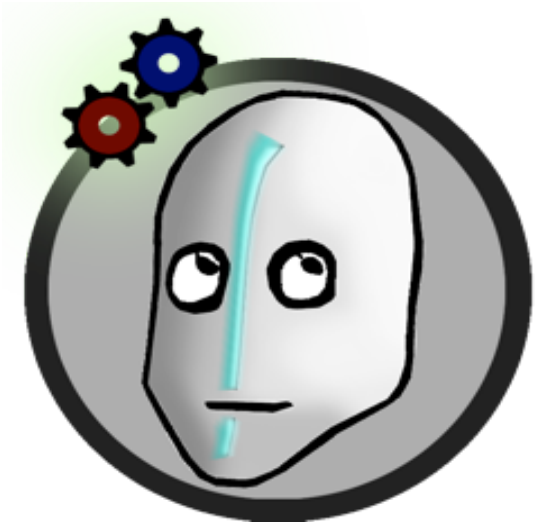


ACRONYME	**ROBOERGOSUM**
NOM DU PROJET	ROBOT CONSCIENTS
REFERENCE	DECISION ANR-12-CORD-0030
NUMERO DE LA TACHE	**T1**
NOM DE LA TACHE	Situation Awareness and Semantic Scene Interpretation
NUMERO DU RAPPORT	**D1.3**
TITRE DU RAPPORT	Inclusion of the Scene Interpretation Processes in the Global Architecture
PARTENAIRES	ISIR, LAAS
DATE	T0+36

Contents

1 Summary

Scene understanding requires the coordination of many simultaneous tasks such as data retrieval, surface orientation estimation, object recognition and tracking, and scene categorization [1]. Scene understanding is the process of perceiving, analysing and elaborating an interpretation of a dynamic scene observed through a network of sensors (such as 3D sensors). Its goal consists mainly in matching signal information coming from sensors observing the scene with models which humans are using to understand the scene. One can say scene understanding is both adding and extracting semantic from the sensor data characterizing a scene.

A cognitive robot may face several types of failures during the execution of its actions in the real world. These failures may arise due to the gap between the real-world facts and their symbolic representations used during planning, unexpected events that may change the current state of the world or internal problems [1][2][3]. In the physical world, there is uncertainty in the data gathered through sensors due to different factors like varying illumination conditions or dynamic environments. This makes bare object recognition results unreliable for cognitive robotic applications. The robot should gain experience from these failures and use this experience in its future tasks, which requires tight integration of continual planning, monitoring, reasoning and lifelong experimental learning [4][5]. Efficient and consistent scene interpretation is a prerequisite for these cognitive abilities.

In these papers, we study, develop, and experimentally evaluate sensorimotor representations and scene interpretation processes based on visual and proprioceptive inputs when the robot physically interacts with objects. This enables robots to understand their environment by interacting with it. Our architecture builds models of objects based on perceptual clues and effects of robot actions on them, which relate to the notion of *affordance*. We employ a Bayesian network that represents with continuous and discrete variables the objects, actions, and effects in the observable environment. We then perform structure learning to identify the most probable Bayesian network that best fits the observed data. The discovered structure of the Bayesian network allows the robot to discover causal relationships in the environment using statistical data.

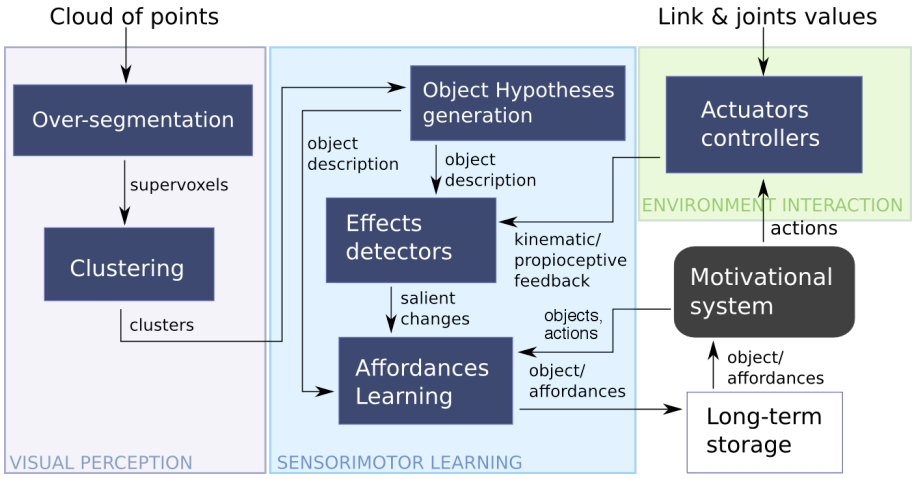


Figure 1: Architecture of the proposed sensorimotor approach for scene affordance learning. Affordances are a key concept in our scene interpretation solution.

