Learning to recognize objects through curiosity-driven manipulation with the iCub humanoid robot

Sao Mai Nguyen, Serena Ivaldi, Natalia Lyubova, Alain Droniou, Damien Gerardeaux-Viret, David Filliat, Vincent Padois, Olivier Sigaud, Pierre-Yves Oudeyer

Learning to recognize objects through curiosity-driven manipulation with the iCub humanoid robot

Sao Mai Nguyen[1], Serena Ivaldi[2], Natalia Lyubova[3], Alain Droniou[2], Damien Gérardeaux-Viret[3], David Filliat[3], Vincent Padois[2] , Olivier Sigaud[2] and Pierre-Yves Oudeyer[1]

***Abstract*—In this paper we address the problem of learning to recognize objects by manipulation in a developmental robotics scenario. In a life-long learning perspective, a humanoid robot should be capable of improving its knowledge of objects with active perception. Our approach stems from the cognitive development of infants, exploiting active curiosity-driven manipulation to improve perceptual learning of objects. These functionalities are implemented as perception, control and active exploration modules as part of the Cognitive Architecture of the MACSi project. In this paper we integrate these functionalities into an active perception system which learns to recognise objects through manipulation. Our work in this paper integrates a bottom-up vision system, a control system of a complex robot system and a top-down interactive exploration method, which actively chooses an exploration method to collect data and whether interacting with humans is profitable or not. Experimental results show that the humanoid robot iCub can learn to recognize 3D objects by manipulation and in interaction with teachers by choosing the adequate exploration strategy to enhance competence progress and by focusing its efforts on the most complex tasks. Thus the learner can learn interactively with humans by actively self-regulating its requests for help.**

I. INTRODUCTION

Motor activity plays a fundamental role in the learning process about objects and their properties. The action-perception coupling is particularly evident during the cognitive development of children, who learn object representations essentially through interaction and manipulation [1]. By means of simple actions like pushing or throwing, infants can perceive an object from different points of view, learn its different "appearances", i.e. improve their representation of the manipulated object. In the early stages of their cognitive development, infants mostly learn the visual properties of objects that are shown by their caregivers. Once they become capable of controlling their body and perform goal-directed actions, they learn to act independently and to explore objects on their own. As children grow, they gradually understand what actions can be associated to objects and learn to predict the outcome of their actions on the objects, i.e. to learn object affordances [2].

Many studies in humanoid robotics have been inspired by such evolution of behaviors, where the quality of manipulation relates to the knowledge about explored objects. For example, in [3] the robot learns objects that are simply shown by the caregiver, whereas in [4] it chooses actions which are expected to reveal the most information about the objects in the scene. In [5] the robot performs simple actions (grasping, pushing, *etc.*) to learn the properties of objects (e.g. a ball and a cylinder can roll). In general, coupling manipulation with vision outperforms passive, vision-only object recognition [6].

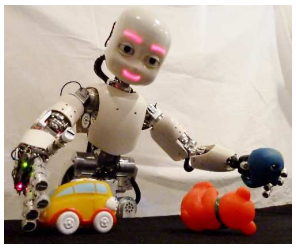


Fig. 1: The humanoid iCub and the experimental context.

The MACSi project[1] continues in this line of research and investigates a mechanism underlying the choice of "the next action to perform to improve the knowledge of objects". In [7], the iCub recognized objects shown by a caregiver. We tested the ability of the perceptual system to track and recognize the object during simple manipulations, such as pushing.

Here, we limit the robot's dependence on the caregiver, by enabling the robot to choose whether to request for help with interactive learning, based on curiosity and intrinsic motivation [8]. More generally, the learner chooses actively between several data collection strategies on a meta level. The experimental scenario consists of a humanoid robot, minimally assisted by a caregiver, manipulating multiple objects to learn to recognize them better. In such a context, the choice of an object (to explore) and the choice of the exploration strategy (e.g. which manipulation to perform on the object that represents a set of possible image generation rules) to adopt is crucial. Intrinsic motivation and socially guided learning have proved to be efficient exploration methods for autonomous agents taking such decisions [9], [10].

In this paper, we address the problem of active object recognition exploiting intrinsic motivation from a developmental perspective. We design a Cognitive Architecture for life-long learning in natural environment, with multiple outcomes to learn and with multiple strategies using Interactive Learning, introduced in Section II. We first illustrate our problem in Section III. We outline the perception, control and decision making components of the Cognitive Architecture of the robot in Section IV. We describe an object recognition experiment with the humanoid robot iCub in Section V and show that our *Socially Guided Intrinsic Motivation with Active Choice of Teacher and Strategy* (SGIM-ACTS) algorithm can recognise

This work was supported by the French ANR program (ANR 2010 BLAN 0216 01) through Project MACSi.

[1] S.M. Nguyen and P.-Y. Oudeyer are with Flowers Team, INRIA, Bordeaux - Sud-Ouest, France. name.surname@inria.fr

[2] S. Ivaldi, A. Droniou, V. Padois and O. Sigaud are with Institut des Systèmes Intelligents et de Robotique, CNRS UMR 7222 & Université Pierre et Marie Curie, Paris, France. name.surname@isir.upmc.fr

[3] N. Lyubova, D. Gérardeaux-Viret and D. Filliat are with Flowers Team, ENSTA ParisTech, Paris, France. name.surname@ensta-paristech.fr

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

efficiently several objects by focusing on the complex objects in Section VI.

