is computed accordingly to the Eq.(4), where v is the number of utterances that are not present in both sets U_{Bq} and U_{Hq} (e.g. either missed or are superfluous utterances interpreting the given features). The globally best fitting organism is chosen as the belief that best explains observations O made (by robot).

$$D(v) = \frac{1}{1+v} \quad \text{with } v = |U_{Hq} \cup U_{Bq}| - |U_{Hq} \cap U_{Bq}| \quad (4)$$

Figure 3 gives the bloc diagram of the designed evolutionary process. It is important to note that here the above-described GA based evolutionary process doesn't operate as only an optimizer but it generates the machines (e.g. robot's) most coherent belief about the observation accomplished by this robot and about the way that the same robot will autonomously construct a human-like description of the observed reality. In other words, it is the GA based evolutionary process that drives the robot's most coherent semantic understanding of the observed reality. It plays also a key role in implementation of the epistemic curiosity because the drop of the search for the most coherent belief, due to leakage of knowledge about the observed reality, makes the robot interacting with its human counterpart and thus drives its epistemic curiosity.

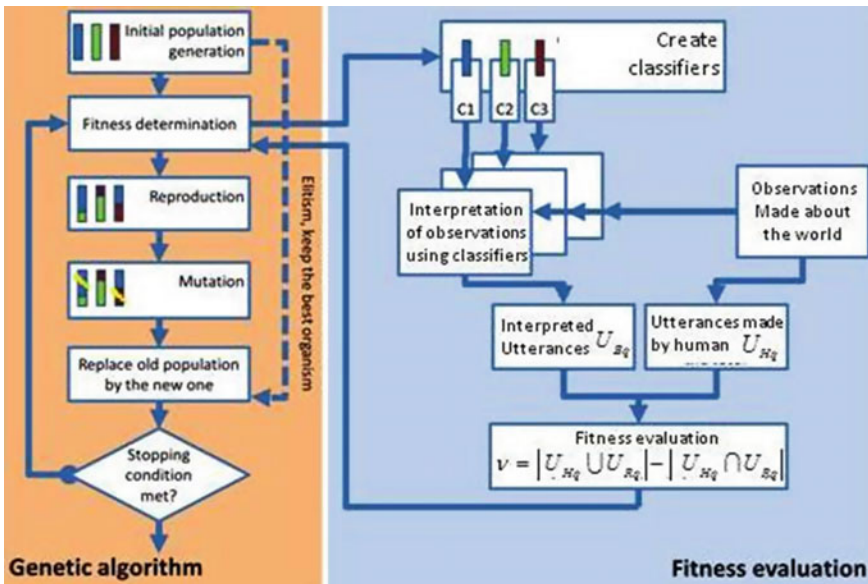


Fig. 3 Bloc diagram of described genetic algorithm's workflow. The left part describes the genetic algorithm itself, while the right part focuses on the fitness evaluation workflow

2.3 Role of Human-Robot Interaction

Human beings learn both by observation and by interaction with the world and with other human beings. The former is captured in our system in the “best interpretation search” outlined previous subsections. The latter type of learning requires that the robot be able to communicate with its environment and is facilitated by learning by observation, which may serve as its bootstrap. In our approach, the learning by interaction is carried out in two kinds of interactions: human-to-robot and robot-to-human. The human-to-robot interaction is activated anytime the robot interprets wrongly the world. When the human receives a wrong response (from robot), he provides the robot a new observation by uttering the desired interpretation. The robot takes this new corrective knowledge about the world into account and searches for a new interpretation of the world conformably to this new observation. The robot-to-human interaction may be activated when the robot attempts to interpret a particular feature classified with a very low confidence: a sign that this feature is a borderline example. In this case, it may be beneficial to clarify its true nature. Thus, led by the epistemic curiosity, the robot asks its human counterpart to make an utterance about the uncertain observation. If the robot’s interpretation is not conforming to the utterance given by the human (robot’s interpretation was wrong), this observation is recorded as a new knowledge and a search for the new interpretation is started.

3 Implementation and Validation Results

The validation of the proposed system has been performed on the basis of both simulation of the designed system as by an implementation on a real humanoid robot. A video capturing different parts of the experiment may be found online on: <http://youtu.be/W5FD6 zXihOo>. As real robot we have considered NAO robot (a small humanoid robot from Aldebaran Robotics) which provides a number of facilities such as onboard camera (vision), communication devices and onboard speech generator. The fact that the above-mentioned facilities been already available offers a huge save of time, even if those faculties remain quite basic in that kind of robots.

