Perception and human interaction for developmental learning of objects and affordances

Serena Ivaldi, Natalya Lyubova, Damien Gérardeaux-Viret, Alain Droniou, Salvatore Anzalone, Mohamed Chetouani, David Filliat, Olivier Sigaud

Perception and human interaction for developmental learning of objects and affordances

Serena Ivaldi[1], Natalia Lyubova[2], Damien Gérardeaux-Viret[2], Alain Droniou[1], Salvatore M. Anzalone[1], Mohamed Chetouani[1], David Filliat[2], Olivier Sigaud[1]

***Abstract*— In this paper we describe a cognitive architecture for humanoids interacting with objects and caregivers in a developmental robotics scenario. The architecture is foundational to the MACSi project: it is designed to support experiments to make a humanoid robot gradually enlarge its repertoire of known objects and skills combining autonomous learning, social guidance and intrinsic motivation. This complex learning process requires the capability to learn affordances first. Here, we present the general framework for achieving these goals, focusing on the elementary action, perception and interaction modules. Preliminary experiments performed on the humanoid robot iCub are also discussed.**

I. INTRODUCTION

Pre-programming behaviors exploiting *a priori* knowledge, as it has been done so far for industrial robots, is not a viable *modus operandi* for personal and service robots coming into everyday life [1]. It is imperative to give robots the ability to acquire new skills to cope with the evolving environment, different partners and a possibly wide variety of tools. Epigenetic or developmental robotics addresses this problem by taking inspiration from the cognitive development of children [2]. Indeed, one striking characteristic of children is the nearly open-ended diversity of the skills they are capable of learning, driven both by internal motives and by social guidance. In a similar manner, the robot may be endowed with learning capabilities and a set of basic primitives for motion, perception and interaction. Then it may gradually enlarge its knowledge of the environment (e.g. objects) and repertoire of skills, through learning and exploration [3]. The robot may therefore find problem-specific solutions autonomously or with a minimal intervention from humans, adapting its behavior on the fly to experienced circumstances. The human acting as a "caregiver" may have a catalytic effect on the learning process, from the bootstrapping to the whole exploratory phase.

A fruitful approach to the acquisition of new skills is the learning of *affordances*. Without going into the intricacies of this notion borrowed from developmental psychology [4], an affordance is intuitively the ability of an object to produce a certain effect or realize a certain goal as a consequence of an action performed with it. Indeed, affordances are generally built as Action-Object-Effect (AOE) complexes, and modeled as inter-relations between the elements, such that one can be predicted given the other two. Their knowledge may, for example, help in choosing an object for a specific task (find O, given A and E), or predict the outcome of a task (infer E, given A and O). In robotics, affordances are used to improve object recognition, categorization, evaluate possible actions and learn new skills [5]–[10]. A typical scenario for affordance learning consists in having the robot surrounded by several objects at reachable distance (for example on a table, see Fig. 1). The robot is then instructed by a caregiver to manipulate those objects so as to identify the outcome of different actions on the objects.

Fig. 1. The humanoid robot iCub and the experimental context.

The design of such complex experiments, where humanoid robots interact with caregivers and tools, necessarily requires a considerable effort. The puzzle has several pieces, like the edification of the basic perceptual and motor primitives of the robot, the choice of an informative representation of the robot state, the correct interpretation of the human intent, *etc*. To provide modularity, these functionalities are then implemented in several software modules, which are typically integrated and executed concurrently on the robotic platform.

Several Cognitive Architectures (CAs) have been recently proposed to answer these needs. For example in [11] the CA is focused on interaction and emotional status, whereas in [12] it focuses on cooperation and shared plans execution. In particular, the latter is an interesting modular and scalable solution, where all perceptual, motor and learning modules are abstracted from the hardware and exchange data according to predefined protocols.

Here, we take inspiration from this work, but we propose a novel architecture which is specifically focused on affordance learning experiments in a developmental robotics context. The realization of such a framework is indeed

[1] Institut des Systèmes Intelligents et de Robotique, CNRS UMR 7222 & Université Pierre et Marie Curie, Paris, France. Email: *name.surname*@isir.upmc.fr

[2] ENSTA ParisTech - INRIA Flowers Team, Paris, France. Email: *name.surname*@ensta-paristech.fr

one of the goals of the MACSi project[1]. We propose a general architecture, intrinsically modular and scalable at the functional level, which benefits from modularity, re-usability and hardware abstraction. The main feature of our solution with respect to the others is that it is natively designed for learning experiments where social guidance [13] is gradually superseded by autonomous behaviors driven by artificial curiosity and motivation [14]. This entails the adaptation and extension of existing learning and control techniques to a challenging context where, rather than a single goal, there is a gradually increasing repertoire of goals.

