Human Robot Interaction can Boost Robot's Affordance Learning: A Proof of Concept

Amit Kumar Pandey
Aldebaran, A-Lab, France
akpandey@aldebaran.com

Rodolphe Gelin
Aldebaran, A-Lab, France
rgelin@aldebaran.com

***Abstract*—Affordance, being one of the key building blocks behind how we interact with the environment, is also studied widely in robotics from different perspectives, for navigation, for task planning, etc. Therefore, the study is mostly focused on affordances of individual objects and for robot environment interaction, and such affordances have been mostly perceived through vision and physical interaction. However, in a human centered environment, for a robot to be socially intelligent and exhibit more natural interaction behavior, it should be able to learn affordances also through day-to-day verbal interaction and that too from the perspective of what does the presence of a specific set of objects affords to provide. In this paper, we will present the novel idea of verbal interaction based multi-object affordance learning and a framework to achieve that. Further, an instantiation of the framework on the real robot within office context is analyzed. Some of the potential future works and applications, such as fusing with activity pattern and interaction grounding will be briefly discussed.**

I. Introduction

Reasoning about affordance - what something can offer or afford to do, i.e. the action possibility - is important to shape our day-to-day interaction with the environment and with others. Hence, the ability to learn affordance is one of the basic ingredients for developing complex behaviors among primates, including human.

In cognitive psychology, Gibson, in his pioneering work on affordance [1], refers affordance as what an object offers, as all action possibilities, independent of the agent's ability to recognize them. Whereas, seeing through Human Computer Interaction (HCI) perspective, Norman [2] tightly couples affordances with past knowledge and experience; hence, sees affordance as perceived and actual properties of the things. Irrespective of shifts in the interpretation, affordance is an important aspect in developmental perceptual learning and action differentiation and selection, [3]. Affordances has been further argued as basic component for socio-cognitive development as a central organizing construct for action, [4], and to be at the root of embodied cognition, in the light of discovery of canonical neurons, [5].

In robotics, affordance has been viewed from different perspectives: agent, observer and environment; hence, the definition depends upon the perspective, [6]. Affordance have been used in robotics for tool use [7], for traversability [8] of the robot, to learn action selection [9], etc. Affordance has already been shown to be an important component for cognitive embodiment in robots, such as cueing and recognition of affordance-based visual entities for robot control [10], agent-object affordance-based anticipation of human activity for reactive robot responses, [11], internal rehearsal based learning of affordance relations to predict the outcomes of its behaviors before executing them, [12], and so on. In [13] through the introduction of affordance graph, the notion of affordance has been further enriched by incorporating other agents and efforts.

In this paper we focus on what we call as *collective affordance*, in the sense, not only using a single object as was the case in the natural environment, but to exploit the idea that a set of objects in the human environment collectively indicates affordance possibility, e.g. presence of a Monitor, Keyboard and Mouse indicates the action possibility of watching movies, checking emails, etc. Hence, we are using the typical notion of affordance, but extending the scope for the objects for human centered environment. And this requires a new kind of reasoning and affordance learning mechanism, based on multi-object reasoning.

Recently, there have been some works in the direction of reasoning about multiple objects from affordance perspective. For example, in [14], using a first order logic, the robot learns manipulation possibility based on spatial relation (relative distance, orientation angle and contact) between object. This affordance model is obtained by interaction with the environment and basically captures the effect of robot's hand motion on a main object and a secondary object, which may interact with the main object through the robot's action. In [15], the notion of multiple objects has been explored from the perspective of predicting effect of actions on paired object. For example, through a bootstrapping process, affordance-features (such as rollability, pushability, etc.) are fused with basic-features (such as size, shape, etc.) to learn and predict more complex affordances involving two objects, such as *stackability* affordances. In [16], it is shown that considering spatio-temporal relationships between objects adds to the functional descriptor for objects to reason about human activities using object context. The key idea exploited in this is, the same object can afford different action possibility when associated with different other objects and depends upon the context. Supporting such argument, we further argue that for human made objects in the human centered environment it is not sufficient to learn or infer about affordance i.e. the action possibilities based on a single object. For example, just the presence of

This work is funded by Romeo2 project, (http://www.projetromeo.com), BPIFrance in the framework of the Structuring Projects of Competitiveness Clusters (PSPC)