II. RELATED WORK

Recognising several objects belongs to the broader challenge learning mappings for various outcomes under time and resources constraints in unstructured environments. In control problems, outcomes are typically the end-effector positions which have to be mapped with control policies. In a classification problem, outcomes are class objects which have to be mapped with element features. Classical approaches to this problem include Intrinsic Motivation, Imitation Learning and in particular Interactive Learning.

A. Active Learning for Producing Varied Outcomes with Multiple Strategies

The learning agent has to decide in which order it should focus on the different outcomes, how much time it can spend on learning a specific outcome or which methods to adopt for a given outcome, as a strategic student would do. These questions can be formalised under the notion of strategic learning [11] and have been addressed in several works.

One perspective is learning varied outcomes. It aims at selecting which outcome to spend time on. A typical classification was proposed in [12], [13] where active learning methods improved the overall quality of the learning. In sequential problems as in robotics, producing an outcome has been modelled as a local predictive forward model [14], an option [15], or a region in a parameterised goal/option space [16]. In these works each sampling of an outcome entails a cost. The learning agent has to decide which outcome to explore/observe next.

Another perspective is learning how to learn, by making explicit the choice and dependence of the learning performance on the method. For instance, [17] selects among different learning strategies depending on the results for different outcomes. [18] implemented a control based on information gain to classify categories of objects in a room. However these studies focus only on the search of an action to perform, and not on the object/outcome the action is performed on.

Here, we study how a learning agent can produce multiple outcomes, and how it can learn for those various outcomes which strategy to adopt simultaneously.

B. Interactive Learning

Imitation learning is an intuitive medium of communication for humans, who already use demonstrations to teach other humans. It thus offers a natural mean of teaching machines that would be accessible to non experts. That is why several works incorporate human input to a machine learning process, such as in some examples of Programming by Demonstration (PbD) [19] or learning by physical guidance [20], [21], where the learner scarcely explores on its own. Prior works have also given a human trainer control of a reinforcement learning reward [22], [23], provided advice [24], or tele-operated the agent during training [25].

Furthermore, an interactive learner which does not only listen to the teacher, but actively requests for the information it needs and when it needs help, has been shown to be a fundamental aspect of social learning [26]. In the interactive learning approach, the robot interacts with the user, combining learning by self-exploration and social guidance. Several works in interactive learning have considered extra reinforcement signals [27], action requests [28], [29] or disambiguation among actions [26]. Interactive systems for multiple outcomes have also been presented in [27], [30].

On the one hand, adding autonomous exploration to socially guided learning averts the case where learning depends too much on the teacher, which is limited by ambiguous human input or the correspondence problem and can quickly turn out to be too time-consuming [10]. While self-exploration fosters a broader outcome repertoire, exploration guided by a human teacher tends to be more specialized, resulting in fewer outcomes that are learnt faster. Combining both can thus bring out a system that acquires a wide range of knowledge which is necessary to scaffold future learning with a human teacher on specifically needed outcomes, as proposed in [27], [21], [29].

On the other hand, adding socially guided learning to autonomous exploration is beneficial on two different levels. First, while a learner might explore an outcome space for its own sake using intrinsic motivation mechanisms, social guidance can introduce it to new outcomes it may not have discovered otherwise. Then, given an outcome, social guidance can introduce new means for achieving intrinsically motivated activities by providing new examples. One might either search in the neighborhood of the good example, or eliminate bad examples from the search space. The structure of demonstrations can also encourage exploration both in the action space and the outcome space, in particular subspaces that are more powerful for generalization as shown in [31].

Thus, interactive learning is an interesting example of Strategic Learning, where the agent decides between autonomous learning and socially guided learning, and uses both strategies to bootstrap each other.

C. Intrinsic Motivation

Intrinsic motivation, a particular example of internal mechanism for guiding exploration, has recently drawn a lot of attention, especially for open-ended cumulative learning of skills in autonomous robots [32], [9]. The expression *intrinsic motivation*, closely related to the concept of *curiosity*, was first used in psychology to describe the spontaneous attraction of humans toward different activities for the pleasure that they experience intrinsically [33]. These mechanisms have been shown crucial for humans to autonomously learn and discover new capabilities [34]. This inspired the creation of fully autonomous robots with meta-exploration mechanisms monitoring the evolution of learning performances [35], [36], [37], with heuristics defining the notion of interest used in an active learning framework [38], [39]. We develop our interactive system, where the learner decides whether to interact with the teacher, and which exploration strategy to use, based on intrinsic motivation, and in particular on measures of competence progress.

Overall, it is critical for life-long learning to decide which outcome to learn to achieve and which learning strategy to adopt. Learning agents taking such decisions can fruitfully

profit of intrinsic motivation and socially guided learning. Furthermore, the combination of both methods into Interactive Learning algorithms has shown better accuracy and less dependence on the human caregiver [40].

III. PROBLEM FORMALIZATION

In this section, we describe shortly our experiment, then formalize the framework of classification with several data sampling strategies which encapsulates our problem.

A. Description of the experiment

The robot learns to associate a camera view with an object by episodes. At each episode, it has to decide which object it wants to learn more, which manipulation to use as an exploration strategy to manipulate the object. Once the object has been manipulated, the robot acquires a new image of the object, for which it computes its competence at recognising the right object. This new data is used to improve the recognition algorithm and learn to better distinguish between objects. In this section, we focus, not on the classification algorithm (which is described in section IV-B), but on the exploration method: how the robot generates new images by deciding a strategy of manipulation.