In this paper, we present the general framework for reaching these goals and illustrate the preliminary experiments performed on the humanoid robot iCub as a first step towards the learning of object affordances. Due to the lack of space, we shall strictly focus on the action-perception aspects of the architecture and defer to future papers the description of the curiosity mechanisms and the affordance learning techniques.

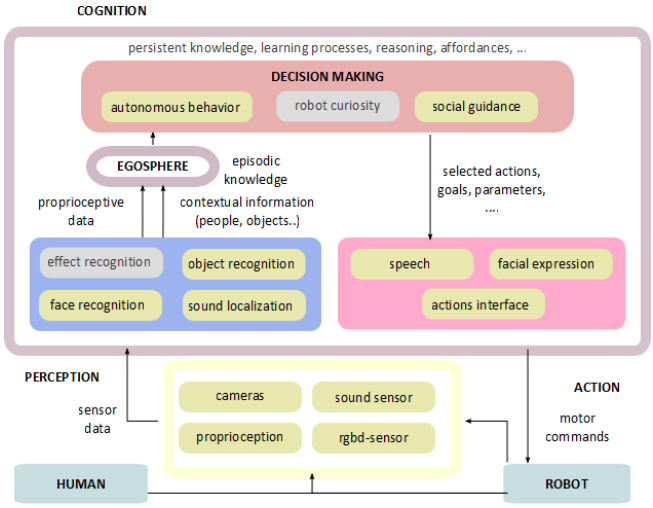


Fig. 2. A functional description of the elementary modules of the cognitive architecture. Grey blocks are not yet described in this paper.

II. COGNITIVE ARCHITECTURE

The CA is an integrated system which orchestrates all the cognitive, perceptive, learning and control modules. In general, the cognitive process is split into different stages, for example the robot "status" is built after visual and proprioceptive information, the motor actions are built in response to estimated states and goals, and so on. Here, we maintain a certain action-perception distinction, but finally a global cognitive process fuses the information of the two (see Fig. 2). This way, perceptual and motor skills of the robot can be increased in a modular fashion, and "improved" exploiting different learning mechanisms. The CA designed for MACSi is thus a growing framework with several cognitive functionalities. In the following we outline its elementary modules, and describe in more details the main ones used in the experiments described in Section IV.

[1]www.macsi.isir.upmc.fr

A. Scene perception

Segmenting the perceptual spaces into discrete objects representations such as the robot's own body, humans or manipulable objects, has been the subject of a lot of research in the computer vision community. However, the developmental approach we seek in MACSi imposes specific constraints: the approach should be generic and apply to many potential "objects" present in the environment (e.g. robot body, human hands, colored or textured objects) and it should perform online and incremental learning of new objects in an open-ended scenario. These constraints imply that such a system will probably be less efficient than algorithms specially tailored for specific tasks in specific environments. But, on the other hand, it will give a generic visual learning capacity that can be applied to novel tasks involving both the robot and the caregiver. In the proposed CA, perceptual information is extracted from several sensory sources. The main source is a rgbd-sensor camera whose data are processed using the approach presented in [15] to perform an efficient online incremental learning and recognition of various elements of the environment.

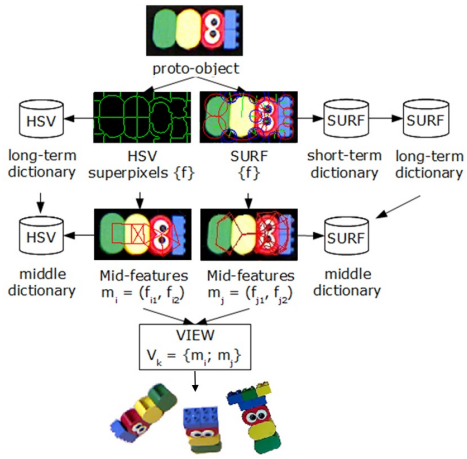


Fig. 3. Procedure for processing visual images.

The processing pipeline (Fig. 3, see details in [15]) starts by segmenting proto-objects [16] using coherent motion of KLT-points [17] and refining their boundaries from contours found in the depth information from the rgbd-sensor. The visual appearance of proto-objects is characterized using complementary features (SURF and color of superpixels [18]) in order to model efficiently both colored and textured objects. These low-level features are grouped into local geometric structures whose occurrence frequencies are used to describe the appearance of views. Views are then grouped using maximum likelihood recognition or tracking to create models of objects from different viewing directions. Beside this unsupervised segmentation approach, the robot cameras are also used to detect the presence of people. In particular, the "active" caregiver, i.e. the human partner who gives feedbacks and instructions to the robot, is tracked through a multimodal approach. The rgbd-sensor is used to estimate the position of the humans in front of the robot. This