978-1-4673-7509-2/15/$31.00

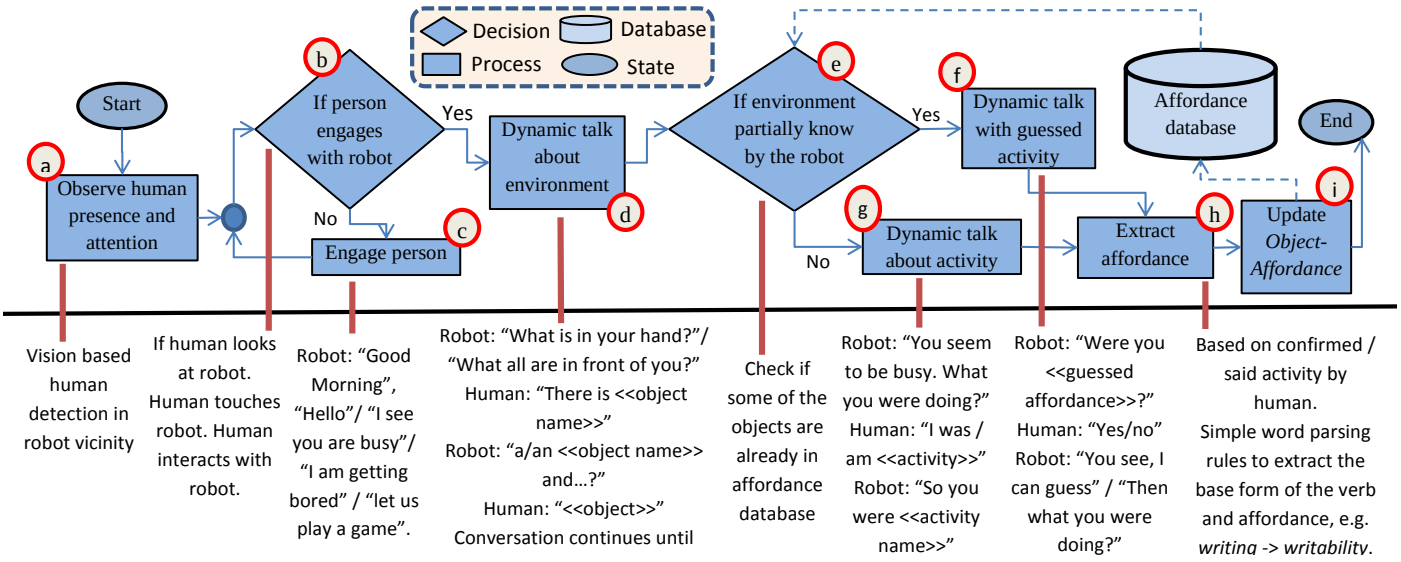

Fig. 1: Outline of the interactive affordance learning framework (top half), and the main aspects of the verbal interaction based instantiation presented in this paper (bottom half).

a notebook does not indicate the affordance of *writability*. There is a need of a pen nearby. Hence, the notebook and the pen collectively indicate an affordance, which they cannot individually. Hence, there is a need to elevate the notion of affordance to multi-object for human centered environment. In this paper, we will introduce a framework for learning such collective multi-object affordance.

From the point of view of the medium of gathering affordance related information, mainly vision based data, such as [17], and by manipulation trials, such as [14] are used. In this paper we will add the dimension of learning affordance through verbal interaction. Verbal interaction has already been shown to be useful for affordance-based human robot interaction, such as linguistic instruction from the human has been grounded based on the affordances of the objects, i.e. the services they provide, [18]. The studies in cognitive neuroscience, such as [19], suggest that language and verbs that are present, reflect some characteristics of action organization and evoke simulations, which in turn activate affordances, more specifically functional affordances. To the best of our knowledge, in robotics, natural language based human-robot interaction has not been exploited for learning functional affordances. The novelties of this paper are to introduce the idea of utilizing human-robot verbal interaction for learning affordance, that too at the level of multi-object joint affordance, and to demonstrate the proof of concept through human-robot interaction over a period of time in an office setup. Another motivation behind the current work is to facilitate affordance learning in the situations when the robot cannot directly perceive the environment, the objects and the activities. This might be because of various reasons, such as the limitations of vision based perception, remote human-robot interaction, etc.

In the next section we present our framework of verbal interaction based affordance learning and the *m-estimate* based approach to take into account the notion of experience. Section III presents instantiation of the framework with Nao robot in an office scenario. Followed by this, section IV presents some of the interesting results of two types of associations of affordance: *Object-Object association* for a particular affordance and *Affordance-Object* association for different affordances. In section V some of the benefits and the potential applications of the framework and the system is discussed, followed by conclusion and future work in section VI.