B. Mathematical Formalisation

Our agent learns a binary relation M between the space A of camera views and the space B of the objects. We suppose in this experiment that only one object is in the robot's field of vision at a time, and that $M : A \rightarrow B$ is a function. A is the space of all possible rgb-d images. In our experimental setting, A is of dimension $4 \times 480 \times 640$. B is the set of objects to be recognized, i.e. $B = \cup_{\text{All Objects}} b_i$ where b_i is the object i. The forward relation M is the true labelling of the images. For an object $b \in B$, $M^{-1}(\{b\})$ is the set of all images of the object b in its different positions and orientations. The binary relation M is a priori very redundant as an infinity of images correspond to the same object seen through different angles, positions and distances. The learner estimates the true labelling M with an estimation L. Let $\gamma_L(a)$ be a measure of competence at recognising the right object in image a with the estimation L. Our goal is to recognize all objects, i.e. to maximize with respect to L:

$$I = \sum_a P(a)\gamma_L(a) \tag{1}$$

where $P(a)$ is a probability over A that a appears to the robot. To learn to recognize different views of objects, the learner must sample more images of each object. While classical active learning methods choose images $a \in A$ and then ask for their objects $b \in B$, our method mainly explores the object space B by choosing first an object, and generating images with 3 different strategies: it can push the object, lift and drop the object, or it can ask a human to manipulate an object. These different strategies σ have different costs $\kappa(\sigma)$ that take into account the time cost, energy cost, caregiver effort of each strategy. In this study $\kappa(\sigma)$ are set to arbitrary constant values.

To summarize, at each episode the robot has to decide which object it wants to learn more, which manipulation to use as a strategy to generate new sample data, and then to learn to distinguish between objects.

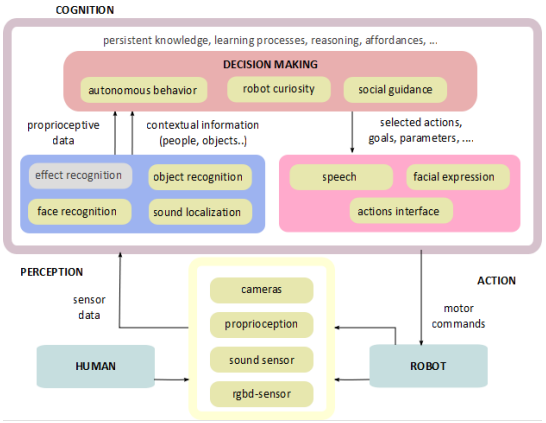


Fig. 2: A functional description of the elementary modules of the cognitive architecture.

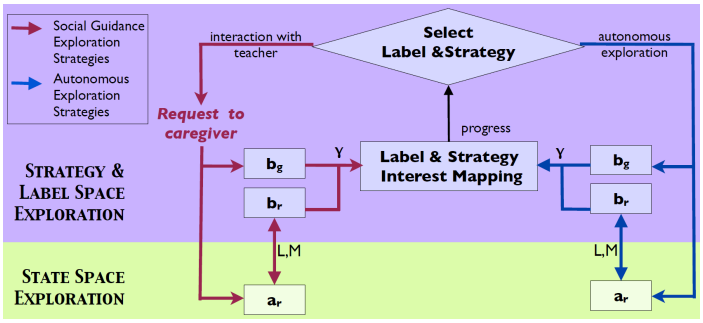


Fig. 3: Time flow chart of SGIM-ACTS, which combines Intrinsic Motivation and Social Guidance exploration strategies into 2 layers: the strategy and object space exploration and the state space exploration.

IV. METHODS

Designing complex experiments where humanoid robots interact with caregivers, manipulate objects and take decision in an autonomous way necessarily requires 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*. These functionalities are implemented in several software modules, which are integrated in the Cognitive Architecture (CA) of the robot and executed concurrently on the robotic platform. In this paper, our experiments are grounded on the MACSi Cognitive Architecture [7]. The main feature of this CA is that it is natively designed for learning experiments in a developmental robotics context, where social guidance [41] is gradually superseded by autonomous behaviors driven by artificial curiosity and motivation [42]. The CA is an integrated system which orchestrates all the perceptive, motor and cognitive modules (see Fig. 2).

A. Action

An action module controlling the robot exposes a set of high level commands to the perceptive and cognitive modules. It acts as intermediate controller for speech, emotion interfaces and motor joints. Modules can send commands to the robot, specifying the type of action and a variable list of parameters (the object properties, e.g. name, location on the table, orientation; the person involved in the action; the type of grasp; the

action timing, *etc.*). Actions can be simple (for example the primitives *grasp*, *touch*, *look*, *reach*, *lift*, *speak*) but also more complex (as taking an object, manipulating it and putting it on a desired location, piling objects, *etc.*). Autonomous reflexes are triggered by unpredictable events that could potentially harm the robot or cause failures in the execution of a given command. Primitives for basic HRI are reported in [43].