information is then combined with the direction of sound sources: sound coming from the same direction of a person is assumed to be his voice, so the talking human will be marked as the active caregiver. The robot's decision system can use this information to focus its attention on the active caregiver, stimulating in this way eye contact sensations by simply gazing at his head. At the same time, the robot's cameras (moved with the eyes by the gazing system) are used to detect the human partner in its field of view. Since caregivers usually focus their attention on the robot, they naturally engage in eye-to-eye contact with the robot during interaction, so that faces perceived from the robot camera will naturally point to the robot [19], [20]. This will be exploited in future works to recognize the identity of the human partners, by the employment of standard facial features, such as Eigen Faces, and machine learning techniques, as Support Vector Machine. Fig. 4 shows the processing of information retrieved from the sensors (rgbd-sensor, robot cameras, etc.) "shared" across different middlewares (YARP and ROS).

B. Egosphere, decision making, HRI

The perceptive modules communicate to the egosphere, which implements the episodic knowledge of the robot. Specifically, it memorizes the objects in the scene and their main features and properties retrieved by the vision modules, such as the 3D position and orientation. Information about people and their location is integrated as well. The connection between the persistent and the episodic knowledge is not made explicit here (it will be covered in further works), because it is not relevant at this current stage for the preliminary experiments.

A planning module is then in charge of the decision-making process. A combination of social guidance, artificial curiosity and autonomous behaviors are entailed at this level.

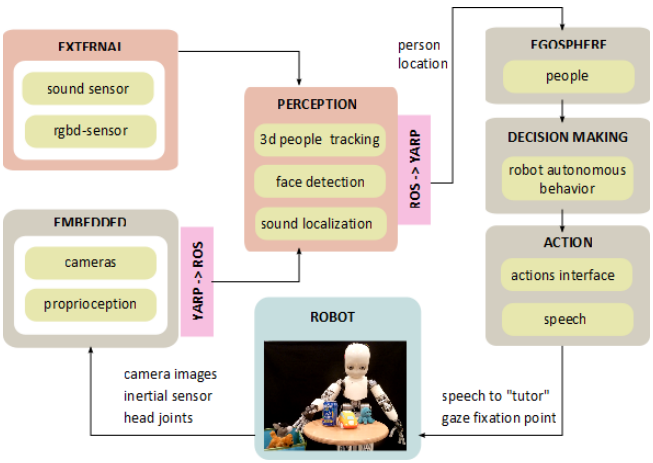
In the experiments described in Section IV, social guidance is restricted to the mere execution of commands received from the caregiver. In future plans, it will account for more complex interactions such as action recognition and negotiation of shared plans. Notably, a feedback from the partner will be used by the robot to improve its learning experience, as it will be described later on. As for now, it is only exploited so as to improve the engagement of the caregiver.

C. Action primitives

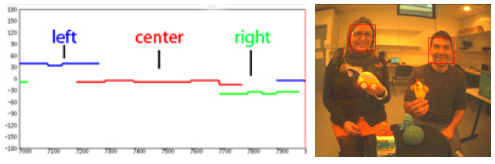
The perceptive and cognitive modules are interfaced to the robot through a pool of action modules, which act as intermediate controllers for speech, emotion interfaces and motor joints. An action interface module exposes a set of high level commands to control the robot. Modules can command simple or complex actions to the robot via a simple communication protocol, specifying the type of action (e.g. *take, grasp*) and a variable list of parameters (the object name, the person involved in the action, the location of the object, the type of grasp, and so on). Differently from [12], we do not define which are the motor primitives, but provide a set of pre-built actions, which can be simple (as the motor primitives grasp, touch, look) but also more complex. The idea is that the notion of motion primitive will become "recursive" when the learning process will start exploring autonomously and will make use of its whole repertoire of built and learnt actions to learn new ones. Pre-built actions are: (simple) *speak, look, grasp, reach, take, rotate, push, put-on, lift, point* (complex) *put object on top of another, lift an object and make it fall, manipulate an object to see all its sides, put the object somewhere, rotate an object*.

If unpredictable events occur during the execution of an action or a task, for example an unsuccessful grasp or a potentially harmful contact with the environment, one or more autonomous reflexes are triggered. These reflexes are pre-coded sequences of actions that may interrupt or change the execution of the current action or task.

A reward mechanism is also implemented. A positive reward is generated when actions are successful, for example when the robot executes a particular action correctly, or when it recognizes an object. A negative one is generated in case of unsuccess, for example when a grasp fails because the object slips. Rewards are not used at this stage, but will be fundamental to harness future learning experiments.