II. Framework for Interaction Based Multi-Object Affordance Learning

A. Overview

The top part of figure 1 outlines our framework of the verbal interaction based affordance learning. In the bottom half of the figure, the connecting text below each of these blocks shows the instantiation aspects of the framework for the current implementation, which will be discussed in section III.

As shown, the robot begins by observing the environment, *block a*. Then at appropriate moments, the robot tries to engage the person who is present in the environment, if the person is not already engaged in interaction with the robot (*blocks b and c*). Once an engaged person is detected, the robot starts affordance discovery oriented verbal interaction with the human, which includes talking about the environment in the vicinity of the human (*block e*). Such interaction will be designed specifically with some dynamic parameters to extract information about objects from the human speech. Once the robot has some information about the objects that are present in the human's vicinity, it checks in its existing affordance database for information about those objects (*block e*). If object is found in the database, some affordance based activities are guesses to synthesize some dynamic interaction with the hu-

man (*block f*). Otherwise, some other dialog for interaction are triggered to get the information about the activity in which the human was engaged (*block g*). Through either of the channels, the current engaged activity is obtained and used to extract the associated affordance (*block h*). The extracted affordance is then inserted in the affordance database with the information about objects and other additional information, such as time, user id, etc. (*block i*). Once a particular affordance has been obtained and inserted in the database, one *interaction instance* is said to be finished.

It is interesting to note that the framework not only brings the notion of experience for internal processing, but also for making the interaction more natural and dynamic. As shown in figure 1, it does not always ask directly about the activity. Based on experience it tried to guess a couple of activities and if the guess is correct it simply updates the knowledge, otherwise tried to ask for the activity directly.

One of the basic requirements of the learning system is to associate the object with an affordance and with another object based on experience over multiple interaction instances. To achieve this, instead of using simple probability based approach, we use *m-estimate* based reasoning, as discussed next.

B. m-estimate based reasoning

This paper is not intended to contribute in machine learning techniques from the point of view of incremental learning of associations between the members of a set of symbols. To fulfill the requirement of incorporating the notion of experience in such learning, instead of using simple probability based reasoning, we chose to use *m-estimate*, which has been shown to be useful for rule evaluation, [20], to avoid premature conclusions [21] and as a tool to learn explanation tree for task understanding from demonstration, [22]. For continuity this subsection outlines the basic idea of *m-estimate*.

Let us assume for a particular object *Obj*, the affordance *af* has been assigned based on *n* number of interactions, out of a total of *N* interactions in which *Obj* has been mentioned. Within the *m-estimate* framework, the likelihood (i.e. the measure of the extent to which a sample provides support for particular affordance) of associating the same affordance *af* to object *Obj* during the next interaction in which object *Obj* will be mentioned, is given as:

$$Q^{af}_{Obj}(n, N) = \frac{n + a}{N + a + b} \tag{1}$$

Where, $a > 0, b > 0, a + b = m \; and \; a = m \times P_{af}$

m is domain dependent, and could also be used to include noise, [23]. From eq. 1 following properties could be deduced:

$$Q^{af}_{Obj}(0, 0) = P_{af,Obj} > 0 \tag{2}$$

$$Q^{af}_{Obj}(0, N) = \frac{a}{N + a + b} > 0 \tag{3}$$

$$Q^{af}_{Obj}(N, N) = \frac{N + a}{N + a + b} < 1 \tag{4}$$

$$Q^{af}_{Obj}(N + 1, N + 1) > Q^{af}_{Obj}(N, N) \tag{5}$$

$$Q^{af}_{Obj}(0, N) < Q^{af}_{Obj}(0, N + 1) \tag{6}$$

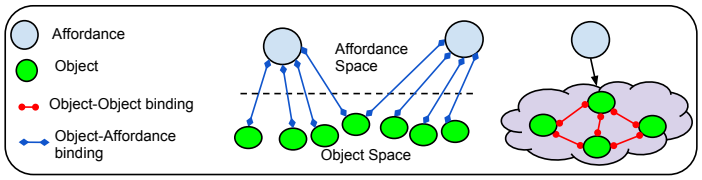

Fig. 2: Types of Object and Affordance bindings. In this paper we capture verbal interaction and experience based binding between (i) object and affordances (middle) and (ii) object pairs for a particular affordance (right).