Fig. 1 shows the proposed architecture for learning affordances. Measurements from Visual perception and Environment interaction are considered as the main inputs of our approach. Visual perception extracts, from clouds of points, a set of clusters. Clusters are then tracked to generate object hypotheses to interact with. A *motivational system* is in charge of selecting objects and actions that will be applied on them. Proprioceptive feedback is retrieved in the form of joint and force measurements. *Effect detectors* analyse the input from perception and action tasks to extract salient changes from the interaction process. Sensorimotor learning is the intersection between the two input processes and represents the fusion between the perception and action components. *Affordances learning* finds the correlations that build the final sensorimotor representation by relating *objects*, *actions* and induced changes considered as *effects*. A *long-term storage* is used to save the final representation and to provide a feedback for the motivational system. Our approach does not rely on *a priori* dependency assumptions between them. It allows the robot to infer the dependencies between the elements while interacting and combining percetual and proprioceptual data. The learned sensorimotor representation along the Bayesian framework will allow the robot's motivational system to make predictions about elements in the environment. Moreover, this inferred information can be used for future planning tasks or to add sensor and motor capabilities to the innate repertoire.

These articles have been presented in the The 2016 International Symposium on Experimental Robotics (ISER 2016) and in the Workshop on Machine Learning Methods for High-Level Cognitive Capabilities in Robotics 2016 and are going to be published in the corresponding proceedings.

2 Publications

Discovering and Manipulating Affordances

R. Omar Chavez-Garcia, Mihai Andries, Pierre Luce-Vayrac, and Raja Chatila

Institut des Systèmes Intelligents et de Robotique (ISIR),
Sorbonne Universités, UPMC Univ Paris 06, CNRS, 75005 Paris, France.
{chavez,andries,luce-vayrac,raja.chatila}@isir.upmc.fr

Abstract. Reasoning jointly on perception and action requires to interpret the scene in terms of the agent's own potential capabilities. We propose a Bayesian architecture for learning sensorimotor representations from the interaction between *perception*, *action*, and *salient changes* generated by robot actions. This connects these three elements in a common representation: *affordances*. In this paper, we are working towards a richer representation and formalization of affordances. Current experimental analysis shows the qualitative and quantitative aspects of affordances. In addition, our formalization motivates several experiments for exploring hypothetical operations between learned affordances. In particular, we infer affordances of composite objects, based on prior knowledge on the affordances of the elementary objects.

Keywords: affordances, sensorimotor representations, developmental robotics

1 Motivation, Problem Statement

The grounding of robotic knowledge [19] is the problem of creating links between the entities and events in the observable environment and their symbolic representations employed by a robot's reasoning algorithms. Solving this problem would allow robots to autonomously discover their environment, without the need of human intervention. Symbolic grounding cannot be achieved by a process of observation alone, and requires interaction between the agent and its environment.

In this paper, we study, develop, and experimentally evaluate sensorimotor representations and scene interpretation processes based on visual and proprioceptive inputs when the robot physically interacts with objects. This enables robots to understand their environment by interacting with it. Our architecture builds models of objects based on perceptual clues and effects of robot actions on them, which relate to the notion of *affordance*. We employ a Bayesian network that represents with continuous and discrete variables the objects, actions, and effects in the observable environment. We then perform structure learning to identify the most probable Bayesian network that best fits with the observed data. The discovered structure of the Bayesian network allows the robot to discover causal relationships in the environment using statistical data.

The remainder of the paper is structured as follows. In Section 2 we discuss related work on the discovery of object affordances, and we introduce our specific

contribution. Section 3 describes our technical approach, including an illustration of the architecture employed for learning affordances. Experimental results are presented in Section 4. We draw a conclusion Section 5 and present ideas for future work.