(a)



(b)

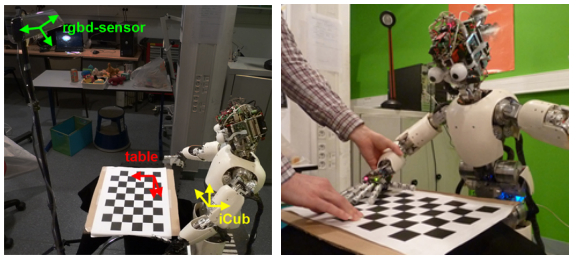
(c)

Fig. 4. Human localization. (4(a)): integration of YARP and ROS modules: two colors are used to distinguish modules and drivers for one middleware or the other. (4(b)): localization of three different sound sources (people talking) in front of the robot; (4(c)) detecting faces as seen from by the robot eyes during interaction.

III. EXPERIMENTAL SETUP

We hereby detail the robotics setup for the experiments of Section IV.

1) Robotic platform: Experiments are carried out with iCub, a 53 DOF full-body humanoid robot shaped as a 3



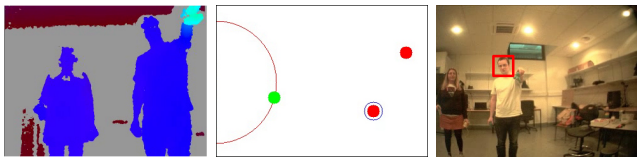
(a) (b)

Fig. 5. Calibration of the rgbd-sensor with respect to the robot reference frame.

years old child [21]. The whole upper-body has been used in our experiments: head (6 DOF), torso (3 DOF), arms (7 DOF each) and hands (9 DOF each), for a total of 41 DOF. The iCub hand allows a considerable dexterity though being small. Thanks to the tendons elasticity, fingers are naturally compliant during contact with objects. A Cartesian controller is used to control the head gaze and steer the robot hand to desired position and orientation in the Cartesian 3D space. An optimization routine takes into account the redundancy of arm and torso [22]. To ensure compliance during motions, impedance control is activated at the main joints (torso and arms). Joint torques and external forces due to contacts are estimated exploiting the proximal Force/Torque sensors placed in the middle of the arms (because the robot is not equipped with joint torque sensors). Precisely, the estimation is performed by iDyn, a C++ library for inverse dynamics which is part of the iCub open-source software project [23]. Several excentric sensors are used: an rgbd-sensor placed over the table to segment objects; a microphone array and a further rgbd-sensor placed behind the robot are used to detect and locate the human caregivers interacting with the robot. Sound (human voices, but also the sound produced by objects during manipulation) is processed by HARK library [24], while rgbd-data are elaborated by OpenNI[2].

2) Software architecture: The CA enjoys a *de facto* parallel execution model. Each component of the architecture provides certain functionalities, seen as services, through a well-defined interface. Communication between modules is usually developed in a relatively fixed topology, sometimes fixed by the modules themselves. All software modules are abstracted from the hardware level thanks to the YARP [25] middleware, which provides an interface to the devices and the robotic platform. Modules running on different middlewares such as ROS[3] can coexist and exchange data with the ones of YARP thanks to a custom developed bridge.

3) Calibration: A calibration procedure is required to match the external sensors data within the robot reference frame, so as to link the perception and the action modules coherently. The rgbd-sensor used to segment and detect the objects on the table is placed in front of the robot, as shown in Fig. 5(a). Its 3D system reference frame is centered on the device sensor. A transformation matrix between the rgbd-sensor and the robot is computed exploiting a third reference frame, usually located on the table and determined as the origin of a chessboard, as shown in Fig. 5(b). The chessboard dimensions being known, the table frame is computed exploiting the chessboard image retrieved by the rgbd-sensor. The position of the chessboard is then retrieved in the robot's coordinates by putting the hand of the robot on a predefined location (i.e. the "origin" of the chessboard). The total transformation matrix is then computed: $T^{robot}_{sensor} = T^{robot}_{table}\, T^{table}_{sensor}$. A similar procedure is used to calibrate other external sensors (e.g. sound arrays).



(a) (b) (c)

Fig. 6. Experiment IV-A: active caregiver tracking. Two caregivers interact with the robot and catch its attention by speaking and showing objects. (6(a)): depth map from the rbgd-sensor. (6(b)): multimodal people tracking, fusing people 3D localization with sound source direction; the "active" caregiver is marked by a circle. (6(c)): the face of the "active" caregiver detected from the embodied robot camera (eye).

IV. EXPERIMENTS

In the following, we describe the preliminary experiments involving action, perception and interaction. From simple to more complex scenarii, the last experiments are the basis for learning affordances.