Above properties show that *m-estimate* does not assume a close world in the sense if an affordance *af* for object *Obj* has not been observed, it does not mean that likelihood of the existence of *af* for that object is NULL (expressions (2), (3)). Eq. 2 also indicates that $P_{af,Obj}$ can be seen as the prior probability of *af* for object *Obj*. On the other hand, it takes into account the likelihood of novel interactions. As evident from eq. 4, if always the same affordance *af* has been observed, that too will not be accepted as universal rule that *Obj* will always have the affordance *af*. Eq. 5 shows that the likelihood will be more if observed in more number of interactions. Eq. 6 reflects that for a never observed affordance *af* for object *Obj*, the likelihood of also *not* being observed in the future will be less if *Obj* has been appeared in less number of interactions. These properties facilitate to incorporate the notion of experience and allow lifelong refinement of the learned affordance concept.

One acceptable instantiation of *m-estimate* is using *Laplace's law of succession*. This states that if in the sample of *N* trials, there were *n* successes, the likelihood of the next trial being successful is *(n+1)/(N+2)*, assuming that the initial distributions of success and failure are uniform. With the similar initial assumption, we also use *a=1* and *a+b=2* for *m-estimate* of eq. 1.

C. Experience based Object-Affordance binding

One of the important notions to capture for affordance learning is, how strongly an object is associated with a particular affordance. Let us say a is an affordance belonging to the affordance space A, see figure 2, and o is an object belonging to the object space O. We define $Q_{OA} = Q^{af=a}_{Obj=o}$ to refer to evidence of object o offering affordance a, see eq. 1. An object can be associated with multiple affordances, as shown in the middle subfigure of figure 2. Hence, Q_{OA} reflects the joint evidence observed across different *interaction instances*[1] across different affordances.

D. Experience based Object-Object binding

Another important aspect is to capture the necessity of *togetherness* of objects for a particular affordance i.e. the notion of multi-object joint affordance. We define $Q_{OOA} = Q^{af=a}_{Obj=o1 \wedge o2}$, which captures the evidence of occurring objects $o1$ and $o2$ together in an instance of interaction when the resulting affordance was a. As *m-estimate* captures the notion of experience, hence repetition of pair $< o1, o2 >$ over

[1]As mentioned one interaction instance is said to be starting from engaging a person to entering the data about one extracted affordance into the database.

multiple interactions for the same affordance will provide stronger evidence of relevance of both being together for that affordance. Each single connecting red line with circular ends of the right subfigure of figure 2 shows this object-object binding. Hence, Q_{OOA} reflects the joint evidence observed across those different interaction instances, which concluded into the discovery of a same affordance.

III. Instantiation

This section provides the technical details of one instantiation of the framework on a real robot, *Nao*, interacting with a person over a period of five working days. One of the motivations behind the present implementation is to make the interaction dynamic and interesting in a game like situation for the person to have some break and encourage to interact with the robot. Therefore, there have been various branches and variations in the dialog pattern of the robot, also the robot has not been programmed to directly ask the question "what are you doing and what are the objects used for that". Therefore, there might be some false positives, but with this the main objective to facilitate extraction of affordance information from day-to-day natural interaction will be achieved. Such false positives could also be used to make the interaction more dynamic, triggering keywords for further interactions.

1) Robot and development environment: Nao robot has been used as platform for the experiment. *Choregraphe* [24] has been used as the development environment, within the *Naoqi* framework of the robot's operating system.

2) Perception of person: For detection of human and detection of human engagement with the robot (as shown in the bottom half of fig. 1), vision and touch based modules of Naoqi are used. It has methods for detecting presence and leaving of people, gaze analysis, person sitting, etc. See the online documentation[2] for details.

3) Dialog engine for verbal interaction and dialog for affordance discovery: As the information about affordance is extracted through verbal interaction, the bottom half of fig. 1 also shows some examples of such human-robot dialog based interaction. Through such natural verbal interaction, the robot extracts information about objects in the environment and the activity the person was involved in. Hence, it greatly reduces the needs of vision based perception of objects and activities. Nevertheless, such vision based recognition system will add to the interaction and affordance learning.