B. Scene perception

The perceptual system of the robot combines several sensory sources. The primary source for object detection is a rgb-d sensor placed over the area where object manipulation occurs. The object recognition system is based on an incremental online learning approach, which is bootstrapped without any *a priori* knowledge about the visual scene or the objects [44]. Visual attention is focused on motion: proto-objects [45] corresponding to regions of interest are identified in each frame thanks to the depth image information, hence they are tracked using KLT-tracking [46]. To suitably describe both homogeneous and complex textured objects, SURF descriptors and HSV colors models are used. Particularly, colors are analyzed on the level of regularly segmented superpixels that corresponds to regions of similar adjacent pixels [47]. Both SURF and color descriptors are quantized into vocabularies that build the basis of our visual system. In order to incorporate geometry into the object model, the closest SURF points and superpixels are grouped into pairs and triples based on their distance in the visual space. The Bag of visual Words (BoW) approach with incremental dictionaries is used to characterize objects appearance through the occurrence of mid-level features [48]. Mid-features are quantized into vocabularies and used to encode an object appearance from different perspectives called *views*. In order to learn an overall appearance of an object (i.e. multiple views of the same object), we accumulate the visual information from different viewing points into a multi-view model. All recognized views are associated with their objects while objects are tracked during manipulations (the identity of the object is taught by the teacher). The object recognition system is based on a voting method using the TF-IDF (Term-Frequency - Inverse Document Frequency) [49] and maximum likelihood approach. Each set of extracted mid-features is labeled according to the maximum likelihood of being one of already learned views. If the probability of recognition is low, the view is stored as novel. Thus, the vision system first uses a bottom-up approach to organize the views it sees, then associates by supervised learning views and objects.

At each image $a \in A$ seen, the iCub computes the likelihood for each already known views, and returns the two highest likelihood measures p_{m1}, p_{m2}, as well as the objects b_{m1}, b_{m2} of the objects associated with the views, and the number n_{m1}, n_{m2} of known views for each of the objects. As through social interaction, the caregiver teaches to the iCub the object b_g of the object he is manipulating, the robot can estimate its competence at distinguishing b_g from other objects, with the dissimilarity of likelihood measures between the 1st object associated and the 2nd object associated, and by estimating its gain of information about the object by collecting new views. The competence at recognising object b_g in image a is thus defined as

$$\gamma(b,a) = \begin{cases} n_{m1} \times p_{m1} + c_1 & \text{if } b_g = b_{m1} = b_{m2} \\ n_{m1} \times p_{m1}/(1+p_{m2}) + c_1 & \text{if } b_g = b_{m1}, b_g \neq b_{m2} \\ n_{m2} \times p_{m2}/(1+p_{m1}) + c_1 & \text{if } b_g \neq b_{m1}, b_g = b_{m2} \\ c_1 & \text{if } b_g \neq b_1, b_g \neq b_{m2} \end{cases}$$

where c_1 is a constant, set to -1 in our experiment.

C. Decision making

Several planning modules constitute the decision-making process: a combination of social guidance, shared plans negotiation, artificial curiosity and autonomous behaviors are entailed at this level. At this stage, the choice and interconnection of these agents is hard-coded, and basically dependent on the experiment to perform. Differently from [7], where social guidance was restricted to the mere execution of commands received from the caregiver, in this paper the robot takes its decisions autonomously based on intrinsic motivation and curiosity. Being the novel contribution of this paper to the CA, we hereinafter describe in detail the *Socially Guided Intrinsic Motivation with Active Choice of Teacher and Strategy* (SGIM-ACTS) algorithm. Our learner improves its estimation L of M to maximize $I = \sum_a P(a)\gamma(a)$ both by self-exploring A and B spaces. It generates new perception samples by manipulating the objects and by asking for help to a caregiver, who hands the objects to the robot. When an object is placed on the table, a rgb-d image $a \in A$ is retrieved at each step. SGIM-ACTS learns by episodes during which it actively chooses both an object $b \in B$ to learn to recognize and a learning strategy σ between: pushing the object, taking and dropping the object or asking the caregiver to manipulate the object. For each object b it has decided to explore, it also decides the strategy σ which maximizes its *competence progress* or $interest$, defined as *the local competence progress, over a sliding time window of δ for an object b with strategy σ at cost $\kappa(\sigma)$*. If the competence measured for object b with strategy σ constitute the list $R(b,\sigma) = \{\gamma_1, ...\gamma_N\}$:

$$interest(b,\sigma) = \frac{1}{\kappa(\sigma)} \frac{\left| \left(\sum_{j=N-\delta}^{N-\frac{\delta}{2}} \gamma_j \right) - \left(\sum_{j=N-\frac{\delta}{2}}^{N} \gamma_j \right) \right|}{\delta} \quad (2)$$

This strategy enables the learner to generate new samples a in subspaces of A. The SGIM-ACTS learner explores preferentially objects where it makes progress the fastest. It samples views from the object to improve its vision system, re-using and optimizing the recognition algorithm built through its different exploration strategies.

This behavioral description of SGIM-ACTS is completed in the next section by the description of its architecture.