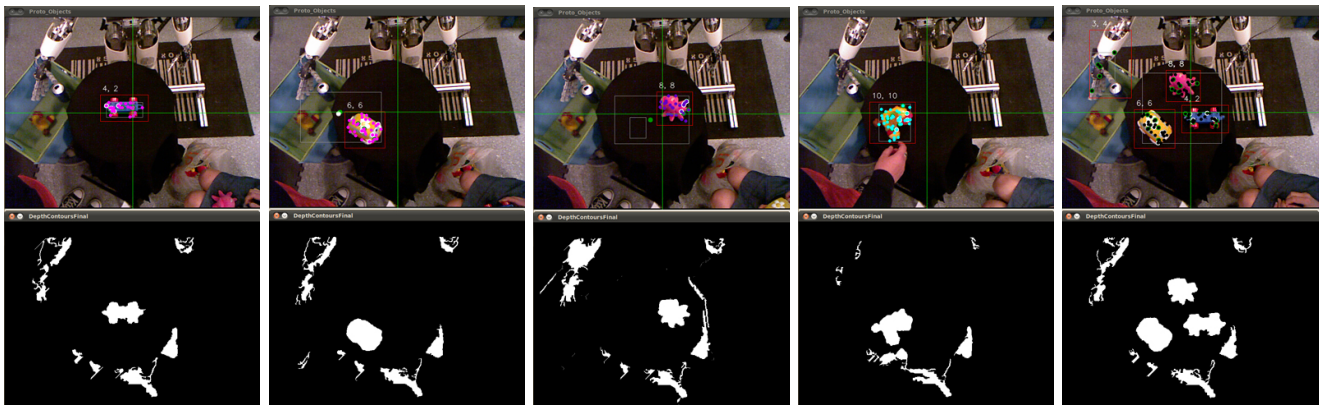
A. Basic human-robot interaction

This experiment is used to test the HRI modules, precisely the perceptual system dedicated to the sound localization and people detection. Two caregivers are in front of the robot: their 3D location is tracked exploiting standard skeleton algorithms and the rgbd-data. As one of the people speaks, the produced sound is detected by the external sensor. The robot then turns its head so as to gaze at the person at once. At this point, the images from the robot eyes (i.e. cameras) are scanned by a face detector to match the sound. Fig. 6 shows the multimodal approach. If the caregiver speaks and moves at the same time, the robot tracks the human gazing at his head. If people talk alternatively, as during a conversation, the robot gazes at the "active" one (i.e. the speaking one).

B. Learning objects

This experiment is used to test the ability of the perceptual system to detect and recognize different objects. An overview of the experiment is shown in Fig. 7. In a preliminary phase, several objects are shown in sequence to the robot, with the purpose to build an episodic knowledge of the objects (see Figures 7(a)-7(c)). Here, we assume that the objects are totally new to the robot, and there is no prior information about the scene or its items. When the human caregiver introduces a new object in the visual field, its features are detected and a number is assigned both to the object and

[2]www.openni.org
[3]www.ros.org



(a) (b) (c) (d) (e)

Fig. 7. Experiment IV-B: detecting and learning objects. Several objects are shown to the robot, for example a blue car 7(a), a yellow car 7(b), an octopus 7(c), a chicken 7(d). Two people interact with the robot, showing the objects on the table, rotating the objects so that the robot can take different views of the items. The name of the object is associated in this preliminary learning phase. In a secondary phase 7(e), one of the human partners asks the robot to recognize the object in a complex scene, where multiple objects are present. The images in the first row show a snapshot of the scene as seen by the rgbd-sensor and the identification of the proto-objects. Each object is labeled by two numbers, corresponding to its "id" and its "view". The images in the second row show the segmentation of the scene after the depth image.

Fig. 8. Experiment IV-B: recognizing objects. The robot is shown two different objects, then it is asked to recognize and point at one of them when both are present on the table.

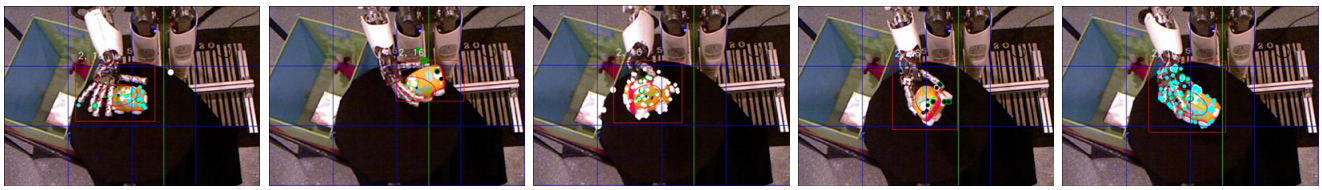


Fig. 9. Experiment IV-C: exploiting action for perception. The robot recognizes an object, grasps it and rotates it in its hand to take more views of it, and unveil hidden parts.