For creating dialog based interaction system, we are using *ALDialog* and *qichat* modules[3] of Naoqi. It provides various functionalities to create and shape natural interaction, such as creation of concept, topics etc. **Concept**: Concept is a list of words and/or phrases that refer to one idea. For example, a list of countries, a list of names, synonyms of a word. They can be used both in the human input or the robot output. We use it to extract the list of actions from the human speech. **Topic**: A Topic is a script box (or file) containing **Rule**. A rule associates a human input (what the human says) with a relevant robot output (what the robot answers). For example a simple rule to capture the names of objects from human speech can be:

[2]http://doc.aldebaran.com/2-1/naoqi/peopleperception/index.html
[3]http://doc.aldebaran.com/2-1/naoqi/audio/dialog/aldialog.html

Fig. 3: An illustrative scenario of the *Nao* robot on an office desk, interacting with the human.

u:(There is a _) Ok, so there is a $1 and what else $response=$1.*

The part within () states the pattern to capture from the human speech. The item followed by "*There is a*" in the human speech will be stored in argument *$1*. The second part of the rule, which in this example is *Ok, so there is a $1 and what else*, defines the robot's response by using the name of the object stored in *$1* obtained from the human speech. And the third part *$response=$1* assigns the object name stored in *$1* to a variable *response*, which will be visible to external modules for further processing.

4) Database: We are using *SQlite*[4] SQL database engine. The table populated through the interaction has the following fields from the affordance learning point of view: *Person ID, Object, Affordance, Time*.

IV. Experimental Result and Analysis

We have collected data through *Nao* robot in an office scenario (see the example scenario of figure 3). The interaction and the learning of affordance have taken place over a period of five weekdays with the same person. The robot has been turned on from 9:30 AM to 6:30 PM during the office hour. To avoid any over-saturation due to a highly repeating nature of the environment and the work pattern, we added some rules to maintain temporal distance in two successive interaction instances. Therefore, in one working day there were *5-6* interactions at most at somewhat equal intervals, such as early office hour, early office hour+2 hours, lunch hours, after lunch work, coffee break hour, evening, etc. Therefore, the total number of interaction instances stored in the database over five days was *32*.

The user was familiar with the robot, and was aware about the purpose of such interaction instances, i.e. to provide the robot with the information asked during the interaction. One example of provided information by the user during such an interaction instance is, the user said that the objects in front of him are *monitor, keyboard, mouse, lunch-box and paper*, and the activity the human said to be engaged in is *checking-mail*. Hence, the robot stores in the database for this particular interaction instance the information as $< (monitor, keyboard, mouse, lunch-box, paper), check-mail, < current-time >>$.

[4]See the details at: http://www.sqlite.org/

TABLE I: Progressive affordance learning through interaction. *N*:Number of interaction instances (See *notes). Note that *N* does not show the number of all the interaction instances unless stated. Also note that for compactness, the suffix *-ability* from the affordance names have been dropped, for example instead of *writability*, *write* is used.

(a) Experience based Object-Object Joint Affordance

Affordance	N_f*	Object Pair	P	Q_{OOA}
check-email	1	<*monitor,mouse*>	1.0	0.66
write	1	<*pen,paper*>	1.0	0.66
check-email	3	<*monitor,mouse*>	1.0	0.80

*in which the **particular affordance** appeared

(b) Experience based avoiding false object-object binding

Affordance	Object Pair	N_f*	Q_{OOA}
check-mail	<*lunch-box,monitor*>	2	0.50
		5	0.28

*in which the **particular affordance** appeared

(c) Experience based avoiding false affordance binding

Object	N_o*	Affordance	Q_{OA}
laptop	1	*talk*	0.66
laptop	2	*eat*	0.50
		talk	0.50
laptop	4	*talk*	0.66
		eat	0.33

*in which the **particular Object** appeared

(d) Experience based Multi-Affordance binding

Object	N_o*	Affordance	Q_{OA}
Keyboard	4	*check-mail*	0.66
		programme	0.33
Keyboard	9	*check-mail*	0.45
		programme	0.27
		write	0.27
		eat	0.18

*in which the **particular Object** appeared

(g) Some Major *Affordance-Object-object* binding after long term interaction, Number of interaction instances ***N=32***

Affordance	Object Pair	Q_{OOA}
check-mail	<*keyboard - monitor* >	0.625
	<*keyboard - mouse* >	0.625
	<*monitor - mouse* >	0.625
	——	
	<*mobile - monitor* >	0.375
	<*pen - water bottle*>	0.125
work	<*monitor - mouse* >	0.714
	<*monitor - paper*>	0.428
	——	
	<*charger - mobile* >	0.142
	<*charger - paper* >	0.142
program	<*keyboard - monitor* >	0.83
	<*monitor - mouse*>	0.66
	——	
	<*mouse - pen* >	0.16
	<*notebook - pen* >	0.16