D. SGIM-ACTS Architecture

SGIM-ACTS is an algorithm based on interactive learning and intrinsic motivation. It learns to recognise different objects by actively choosing which object $b \in B$ to focus on, and which learning strategy σ to adopt to learn local inverse and forward models. Its architecture is separated into two levels as described in Alg. IV.1:

Algorithm IV.1 SGIM-ACTS

Input: $s_1, s_2, ...$: available strategies with cost κ_i.
Initialization: $\mathcal{R} \leftarrow$ singleton $\{B\}$.
Initialization: $Memo \leftarrow$ empty episodic memory.
loop
  $\sigma_i, b_g \leftarrow$ **Select Label and Strategy**$(\mathcal{R})$
  repeat
    if σ_i is a Social Guidance learning strategy **then**
      $(a_r, b_r, b_g) \leftarrow$ Interact with caregiver with strategy σ_i.
    else if σ_i is an Auton. Exploration learning strategy **then**
      $(a_r, b_r, b_g) \leftarrow$ Perform action with strategy σ_i.
    end if
    Update L and L^{-1} with (a_r, b_r).
    $\gamma \leftarrow$ Competence for b_g
  until end of trials for the same object
  $\mathcal{R} \leftarrow$ Update **Goal Interest Mapping**$(\mathcal{R}, Memo, b_g, \gamma)$
end loop

Algorithm IV.2 $[\mathcal{R}]$ **= GoalInterestMapping**$(\mathcal{R}, Memo, b, \gamma)$

input: b_i: set of labels and corresponding $interest(b, \sigma)$ for each strategy σ.
input: δ : a time window used to compute the interest.
Add γ to $R(b, \sigma)$, the list of competence measures for $b \in B$ with strategy σ.
Compute the new value of competence progress of b:

$$interest(b, \sigma) = \frac{1}{\kappa(\sigma)} \frac{\left| \left(\sum_{j=N-\delta}^{N-\frac{\delta}{2}} \gamma_j \right) - \left(\sum_{j=N-\frac{\delta}{2}}^{N} \gamma_j \right) \right|}{\delta}$$

return $\mathcal{R}$ the set of all $R(b, \sigma)$

Algorithm IV.3 $[\sigma, b_g]$ **= SelectLabelAndStrategy**$(\mathcal{R}$ $)$

input: $\mathcal{R}$: set of regions R_n and corresponding $interest_{R_n}(\sigma)$ for each strategy σ.
parameters: $0 \leq p_1 \leq$: probability for random mode.
$p \leftarrow$ random value between 0 and 1.
if $p < p_1$ **then**
  Ensure a minimum of exploration, i.e. :
  Choose σ and $b_g \in B$ randomly
else
  Focus on areas of highest competence progress, i.e. :
  $\forall(\sigma, n)$, $P_n(\sigma) \leftarrow \frac{interest_{R_n}(\sigma) - \mathbf{min}(interest_{R_i})}{\sum_{i=1}^{|R_n|} interest_{R_i}(\sigma) - \mathbf{min}(interest_{R_i})}$
  $(b, \sigma) \leftarrow argmax_{n,\sigma} P_n(\sigma)$
end if
return (b, σ)

- A *Strategy and Label Space Exploration* level which decides actively which object b_g to set as a goal, which strategy σ to adopt, and which object to manipulate (*Select Label and Strategy*). To motivate its choice, it maps B in terms of interest level for each strategy (*Goal Interest Mapping*) as detailed in Alg. IV.2.
- A *State Space Exploration* level that explores A, according to the object b_g and strategy σ chosen by the Strategy and Label Space Exploration level. With each chosen strategy, different samples (a_r, b_r) are generated to minimise γ, while improving its estimation of $M(a_r)$ which it can use later on to reach other goals. It finally returns the competence measure $\gamma(b_g)$ to the Strategy and Label Space Exploration level.

V. EXPERIMENTAL SCENARIO

A. Experimental Platform

Experiments are carried out with iCub, a 53 DOF full-body humanoid robot [50]. The whole upper-body has been used in our experiments: head, torso, arms and hands, for a total of 41 DOF. Thanks to proximal force sensing, the main joints (arms, torso) are compliant [51]. All software modules used in the experiments of Section VI belong to the MACSi software architecture [7].

B. Experimental Protocol

An experiment consists of a sequence of interactions with an object. The robot can decide to perform the actions autonomously or to ask the caregiver. Precisely, the curiosity system chooses an object to manipulate and a strategy among the following:

- push the object
- take the object, lift it, and let it fall on the table
- ask the human to manipulate a specified object

In an experiment, the human first presents and labels each of the objects one by one and lets the iCub manipulate them. At any time, the robot can ask the caregiver to switch to a specific object. It thus knows which object b it is manipulating. During the execution of the action, the vision processing system is inactive. When the action is completed, and the object is generally immobile on the table (notably, in a different pose), the vision system is triggered. After each manipulation, the robot tests which object it associates with the new object image, computes a confidence measure on its capability to recognize the object, and sends the evaluation results to the curiosity system, before gathering new knowledge about the object and updating its recognition model L with the known object b. Depending on the progress, the curiosity system decides the next action to trigger. The objects used in the experiments are shown in Fig. 4: remarkably, some objects are more "challenging" to recognize because their appearance is different depending on their side (generally their color, but also their size - in the case of cubes and bear):

- a gray dog-shaped stuffed toy. Its color and shape are quite different from the others, and it is therefore easy to recognize it.
- a purple and blue colored ball. The colors and shape are quite different from the other objects, so it is quite easy to distinguish. However, because the two sides of the ball are of different colors, more samples are required to associate the different views to the ball.
- a red teddy bear. Its color and shape are quite easy to recognize, but it can be confused with the cubes which also have red parts.
- a yellow car. This toy offers numerous views depending on its orientation and position on the table. We expect such a toy to arouse the interest of an agent because of its rich "perceptive affordance". Moreover, the toy has the same color as parts of the cubes, and almost the same shape as some views of the cubes (when a lateral view shows only the yellow cubes). Thus its classification may be difficult.
- a patchwork of yellow-red-green cubes. This toy also offers numerous views depending on its orientation and position. This object is the most tricky to recognize as it can be confused with both the car and the teddy bear.