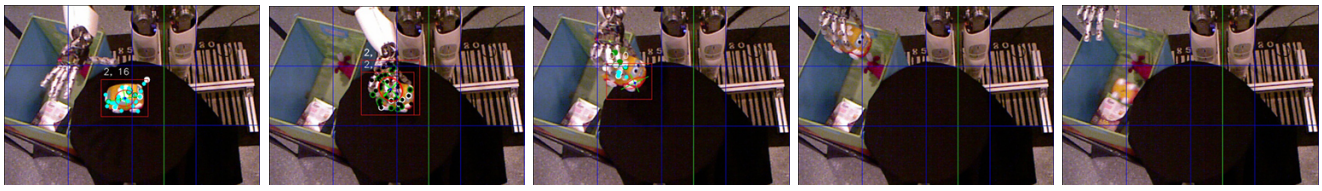


Fig. 10. Experiment IV-C: a task combining perception and action. The robot recognizes an object, grasps it and puts it in a box.

the "view" of the object. The human then moves the object around the scene, rotating it so that the vision system can associate more "views" to it (see Fig. 7(d)). A name is also assigned to the object (e.g. *octopus*) by the caregiver. In a secondary phase, the caregiver shows the robot a set of objects, and asks the robot to point a particular object. The robot recognizes the object and moves its hand to point the object to the caregiver. The robot then asks for a feedback: in case of success, a reward is generated, which feeds the "intelligence" modules (e.g. curiosity, autonomous behavior and learning). An example, illustrated in Fig. 8:

H: This is an octopus. *(caregiver shows the octopus, moves it on the table)*
H: *(caregiver adds the car on the table)* This is a blue car. *(caregiver removes the octopus, then moves the car on the table)*

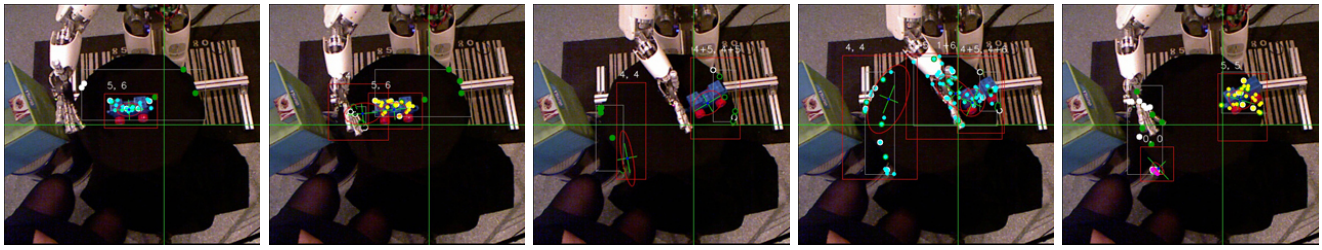


Fig. 11. Experiment IV-D: towards learning affordances. The robot pushes an object to detect its motion and learn the effect of its action (e.g. *the blue car rolls*).

H: *(caregiver places both octopus and car on the table)* Where is the octopus?
R: I found the octopus. This is the octopus. *(robot points at the octopus)* Serena, was it good?
H: Yes. *(positive reward for the success in recognizing the object)*
R: I am happy I did good.

In case of error (e.g. failure in the recognition of the object) the robot is given a negative feedback:

R: I found the octopus. This is the octopus. *(robot points at a wrong location)* Serena, was it good?
H: No. *(negative feedback)*
R: I am sorry.

Remarkably, the robot always refers to the caregiver by his/her name, and asks his/her feedback at the end of its actions. This is done to enhance the engagement of the caregiver during the interaction and learning experiments.

C. Combining action and perception

This experiment is used to integrate action and perception modules. In a first phase, objects are learnt such as described in Experiment IV-B. In a second phase, the caregiver asks the robot to perform some simple actions to manipulate the objects on the table. Such actions can be simple atomic actions (e.g. *reach the object*, *grasp the object*) or more complex actions (e.g. *take the object and put it in the box*, *take the object and look at it closer*). The purpose is to make the robot interact, manipulate and perform different actions on different (partially) known objects, to increase the knowledge of the objects (e.g. explore all sides). Indeed, it is worth noticing that manipulating objects may also affect their recognition: after an action is performed, the object's properties generally change (e.g. its view, position or orientation). Simple manipulations can also improve the learning capabilities of the perceptual modules: for example, if the object is grasped and rotated by the robot, it may allow the object recognition module to learn about its hidden parts. A simple example, as shown in Fig. 9, is the following:

H: Take a look at the yellow car.
R: I found the yellow car. I take the yellow car and look at it closer. *(robot reaches the car, grasps it, lifts it and brings it closer to its head)*
R: Interesting object. *(the robot rotates the car in its hand)* Here it is. *(the robot puts the car back where he took it)*

When simple (or more complex) actions are chained so as to realize a desired effect or accomplish a task, errors are associated to each simple action and a final reward is given or not depending on the success of the task. For example, the caregiver may want the robot to clean the table, and command the robot to put all the objects on the table inside a box. In this case, as shown in Fig. 10, a possible experiment could be:

H: Take the yellow car and put it in the box.
R: I take the yellow car and put it in the box. *(robot reaches the car, grasps it, lifts it, reaches the top of the box, opens its hand and makes the car fall into the box)* Serena, was it good?
H: Yes. *(positive reward for the success of completing the task)*

The error/reward mechanism is not yet used at this stage, but will be crucial for future experiments involving autonomous learning and discovery of actions and affordances.

D. Towards learning affordances

This experiment is a first step towards learning of objects and affordances. The purpose of this experiment is to exploit the robot actions to improve its knowledge of objects and understanding of the effects of its actions on the objects. In a first preliminary phase objects are learnt such as described in Experiment IV-B. In a secondary phase, the caregiver suggests the robot a certain set of actions to interact with the objects on the table and perform simple manipulation tasks. Such actions are generally complex actions (e.g. *push the object towards the left*, *take the object and make it fall*, *put the object on top of another one*) which have the purpose to change the "properties" of the object in the egosphere representation: for example, we may want to change its position on the table, discover if it bounces when falling, if it rolls when pushed, and so on. An example, illustrated in Fig. 11 is the following:

H: Where is the blue car?
R: I found the blue car. Here it is. *(robot points at the car)* Am I right?
H: Yes. *(positive reward for the success in recognizing the object)*
R: I am happy I did good.
H: Push the blue car.
R: I push the blue car and see what happens. *(robot pushes the car)* Serena, was it good?

H: Yes. *(positive reward for the success of pushing the object)*

The learning process is guided by the caregiver, instructing the robot how to manipulate the object: in the future, it will be gradually replaced by artificial curiosity and autonomous robot behavior.

E. Video and code

A video demonstrating the experiments is attached to the paper. The software for the architecture and the experiments is available under GPL license, together with the full video, at `http://macsi.isir.upmc.fr`.

V. CONCLUSIONS

In this paper, we presented the architecture we are using in MACSi to design experiments in developmental robotics scenarii, where the robot interacts with caregivers to learn objects and their affordances, and eventually to take decisions autonomously. The main feature of our solution is that it is natively designed for learning experiments, where social guidance is combined and gradually replaced by artificial curiosity and autonomous exploration. We performed several experiments to assess the efficiency of the proposed solution, setting a solid base for future research in affordance recognition. We focused on the perception and interaction modules; for lack of space, we did not introduced how objects, actions and caregivers are formally represented in the architecture. The internal representation and the techniques for active learning of affordances will be object of forthcoming papers. The next step in the evolution of the CA will be to integrate intrinsic motivation in the decision making process [14], to gradually diminish the role of the caregiver and make the robot take its own decisions autonomously, driven by its artificial curiosity.

ACKNOWLEDGMENT

This work is supported by the French ANR program (ANR 2010 BLAN 0216 01).

REFERENCES

[1] O. Sigaud and J. Peters, "From motor learning to interaction learning in robots," in *From Motor Learning to Interaction Learning in Robots*, O. Sigaud and J. Peters, Eds. Springer, 2010, ch. 1, pp. 1–12.

[2] P. H. Miller, *Theories of developmental psychology*, 5th ed. Worth Publishers, 2010.

[3] M. Lopes and P. Oudeyer, "Active learning and intrinsically motivated exploration in robots: Advances and challenges (guest editorial)," *IEEE Trans. Aut. Mental Development*, vol. 2, no. 2, pp. 65–69, 2010.

[4] J. Gibson, *Perceiving, Acting, and Knowing: Toward an Ecological Psychology (R. Shaw & J. Bransford Eds.)*. Lawrence Erlbaum, 1977, ch. The Theory of Affordances, pp. 67–82.

[5] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, "Learning object affordances: From sensory–motor coordination to imitation," *IEEE Trans. on Robotics*, vol. 24, no. 1, pp. 15–26, 2008.

[6] C. Castellini, T. Tommasi, N. Noceti, F. Odone, and B. Caputo, "Using object affordances to improve object recognition," *IEEE Trans. Aut. Mental Development*, vol. 3, no. 3, pp. 207 –215, 2011.

[7] S. Hart and R. Grupen, "Learning generalizable control programs," *IEEE Trans. Aut. Mental Development*, vol. 3, no. 3, pp. 216 –231, 2011.