Although the usage of the presented system is not specifically bound to humanoid robots, it is pertinent to state two main reasons why a humanoid robot is used for the system’s validation. The first reason for this is that from the definition of the term “humanoid”, a humanoid robot is aspired to make its perception close to the human’s one, entailing a more human-like experience of the world. This is an important aspect to be considered in context of sharing knowledge between a human and a robot. The second reason is that humanoid robots are specifically designed to interact with humans in a “natural” way by using e.g. a loudspeaker and microphone set in order to allow for a bi-directional communication with human by speech synthesis and speech analysis and recognition. This is of importance when speaking about a natural human-robot interaction during learning.

3.1 Simulation Based Validation and Results

The simulation based validation finds its pertinence in assessment of the investigated cognitive-system’s performances. In fact, due to difficulties inherent to organization of strictly same experimental protocols on different real robots and within various realistic contexts, the simulated validation becomes an appealing way to ensure that the protocol remains the same. For simulation based evaluation of the behavior of the above-described system, we have considered color names learning problem. In everyday dialogs, people tend to describe objects, which they see, with only a few color terms (usually only one or two), although the objects in itself contains many more colors. Also different people can have slightly different preferences on what names to use for which color. Due to this, learning color names is a difficult task and it is a relevant sample problem to test our system.

In the simulated environment, images of real-world objects were presented to the system alongside with textual tags describing colors present on each object. The images were taken from the Columbia Object Image Library (COIL) contains 1000 color images of different views of 100 objects database. Five fluent English speakers were asked to describe each object in terms of colors. We restricted the choice of colors to “Black”, “Gray”, “White”, “Red”, “Green”, “Blue” and “Yellow”, based on the color opponent process theory [19] (Schindler 1964). The tagging of the entire set of images was highly coherent across the subjects. In each run of the experiment, we have randomly chosen a tagged set.

The utterances were given in the form of text extracted from the descriptions. The object was accepted as correctly interpreted if the system’s and the human’s interpretations were equal. The rate of correctly described objects from the test set was approximately 91 %. Figure 4 gives the result of interpretation by the system of the colors of the WCS table. Figure 5 shows the learning rate versus the increasing number of exposures of each color.

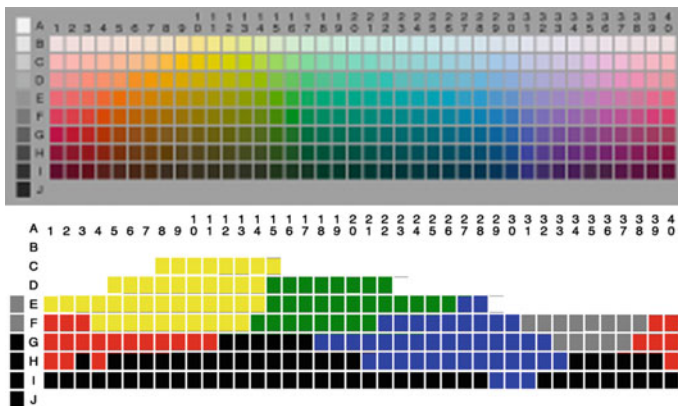
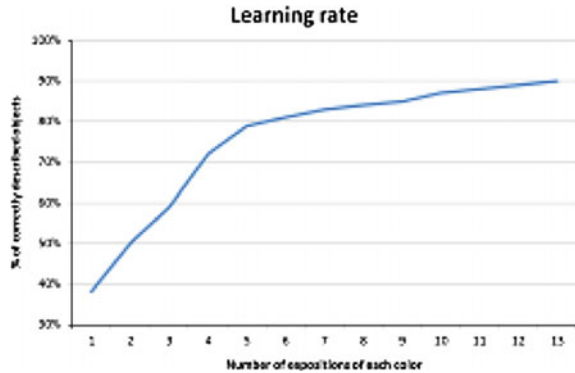


Fig. 4 Original WCS table (upper image), its system’s made interpretation (lower image)

Fig. 5 The rate of correct learning versus the number of presented examples (of the same object) to the system



3.2 Implementation on Real Robot

The designed system has been implemented on NAO robot (from Aldebaran Robotics). It is a small humanoid robot which provides a number of facilities such as onboard camera (vision), communication devices and onboard speech generator. The fact that the above-mentioned facilities are already available offers a huge save of time, even if those faculties remain quite basic in that kind of robots. If NAO robot integrates an onboard speech-recognition algorithm (e.g. some kind of speech-to-text converter) which is sufficient for “hearing” the tutor, however its onboard speech generator is a basic text-to-speech converter. It is not sufficient to allow the tutor and the robot conversing in natural speech. To overcome NAO’s limitations relating this purpose, the TreeTagger tool² was used in combination with robot’s speech-recognition system to obtain the part-of-speech information from situated dialogs. Standard English grammar rules were used to determine whether the sentence is demonstrative (e.g. for example: “This is an apple.”), descriptive (e.g. for example: “The apple is red.”) or an order (e.g. for example: “Describe this thing!”). To communicate with the tutor, the robot used its text-to-speech engine.