C. Evaluation of the Learning process

To evaluate the efficiency of our algorithm, we compare our SGIM-ACTS with the random algorithm where the agent would choose at each episode a random object and a random strategy. To evaluate the efficiency of each algorithm, we freeze the learning process after each episode and evaluate the classification accuracy on an image database, made up of 64 images of each object in different positions and orientations built independently from the learning process (see Fig. 5 for a sample).

Fig. 4: The objects used during the experiments: some colored cubes, a yellow car, a grey dog, a violet/blue ball, a red bear. Left and right images respectively show the front/rear sides of the objects.

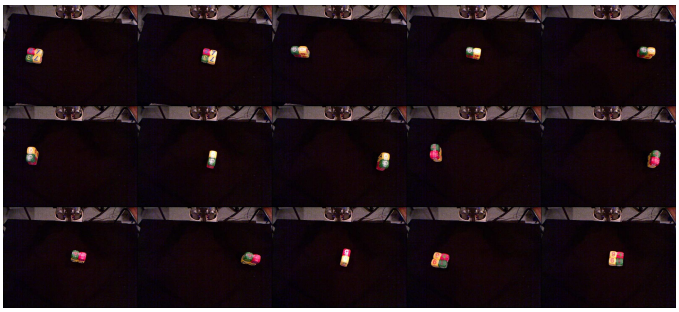


Fig. 5: A portion of the database of objects views used for evaluating the recognition performance: precisely, the images related to the cubes.

VI. EXPERIMENTAL RESULTS

We conducted the experiments with each of the algorithms (SGIM-ACTS and random) under two conditions: with an unbiased teacher who shows objects to the learner under different angles; and with a biased teacher who always shows the same view of each object. We plot results for each case of exploration, strategy and teacher, detailing the learning performance separated by object. We plot the f-measure (i.e. the harmonic mean of precision and recall [52]) and the number of images correctly recognized in the evaluation database.

As shown in Fig. 6 the progress in recognition is better with SGIM-ACTS than with random exploration, for both teachers. At the end of the experiments, the SGIM-ACTS learner is able to correctly recognize the objects in 57 over 64 images, against 50 in the case of the random learner.

Fig. 7 plots how well the system can distinguish objects, and which objects it maniplulates for example experiments under different conditions. We can see in Fig. 7a and 7b that the random learner often switches objects, and explores equally all objects, while the SGIM-ACTS learner focuses on objects for longer periods of time. We note that SGIM-ACTS

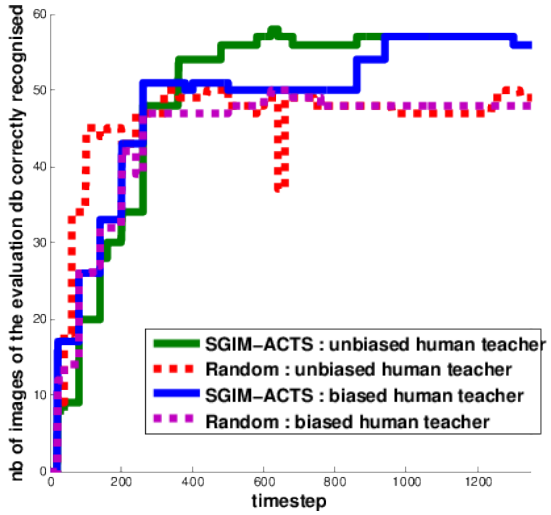
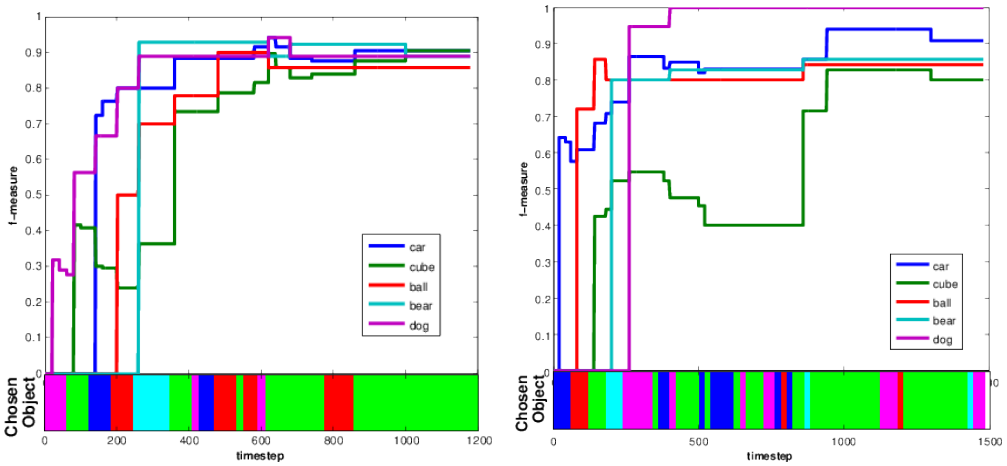


Fig. 6: SGIM-ACTS vs Random: recognition performance, i.e. the number of images of the evaluation database correctly recognized by the two exploration strategies with two different behaviors of the teacher (see text).

manipulates more the cubes, especially when its competence progress increases. Indeed, as stated above, the cube is the most complex of objects because it offers very different views due to its various colors, but also because it can easily be confused with other objects that bear the same colors. Manipulating it brings every time more information about the object since their appearance changes substantially depending on the action (a frontal view consists of four cubes, while a lateral view consists of two cubes only, and depending on the side it could be yellow or red/green), and improves its discrimination from other objects. The iCub has spent 54% and 51% of its time learning about cubes with SGIM-ACTS for both teachers. The system thus allocates more time for the difficulties.