[8] E. Ugur and E. Sahin, "Traversability: A case study for learning and perceiving affordances in robots," *Adaptive Behavior*, vol. 18, pp. 258–284, 2010.

[9] S. Griffith, J. Sinapov, V. Sukhoy, and A. Stoytchev, "A behavior-grounded approach to forming object categories: Separating containers from noncontainers," *IEEE Trans. Aut. Mental Development*, vol. 4, no. 1, pp. 54 –69, 2012.

[10] J. Mugan and B. Kuipers, "Autonomous learning of high-level states and actions in continuous environments," *IEEE Trans. on Autonomous Mental Development*, vol. 4, no. 1, pp. 70 –86, 2012.

[11] M. Malfaz, A. Castro-Gonzalez, R. Barber, and M. Salichs, "A biologically inspired architecture for an autonomous and social robot," *IEEE Trans. Aut. Mental Development*, vol. 3, no. 3, pp. 232 –246, 2011.

[12] S. Lallée, U. Pattacini, J.-D. Boucher, S. Lemaignan, A. Lenz, C. Melhuish, L. Natale, S. Skachek, K. Hamann, J. Steinwender, E. A. Sisbot, G. Metta, R. Alami, M. Warnier, J. Guitton, F. Warneken, and P. F. Dominey, "Towards a platform-independent cooperative human-robot interaction system: II. perception, execution and imitation of goal directed actions," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2011, pp. 2895–2902.

[13] C. Rich, A. Holroyd, B. Ponsler, and C. Sidner, "Recognizing engagement in human-robot interaction," in *ACM/IEEE International Conference on Human Robot Interaction*, 2010, pp. 375–382.

[14] P.-Y. Oudeyer, F. Kaplan, and V. Hafner, "Intrinsic motivation systems for autonomous mental development," *IEEE Trans. on Evolutionary Computation*, vol. 11, no. 2, pp. 265–286, 2007.

[15] N. Lyubova and D. Filliat, "Developmental approach for interactive object discovery," in *Int. Joint Conf. on Neural Networks*, 2012.

[16] Z. W. Pylyshyn, "Visual indexes, preconceptual objects, and situated vision," *Cognition*, vol. 80, pp. 127–158, 2001.

[17] C. Tomasi and T. Kanade, "Detection and tracking of point features," Carnegie Mellon University, Tech. Rep., 1991.

[18] B. Micusik and J. Kosecka, "Semantic segmentation of street scenes by superpixel co-occurrence and 3d geometry," in *IEEE Int. Conf. on Computer Vision*, 2009, pp. 625–632.

[19] S. Al Moubayed, M. Baklouti, M. Chetouani, T. Dutoit, A. Mahdhaoui, J.-C. Martin, S. Ondas, C. Pelachaud, J. Urbain, and M. Yilmaz, "Generating robot/agent backchannels during a storytelling experiment," in *IEEE Int. Conf. on Robotics and Automation*, 2009, pp. 3749– 3754.

[20] S. Anzalone, S. Ivaldi, O. Sigaud, and M. Chetouani, "Multimodal people engagement with icub," in *Int. Conf. on Biologically Inspired Cognitive Architectures*, Palermo, Italy, 2012.

[21] L. Natale, F. Nori, G. Metta, M. Fumagalli, S. Ivaldi, U. Pattacini, M. Randazzo, A. Schmitz, and G. G. Sandini, *Intrinsically motivated learning in natural and artificial systems*. Springer-Verlag, 2012, ch. The iCub platform: a tool for studying intrinsically motivated learning.

[22] U. Pattacini, F. Nori, L. Natale, G. Metta, and G. Sandini, "An experimental evaluation of a novel minimum-jerk cartesian controller for humanoid robots." in *Proc. of IEEE/RSJ Int. Conf. Intelligents Robots and Systems*, 2010, pp. 1668–1674.

[23] S. Ivaldi, M. Fumagalli, M. Randazzo, F. Nori, G. Metta, and G. Sandini, "Computing robot internal/external wrenches by means of inertial, tactile and f/t sensors: theory and implementation on the icub," in *Proc. of the 11th IEEE-RAS Int. Conf. on Humanoid Robots*, 2011, pp. 521–528.

[24] K. Nakadai, T. Takahashi, H. Okuno, H. Nakajima, Y. Hasegawa, and H. Tsujino, "Design and implementation of robot audition system hark," *Advanced Robotics*, vol. 5, no. 6, pp. 739–761, 2010.

[25] P. Fitzpatrick, G. Metta, and L. Natale, "Towards long-lived robot genes," *Robotics and Autonomous Systems*, vol. 56, no. 1, pp. 29–45, 2008.