Affordance	Object Pair	Q_{OOA}
talk	<*laptop - mouse* >	0.66
	<*headphone - laptop*>	0.33
	<*mobile - monitor*>	0.33
	——	
	<*headphone - paper* >	0.16
	<*monitor - pen* >	0.16
eat	<*bottle-lunch-box*>	0.8
	<*bag-biscuit*>	0.4
	——	
	<*biscuit - monitor* >	0.2
write	<*notebook - pen* >	0.6
	<*monitor - pen*>	0.5
	<*keyboard - mouse*>	0.4
	<*paper - pen*>	0.4
	——	
	<*notebook - paper* >	0.1
	<*bottle - mouse* >	0.1

Since it is natural interaction based, so there will be ambiguity in the input from user about objects, activities, etc., but as such ambiguities will not be consistent, hence over a period of time, those will be refined. From the activity, simple parsing rules are used to extract affordance, such as the activity of *writing* results into the affordance *writability*.

Below we will highlight some of the results based on the experimental setup and the data obtained.

Notion of experience compared to simple probability: *m-estimate* captures the notion of experience with multiple interaction instances. Table Ia shows both the probability and *m-estimate* for different $< object, object, affordance >$ tuples. Note that after the first occurrence of affordance *check-mail*, for the object pair $< monitor, mouse >$ the *m-estimate* $Q_{OOA} = Q^{af=check-mail}_{Obj=monitor \wedge mouse}$ is 0.66, whereas the probability is 1. However, in the case when the affordance had occurred three times (last row of Ia) and the same objects pair has been observed, the Q_{OOA} becomes 0.80, but the probability remains the same as 1, because in all the three instances, the pair occurred. Hence, as the number of supporting interaction instances increases, the greater evidence for the same tuple is getting captured in the *m-estimate*. Moreover, after four interaction instances, $< paper, pen >$ appeared once for the only instance of $write$ affordance, and $< monitor, mouse >$ appeared for the three instances for $check - mail$ affordance, hence the higher evidence of $< monitor, mouse, check - mail >$ as compared to $< paper, pen, write >$ has been captured by *m-estimate*, whereas the probabilities are indifferent for both affordances.

Weakening of a false association over interaction: Table Ib shows an interesting example of how the false association can be observed in natural interaction based information extraction and how it can weaken over multiple interactions. For example, after two interaction instances for the affordance $check-mail$, the objects pair $< lunch-box, monitor >$ has also been associated with high *m-estimate* of $Q_{OOA} = 0.5$. This is due to those interaction instances when the user was taking a quick lunch on the work desk while also checking mails on the computer. However, as further 3 more instances of the affordance $check-mail$ haven been observed, during non-lunch hours, the *m-estimate* of the $< lunch-box, monitor >$ pair for that affordance has been reduced to 0.28, hence weakening the togetherness of the objects pair.

Table Ic shows an example of false higher association of *laptop* with *eat* affordance ($Q_{OA} = 0.5$) after second interaction instance corresponding to that affordance, which has been again reduced after four interaction instances.

Multi-affordance binding: Table Id shows how the learned affordances of a same object *keyboard* have been evolving over multiple interaction instances and how the associated *m-estimates* are getting adjusted by distributing the notion of experience over different affordances. Such, multi-affordance binding could be used to make the interaction more dynamic in the presented framework, for example to generate the content for the "dynamic talk with guessed activity" in block (f) of the framework in fig. 1.

Overall learned affordances: The table Ig shows the summary of some of the learned affordances and associated object-object pairs with the corresponding *m-estimates*. For each learned affordances, a few highest and lowest *m-estimate* based object pairs have been shown. It is evident that in long term interaction, such learned affordances are reasonable for human-robot interaction, as they are able to captured affordance-wise some of the most relevant objects pairs. However, it should not be assumed that the robot will not make mistakes, but with mistakes, its knowledge will evolve.

Note that we are deliberately not presenting any performance and time analysis of the system, as we did not notice any computation and resource hungry logs. The learning and the presented results are obtained in almost real-time and human robot verbal interaction has not been seen to be blocked because of any ongoing computation.