The core of the implementation’s architecture is split into five main units: Communication Unit (CU), Navigation Unit (NU), Low-level Knowledge Acquisition Unit (LKAU), High-level Knowledge Acquisition Unit (HLAU) and Behavior Control Unit (BCU). Figure 6 illustrates the bloc-diagram of the implementation’s architecture. The aforementioned units control NAO robot (symbolized by its sensors, its actuators and its interfaces in Fig. 6) through its already available hardware and software facilities. In other words, the above-mentioned architecture controls the whole robot’s behavior.

The purpose of NU is to allow the robot to position itself in space with respect to objects around it and to use this knowledge to navigate within the surrounding environment. Capacities needed in this context are obstacle avoidance and

² Developed by the ICL at University of Stuttgart, available online at: <http://www.ims.uni-stuttgart.de/projekte/complex/TreeTagger>.

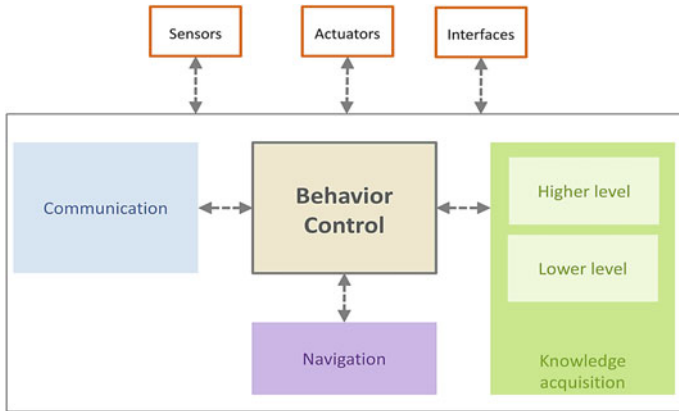


Fig. 6 Block diagram of the implementation’s architecture

determination of distance to objects. Its sub-unit handling spatial orientation receives its inputs from the camera and from the LKAU. To get to the bottom of the obstacle avoidance problem, we have adopted a technique based on ground color modeling. Inspired by the work presented in [20], color model of the ground helps the robot to distinguish free-space from obstacles.

The LKAU ensures gathering of visual knowledge, such as detection of salient objects and their learning (by the sub-unit in charge of salient object detection) and sub-recognition (see [18, 21]). Those activities are carried out mostly in an “unconscious” manner, i.e. they are run as an automatism in “background” while collecting salient objects and learning them. The learned knowledge is stored in Long-term Memory for further use.

The HKAU is the center where the intellectual behavior of the robot is constructed. Receiving its features from the LKAU (visual features) and from the CU (linguistic features), this unit processes the beliefs’ generation, the most coherent belief’s emergence and constructs the high-level semantic representation of acquired visual knowledge. Unlike the LKAU, this unit represents conscious and intentional cognitive activity. In some way, it operates as a baby who learns from observation and from verbal interaction with adults about what he observes developing in this way his own representation and his own opinion about the observed world [22].

The CU is in charge of robots communication. It includes an output communication channel and an input communication channel. The output channel is composed of a Text-To-Speech engine which generates human voice through loud-speakers. It receives the text from the BCU. The input channel takes its input from a microphone and through an Automated Speech Recognition engine (available in NAO) the syntax and semantic analysis (designed and incorporated in BCU) it provides the BCU labeled chain of strings representing the heard speech.

The BCU plays the role of a coordinator of robot’s behavior. It handles data flows and issues command signals for other units, controlling the behavior of the robot and



Fig. 7 Robot and the subset of collected objects used for learning (*left-side picture*) and the recognition by robot of two objects among those learned by the robot in different posture and in different location (*right-side picture*)

its suitable reactions to external events (including its interaction with humans). BCU received its inputs from all other units and returns its outputs to each concerned unit including robot's devices (e.g. sensors, actuators and interfaces) [22]. The human-robot interaction is performed by this unit in cooperation with HLAU. In other words, driven by HLAU, a part of the robot's epistemic curiosity based behavior is handled by BCU.