Overall, the iCub focuses its attention on complex objects, asking human intervention or manipulating autonomously to improve its recognition capability. Fig. 7d clearly illustrates this mechanism: the red bear (cyan line) is easily recognized, hence the robot does not ask again to interact with the object once it is learnt; conversely, the cubes (green line) are difficult to recognize, hence the robot focuses more on them.

Conversely, as shown in Fig. 7b, in the "random" case the robot does not focus on any particular object. Hence, the recognition performance at the end of the experiment is worse, because the "difficult" objects (such as the cubes - green line) are not sufficiently explored. Furthermore, the SGIM-ACTS algorithm is robust to the quality of the teaching, for the recognition performance is high in both cases. Whether the teacher helps by showing new views of objects and bringing new information, the learner improves its discrimination of objects. This is to contrast with the case of the random algorithm which is dependent on the teacher. We can see that the f-measures of Fig. 7a are lower than in Fig. 7b. SGIM-ACTS is able to recognise how profitable a teacher can be, and choose to take advantage of him or not. It is able to use the human to perform actions on objects and generate datasets that it can not perform by itself.



(a) SGIM-ACTS with unbiased teacher

(b) SGIM-ACTS with biased teacher

(c) Random exploration with unbiased teacher

(d) Random exploration with biased teacher

Fig. 7: f-measure on the evaluation database, with respect to time. The bottom part of the plot shows the manipulated object at each timestep.

In conclusion, on the long-term SGIM-ACTS strategy yields better performances, because it facilitates learning all objects dedicating more time and efforts to the complicated objects.

A. *Video and code*

The software for the architecture and the experiments is available under GPL license at http://macsi.isir.upmc.fr. A video demonstrating the experiments of the paper is available at https://www.youtube.com/iCubParis.

VII. CONCLUSIONS

In this paper we described a method to choose actively data collection strategy in order to learn fast how to recognize object, which exploits curiosity to guide exploration and manipulation, such that the robot can improve its knowledge of objects in an autonomous and efficient way. The autonomous behavior driven by intrinsic motivation has been fruitfully integrated in the MACSi Cognitive Architecture. Experimental results show the effectiveness of our approach: the humanoid iCub is now capable of deciding autonomously which actions must be performed on objects in order to improve its knowledge, requiring a minimal assistance from its caregiver. This work constitutes the base for forthcoming research in autonomous learning of affordances. The next step in the evolution of the cognitive architecture will be to integrate in the current framework a high-level representation of actions, objects, features, effects, so that the robot can gradually make progress by trying to discover relationships among these elements, setting the base for affordance learning.

REFERENCES

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

[2] 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.

[3] M. Rudinac, G. Koostra, D. Kragic, and P. Jonker, "Learning and recognition of objects inspired by early cognition," in *Proc. IEEE/RSJ Int. Conf. on Intelligents Robots and Systems*, 2012.

[4] H. van Hoof, O. Kroemer, H. Ben Amor, and J. Peters, "Maximally informative interaction learning for scene exploration," in *Proc. IEEE/RSJ Conf. on Intelligent Robots and Systems*, 2012.

[5] P. Fitzpatrick, G. Metta, L. Natale, S. Rao, and G. Sandini, "Learning about objects through action - initial steps towards artificial cognition," in *IEEE Int. Conf. on Robotics and Automation*, 2003, pp. 3140–3145.

[6] B. Browatzki, V. Tikhanoff, G. Metta, H. Bulthoff, and C. Wallraven, "Active object recognition on a humanoid robot," in *IEEE International Conference on Robotics and Automation*, 2012, pp. 2021–2028.

[7] S. Ivaldi, N. Lyubova, D. Gérardeaux-Viret, A. Droniou, S. M. Anzalone, M. Chetouani, D. Filliat, and O. Sigaud, "Perception and human interaction for developmental learning of objects and affordances," in *Proc. IEEE-RAS Int. Conf. on Humanoid Robots*, Osaka, Japan, 2012.

[8] P.-Y. Oudeyer and F. Kaplan, "How can we define intrinsic motivations ?" in *8th Int. Conf. on Epigenetic Robotics*, 2008.

[9] M. Lopes and P.-Y. 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.

[10] C. L. Nehaniv and K. Dautenhahn, *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions*. Cambridge: Cambridge Univ. Press, March 2007.

[11] M. Lopes and P. Oudeyer, "The strategic student approach for life-long exploration and learning," 2012.

[12] R. Reichart, K. Tomanek, U. Hahn, and A. Rappoport, "Multi-task active learning for linguistic annotations," *ACL*, 2008.

[13] G. Qi, X. Hua, Y. Rui, J. Tang, and H. Zhang, "Two-dimensional active learning for image classification," *Computer Vision and Pattern Recognition*, 2008.

[14] P.-Y. Oudeyer, F. Kaplan, V. Hafner, and A. Whyte, "The playground experiment: Task-independent development of a curious robot," in *Symp. on Developmental Robotics*. AAAI, 2005, pp. 42–47.