V. Potential Applications and Benefits

Some of the benefits of the presented framework are:

Natural and human understandable level of learning: The framework discovers relevant information from day-to-day interaction. Such interactions are natural and at human understandable level of abstraction, so as the learned concept of affordance. This facilitates for smooth verbal human-robot interaction too.

No prior database: No need to have any prior knowledge of affordance. Robot evolves its knowledge about multi-object affordance over the course of interaction.

Complementary to vision based semantic perception: The approach does not dependent heavily upon vision based information extraction about the objects, instead uses dialog based interaction, which in fact is the one of the novelties of the system. However, vision based object perception can be used as complementary capability to enrich the interpretation about the scene, hence the knowledge about affordances.

Some of the potential applications could be:

Predicting affordance based activity and activity based objects: The two aspects perceived in this paper, which are coupling of object with various affordances and coupling of object with other object for a particular affordance, together can be used to predict the most feasible activity the human might be involved in. As well as it can be used to guess the object in front of the human for a particular activity detected. However, this needs further investigation to come up with metric by combining both the types of associations in a legitimate manner. Note that this prediction can be also used to make the interaction more dynamic, during affordance learning in our framework.

Confusion based and active learning: The learned affordance can be used to detect confusion and to trigger keywords for interaction and clarifications. E.g. a false guess of affordance of mail checkability because of presence of a laptop might triggers the robot to say "If you were not ¡¡affordance based activity¿¿ (checking mails) with your ¡¡Object¿¿ (laptop) then what you were doing?" Work on active learning such as [25] can further be adapted to make the robot ask the right questions in such cases.

Temporal reasoning: Such learned affordances could also be coupled with the information of time of the day, to even learn the temporal activity pattern as well as to guess the

objects, which might be in front of the human during a particular period of time.

Task planning for human-robot interaction: Such notion of affordance learned at human level of abstraction can be used to understand human instruction and task specifications and plan for the task. Work such as [26] and [27] can adapted for such task planning requirements, which try to ground the affordances in the sensorimotor system of the robot, and fuses symbolic planning and geometric planning to come up with a feasible and executable solution.

VI. Conclusion and Future Work

In this paper we exploited the notion of multi-object affordances and proposed a novel way of affordance learning based on verbal human-robot interaction. We have shown the feasibility of the framework, which also captures the notion of experience, through an instantiation of robot in office scenario. The two types of associations are captured, *object-affordance* and *object-object coupling for a particular affordance*. As mentioned in the potential application section, one interesting future work is to come up with a metric to fuse these two estimations and develop a mechanism to predict the activity and ground the object during human-robot interaction.

Another extension is to capture the link between more than two objects for a particular affordance as well as to take into account inter-object spatial relations. Also it will be interesting to develop a reasoning mechanism, which will explicitly consider *object in hand* and *object in front* notions as two different indicators for reasoning.

As the approach does not require any prior database about objects, activities; hence the system can be easily used in different scenarios. Another interesting work is to further experiment with robot at home scenario, where it will have more variations in the affordance, objects and activities. It will be also interesting to integrate the system with short and long term memory with memorization, forgetting and abstraction mechanism to store the compact and useful information in the database to facilitate more efficient lifelong refinement.

References

[1] J. J. Gibson, "The theory of affordances," in *The Ecological Approach To Visual Perception*. Psychology Press, Sep. 1986, pp. 127–143.

[2] D. Norman, *The psychology of everyday things*. Basic Books, 1988.

[3] E. J. Gibson, "Perceptual learning in development: Some basic concepts," *Ecological Psychology*, vol. 12, no. 4, pp. 295–302, 2000.

[4] A. Clark, "An embodied cognitive science?" *Trends in cognitive sciences*, vol. 3, no. 9, pp. 345–351, 1999.

[5] F. Garbarini and M. Adenzato, "At the root of embodied cognition: Cognitive science meets neurophysiology," *Brain and cognition*, vol. 56, no. 1, pp. 100–106, 2004.

[6] E. Şahin, M. Çakmak, M. R. Doğar, E. Uğur, and G. Üçoluk, "To afford or not to afford: A new formalization of affordances toward affordance-based robot control," *Adaptive Behavior - Animals, Animats, Software Agents, Robots, Adaptive Systems*, vol. 15, no. 4, pp. 447–472, Dec. 2007.

[7] A. Stoytchev, "Behavior-grounded representation of tool affordances," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, april 2005, pp. 3060 – 3065.