3.3 Experimental Validation

The total of 25 every-day objects was collected for purposes of the experiment. The collected set has been randomly divided into two sets for training and for testing (Fig. 7). The learning set objects were placed around the robot and then a human tutor pointed to each of them calling it by its name. Using its 640×480 monocular color camera, the robot discovered and learned the objects around it by the salient object detection approach we have described in [18]. Here, this approach has been extended by detecting the movement of the human's hand to achieve joint attention. In this way, the robot was able to determine what object the tutor is referring to and to learn its name. The right-side picture in Fig. 7 shows the recognition by robot of two objects among those learned by the robot in different posture and in different location.

During the experiment, the robot has been asked to learn a subset among the 25 considered objects: in term of associating the name of each detected object to that object. At the same time, a second learning has been performed involving the interaction with the tutor who has successively pointed the above-learned objects describing (e.g. telling) to the robot the color of each object. Extracted from the video of the experimental validation, Fig. 8 shows the robot observing and learning



Fig. 8 Experimental setup showing tutor pointing different objects from learning set and robot learning those objects

different objects chosen by the human tutor. Here-bellow an example of the Human-Robot interactive learning is reported:

- **Human** [*pointing a red aid-kit*]: “This is a first-aid-kit!”
- **Robot**: “I will remember that this is a first-aid-kit.”
- **Human**: “It is red and white”.
- **Robot**: “OK, the first-aid-kit is red and the white.”

After learning the names and colors of the discovered objects, the robot is asked to describe a number of objects including as well some of already learned objects but in different posture (for example the yellow box presented in reverse posture) as a number of still unseen objects (as for example a red apple or a white teddy-bear). The robot has successfully described, in a coherent linguistics, the presented seen and unseen objects. Extracted from the video of the experimental validation, Fig. 9 shows the human tutor asking the robot to describe the pointed object (which is a red apple) in term of colors (left-side picture of Fig. 9) and the ground truth detected objects as the robot perceives them. Finally, Fig. 10 shows two examples of observed objects’ interpretation by the robot. Here-bellow is the Human-Robot interaction during the experiment:

- **Human** [*pointing the unseen white teddy-bear*]: “Describe this!”
- **Robot**: “It is white!”
- **Human**: [*pointing the already seen, but reversed, yellow box*]: “Describe this!”
- **Robot**: “It is yellow!”
- **Human**: [*pointing the unseen apple*]: “Describe this!”
- **Robot**: “It is red!”



Fig. 9 Experimental setup showing tutor pointing a red apple which has not been seen before, (by the robot) asking the robot to describe that object in term of colors (*left-side picture*) and the ground truth detected objects as the robot perceives them (*right-side picture*)



Fig. 10 Two objects observed and interpreted by the robot: the original image provided by robot's camera (*left-side pictures*) and the interpretation of those objects by the robot (*right-side pictures*). For the "apple", the robot's given description was "the object is red". For the box, the description was "the object is blue and white"

4 Conclusion and Perspectives

This chapter has presented, discussed and validated a cognitive system for high-level knowledge acquisition based on the notion of artificial curiosity. Driving as well the lower as the higher levels of the presented cognitive system, the emergent artificial curiosity allow such a system to learn in an autonomous manner new knowledge about unknown surrounding world and to complete (enrich or correct) its knowledge by interacting with a human. Experimental results, performed as well on a simulation platform as using the NAO robot show the pertinence of the investigated concepts as well as the effectiveness of the designed system. Although it is difficult to make a precise comparison due to different experimental protocols, the results we obtained

show that our system is able to learn faster and from significantly fewer examples, than the most of more-or-less similar implementations.

Based on obtained results, it is thus justified to say, that a robot endowed with such artificial curiosity based intelligence will necessarily include autonomous cognitive capabilities. With respect to this, several appealing perspectives are pursuing to push further the presented work. The current implemented version allows the robot to work with a single category or property at a time (e.g. for example the color in utterances like “it is red”). We are working on extending its ability to allow the learning of multiple categories at the same time and to distinguish which of the used words are related to which category. While, concerning the middle-term perspectives of this work, they will focus aspects reinforcing the autonomy of such cognitive robots. The ambition here is integration of the designed system to a system of larger capabilities realizing multi-sensor artificial machine-intelligence. There, it will play the role of an underlying part for machine cognition and knowledge acquisition.