[15] A. G. Barto, S. Singh, and N. Chentanez, "Intrinsically motivated learning of hierarchical collections of skills," in *IEEE International Conference on Development and Learning*, 2004.

[16] A. Baranes and P.-Y. Oudeyer, "Active learning of inverse models with intrinsically motivated goal exploration in robots," *Robotics and Autonomous Systems*, vol. 61, no. 1, pp. 49–73, 2013.

[17] Y. Baram, R. El-Yaniv, and K. Luz, "Online choice of active learning algorithms," *The Journal of Machine Learning Research,*, vol. 5, pp. 255–291, 2004.

[18] A. Rebguns, D. Ford, and I. Fasel, "Infomax control for acoustic exploration of objects by a mobile robot," in *AAAI Conference on Artificial Intelligence*, 2011, pp. 22–28.

[19] S. Calinon, *Robot Programming by Demonstration: A Probabilistic Approach*. EPFL/CRC Press, 2009.

[20] S. Calinon, G. F., and A. Billard, "On learning, representing and generalizing a task in a humanoid robot," *IEEE Transactions on Systems, Man and Cybernetics, Part B.*, 2007.

[21] J. Peters and S. Schaal, "Reinforcement learning of motor skills with policy gradients," *Neural Networks*, vol. 21, no. 4, pp. 682–697, 2008.

[22] B. Blumberg, M. Downie, Y. Ivanov, M. Berlin, M. P. Johnson, and B. Tomlinson, "Integrated learning for interactive synthetic characters," *ACM Trans. Graph.*, vol. 21, pp. 417–426, July 2002.

[23] F. Kaplan, P.-Y. Oudeyer, E. Kubinyi, and A. Miklosi, "Robotic clicker training," *Robotics and Autonomous Systems*, vol. 38, no. 3-4, pp. 197–206, 2002.

[24] J. Clouse and P. Utgoff, "A teaching method for reinforcement learning,," *9th Int. Conf. on Machine Learning*, 1992.

[25] W. Smart and L. Kaelbling, "Effective reinforcement learning for mobile robots,," *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 3404–3410., 2002.

[26] S. Chernova and M. Veloso, "Interactive policy learning through confidence-based autonomy," *Journal of Artificial Intelligence Research*, vol. 34, 2009.

[27] A. L. Thomaz and C. Breazeal, "Experiments in socially guided exploration: Lessons learned in building robots that learn with and without human teachers," *Connection Science*, vol. 20 Special Issue on Social Learning in Embodied Agents, no. 2,3, pp. 91–110, 2008.

[28] D. H. Grollman and O. C. Jenkins, "Incremental learning of subtasks from unsegmented demonstration," 2010.

[29] M. Lopes, F. Melo, and L. Montesano, "Active learning for reward estimation in inverse reinforcement learning," in *European Conference on Machine Learning*, 2009.

[30] S. M. Nguyen and P.-Y. Oudeyer, "Interactive learning gives the tempo to an intrinsically motivated robot learner," in *IEEE-RAS International Conference on Humanoid Robots*, 2012.

[31] ——, "Properties for efficient demonstrations to a socially guided intrinsically motivated learner," in *21st IEEE International Symposium on Robot and Human Interactive Communication*, 2012.

[32] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen, "Autonomous mental development by robots and animals," *Science*, vol. 291, no. 599-600, 2001.

[33] E. Deci and R. M. Ryan, *Intrinsic Motivation and self-determination in human behavior*. New York: Plenum Press, 1985.

[34] R. M. Ryan and E. L. Deci, "Intrinsic and extrinsic motivations: Classic definitions and new directions," *Contemporary Educational Psychology*, vol. 25, no. 1, pp. 54 – 67, 2000.

[35] P.-Y. Oudeyer, F. Kaplan, and V. Hafner, "Intrinsic motivation systems for autonomous mental development," *IEEE Transactions on Evolutionary Computation*, vol. 11(2), pp. pp. 265–286, 2007.

[36] J. Schmidhuber, "Formal theory of creativity, fun, and intrinsic motivation (1990-2010)," *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 3, pp. 230–247, 2010.

[37] ——, "Curious model-building control systems," in *Proc. Int. Joint Conf. Neural Netw.*, vol. 2, 1991, pp. 1458–1463.

[38] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *Journal of Artificial Intelligence Research*, vol. 4, pp. 129–145, 1996.

[39] N. Roy and A. McCallum, "Towards optimal active learning through sampling estimation of error reduction," in *Proc. 18th Int. Conf. Mach. Learn.*, vol. 1, 2001, pp. 143–160.

[40] S. M. Nguyen, A. Baranes, and P.-Y. Oudeyer, "Bootstrapping intrinsically motivated learning with human demonstrations," in *IEEE Int. Conf. on Development and Learning*, 2011.

[41] 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.

[42] 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.

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

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

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

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

[47] 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.

[48] D. Filliat, "A visual bag of words method for interactive qualitative localization and mapping," in *IEEE Int. Conf. on Robotics and Automation*, 2007, pp. 3921–3926.

[49] J. Sivic and A. Zisserman, "Video google: Text retrieval approach to object matching in videos," in *Int. Conf. on Computer Vision*, vol. 2, 2003, pp. 1470–1477.

[50] 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.

[51] 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 *IEEE-RAS Int. Conf. on Humanoid Robots*, 2011, pp. 521–528.

[52] C. J. van Rijsbergen, *Information Retrieval*. Butterworth, 1979.