[8] E. Ugur, M. R. Dogar, M. Cakmak, and E. Sahin, "Curiosity-driven learning of traversability affordance on a mobile robot," in *IEEE 6th International Conference on Development and Learning (ICDL)*, july 2007, pp. 13–18.

[9] M. Lopes, F. S. Melo, and L. Montesano, "Affordance-based imitation learning in robots," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2007, pp. 1015–1021.

[10] L. Paletta, G. Fritz, F. Kintzler, J. Irran, and G. Dorffner, "Learning to perceive affordances in a framework of developmental embodied cognition," in *Development and Learning, 2007. ICDL 2007. IEEE 6th International Conference on*, 2007, pp. 110–115.

[11] H. Koppula and A. Saxena, "Anticipating human activities using object affordances for reactive robotic response," in *RSS*, 2013.

[12] E. Erdemir, C. Frankel, K. Kawamura, S. Gordon, S. Thornton, and B. Ulutas, "Towards a cognitive robot that uses internal rehearsal to learn affordance relations," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).*, 2008, pp. 2016–2021.

[13] A. K. Pandey and R. Alami, "Affordance graph: A framework to encode perspective taking and effort based affordances for day-to-day human-robot interaction," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 2180–2187.

[14] B. Moldovan, P. Moreno, M. van Otterlo, J. Santos-Victor, and L. De Raedt, "Learning relational affordance models for robots in multi-object manipulation tasks," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 4373–4378.

[15] E. Ugur, S. Szedmak, and J. Piater, "Bootstrapping paired-object affordance learning with learned single-affordance features," in *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2014 Joint IEEE International Conferences on*, Oct 2014, pp. 476–481.

[16] H. A.Pieropan, C.H.Ek, "Recognizing object affordances in terms of spatio-temporal object-object relationships," in *International Conference on Humanoid Robots, Madrid, Spain*, 2014.

[17] H. S. Koppula, R. Gupta, and A. Saxena, "Learning human activities and object affordances from rgb-d videos," *Int. J. Rob. Res.*, vol. 32, no. 8, pp. 951–970, Jul. 2013.

[18] R. Moratz and T. Tenbrink, "Affordance-based human-robot interaction," in *Towards Affordance-Based Robot Control*, ser. Lecture Notes in Computer Science, E. Rome, J. Hertzberg, and G. Dorffner, Eds. Springer Berlin Heidelberg, 2008, vol. 4760, pp. 63–76.

[19] A. M. Borghi, "Anna m. borghi. action language comprehension, affordances and goals. in yan coello, angela bartolo (eds).(in press early 2012). language and action in cognitive neuroscience. contemporary topics in cognitive neuroscience series. psychology press."

[20] J. Furnkranz and P. Flach, "An analysis of rule evaluation metrics," in *Proc. 20th International Conference on Machine Learning (ICML)*. AAAI Press, January 2003, pp. 202–209.

[21] A. Agostini, C. Torras, and F. Wörgötter, "Integrating task planning and interactive learning for robots to work in human environments," in *Proceedings of the 22nd International Conference on Artificial Intelligence (IJCAI)*, 2011, pp. 2386–2391.

[22] A. Pandey and R. Alami, "Towards human-level semantics understanding of human-centered object manipulation tasks for hri: Reasoning about effect, ability, effort and perspective taking," *International Journal of Social Robotics*, vol. 6, no. 4, pp. 593–620, 2014.

[23] B. Cestnik, "Estimating probabilities: A crucial task in machine learning," in *Proceedings of the Ninth European Conference on Artificial Intelligence, ECAI*, 1990, pp. 147–149.

[24] E. Pot, J. Monceaux, R. Gelin, and B. Maisonnier, "Choregraphe: a graphical tool for humanoid robot programming," in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, Sept 2009, pp. 46–51.

[25] M. Cakmak and A. L. Thomaz, "Designing robot learners that ask good questions," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '12. New York, NY, USA: ACM, 2012, pp. 17–24.

[26] S. Kalkan, N. Dag, O. Yürüten, A. M. Borghi, and E. Şahin, "Verb concepts from affordances," *Interaction Studies*, vol. 15, no. 1, pp. 1–37, 2014.

[27] L. de Silva, A. K. Pandey, and R. Alami, "An interface for interleaved symbolic-geometric planning and backtracking," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 232–239.