References

1. Vernon, D., Metta, G., Sandini, G.: A survey of artificial cognitive systems: implications for the autonomous development of mental capabilities in computational agents. *IEEE Trans. Evol. Comput.* **11**(2), 151–180 (2007)
2. Langley, P., Laird, J.E., Rogers, S.: Cognitive architectures: research issues and challenges. *Cogn. Syst. Res.* **10**(2), 141–160 (2009)
3. Levesque, H.J., Lakemeyer, G.: “Cognitive robotics”, *Handbook of Knowledge Representation*, Dagstuhl: Schloss Dagstuhl—Leibniz-Zentrum fuer Informatik (2010)
4. Berlyne, D.E.: A theory of human curiosity. *British J. Psychol.* **45**(3), 180–191 (1954)
5. Schmidhuber, J.: Curious model-building control systems. In: *Proceedings of International Joint Conference on Neural Networks (IEEE-IJCNN 1991)*, Vol. 2, pp. 1458–1463 (1991)
6. Oudeyer, P.-Y., Kaplan, F., Hafner, V.V.: Intrinsic motivation systems for autonomous mental development. *IEEE Trans. Evol. Comput.* **11**(2), 265–286 (2007)
7. Macedo, L., Cardoso, A.: The exploration of unknown environments populated with entities by a surprise-curiosity-based agent. *Cogn. Syst. Res.* **19–20**, 62–87 (2012)
8. Maier, W., Steinbach, E.G.: Surprise-driven acquisition of visual object representations for cognitive mobile robots, pp. 1621–1626. In: *Proceedings of IEEE International Conference on Robotics and Automation*, Shanghai (2011)
9. Ugur, E., Dogar, M.R., Cakmak, M., Sahin, E.: Curiosity-driven learning of traversability affordance on a mobile robot. In: *Proceedings of IEEE 6th International Conference on Development and Learning*, pp. 13–18 (2007)
10. Yu, C.: The emergence of links between lexical acquisition and object categorization: a computational study. *Connection Sci.* **17**(3–4), 381–397 (2005)
11. Waxman, S.R., Gelman, S.A.: Early word-learning entails reference, not merely associations. *Trends Cogn. Sci.* (2009)
12. Litman, J.A.: Interest and deprivation factors of epistemic curiosity. *Personality Individ. Differ.* **44**(7), 1585–1595 (2008)
13. Kang, M.J.J., Hsu, M., Krajbich, I.M., Loewenstein, G., McClure, S.M., Wang, J.T.T., Camerer, C.F.: The wick in the candle of learning: epistemic curiosity activates reward circuitry and enhances memory. *Psychol. Sci.* **20**(8), 963–973 (2009)
14. Madani, K., Sabourin, C.: Multi-level cognitive machine-learning based concept for human-like artificial walking: application to autonomous stroll of humanoid robots. *Neurocomputing*, 1213–1228 (2011)

15. Ramik, D.-M., Sabourin, C., Madani, K.: From visual patterns to semantic description: a cognitive approach using artificial curiosity as the foundation. *Pattern Recogn. Lett.* **34**(14), 1577–1588 (2013). Elsevier
16. Ramík, D.M., Sabourin, C., Madani, K.: A real-time robot vision approach combining visual saliency and unsupervised learning. In: *Proceedings of 14th International Conference on CLAWAR, Paris, France*, pp. 241–248 (2011)
17. Ramík, D.M., Sabourin, C., Madani, K.: Hybrid salient object extraction approach with automatic estimation of visual attention scale. In: *Proceedings of Seventh International Conference on Signal Image Technology & Internet-Based Systems, Dijon, France*, pp. 438–445 (2011)
18. Ramik, D.M., Sabourin, C., Moreno, R., Madani, K.: A machine learning based intelligent vision system for autonomous object detection and recognition. *J. Appl. Intell.* (2013). Springer, doi:[10.1007/s10489-013-0461-5](https://doi.org/10.1007/s10489-013-0461-5)
19. Schindler, M., Von Goethe, J.W.: *Goethe's theory of colour applied by Maria Schindler*. New Knowledge Books, East Grinstead, Eng (1964)
20. Moreno, R., Ramik, D.M., Graña, M., Madani, K.: Image segmentation on the spherical coordinate representation of the RGB color space. *IET Image Process.* **6**(9), 1275–1283 (2012)
21. Madani, K., Sabourin, C.: Multi-level cognitive machine-learning based concept for human-like "artificial" walking: application to autonomous stroll of humanoid robots. *Neurocomputing* **74**, 1213–1228 (2011)
22. Ramik, D.M., Sabourin, C., Madani, K.: Autonomous knowledge acquisition based on artificial curiosity: application to mobile robots in indoor environment. *J. Robot. Auton. Syst.* **61**(12), 1680–1695 (2013)