2 Related work on object affordance discovery

From the seminal work of Gibson, *the affordance of anything is a specific combination of the properties of its substance and its surfaces taken with reference to an animal* [1]. Sahin et al. discuss on the former definition for the domain of autonomous robot control. They introduce the acquired aspect of an affordance, such that when an agent applies a behavior on an entity, an effect is generated [2].

At the same time, several efforts have spawned from the domain of developmental robotics, for exploring and learning robots' environment. The approach was based on a cycle of exploration-manipulation, initialized with a collection of minimal knowledge and innate capabilities. These works studied the discovery of meaningful discrete motion primitives [10] or sequences thereof [8], using stochastic and deterministic [16] approaches. This allowed a robot to learn object affordances and the predictors that anticipate the effects that these action primitives create.

Stoytchev [4] suggested that the autonomous learning of affordances by a robot provides representations of the observed objects, actions, and effects that are grounded in the environment. This hinted that the affordances can be used to create grounded symbolic representations for the observed entities and events, both in the physical and abstract world.

Montesano et. al. [3] modeled affordances with Bayesian networks. The unsupervised learning of affordances was formulated as a structure learning algorithm, where affordances were encoded in the probabilistic relations between actions, object features and effects.

Hermans et al. [6] suggested to learn affordances in 2 steps: first generating object attributes from the observed visual features, and then linking these object attributes to affordances. In their case, they employed 2D visual features of objects (shape, color, material) and weight features to learn and predict affordances such as pushable, rollable, graspable, liftable, dragable, carryable, and traversable.

Similarly, Jain et al. attempted to estimate the affordances of previously unknown tools, based on the assumption that functional features remain distinctive and invariant across different tools used to perform similar tasks [7]. The proposed system learns bi-directional causal relationships between actions, functional features and the effects of tools, and uses a Bayesian network to model the probabilistic dependencies in the observation data.

Zhu et al. [9] inferred the affordances of objects (with a particular interest for tools) based on their resemblance to other objects observed while in use by a human during a learning phase, using RGB-D data. They made the hypothesis that the object use demonstrated by the human was optimal.

The main novelty of this paper consists in predicting the affordances of combinations of objects, based on prior knowledge on the affordances of the constituent parts of the composite object. We employ a probabilistic architecture, that generates a sensorimotor representation which encodes, through the learning of affordances, effects, objects and actions using the same formalism. The architecture spans from low-level data acquired from sensors and actuators, up to learning relations between higher-level representations. We use 3D visual features, as well as force measurements to create a description of the objects and effects generated through interactions with objects. Although we assume that the agent has a limited innate set of sensors and motor capabilities, the architecture allows for learning and extending these capabilities as well. We employed a continuous Bayesian network (as opposed to discrete Bayesian networks) to work with the quantitative aspect of affordances (i.e. to measure, learn and predict intensities of effects).

3 Technical Approach

Fig. 1 shows the proposed architecture for learning affordances. Measurements from VISUAL PERCEPTION and ENVIRONMENT INTERACTION are considered as the main inputs of our approach. VISUAL PERCEPTION extracts, from clouds of points, a set of clusters. Clusters are then tracked to generate object hypotheses to interact with. A *motivational system* is in charge of selecting objects and actions that will be applied on them. Proprioceptive feedback is retrieved in the form of joint and force measurements. *Effect detectors* analyze the input from perception and action tasks to extract salient changes. SENSORIMOTOR LEARNING is the intersection between the two input processes and represents the fusion between the perception and action components. *Affordances learning* finds the correlations that build the final sensorimotor representation by relating *objects*, *actions* and induced salient changes considered as *effects*. A *long-term storage* is used to save the final representation and to provide a feedback for the motivational system.

3.1 Visual perception

In order to interact with the environment, a segmentation process is performed to identify the objects in the scene. Voxel Cloud Connectivity Segmentation (VCCS), presented in [11], benefits from 3D geometry provided by RGB+D cameras to generate an even distribution of *supervoxels* in the observed space. The seeding methodology to find the supervoxels is based on 3D space and a flow-constrained local iterative clustering which uses color and geometric features. As this algorithm relies on strict partial connectivity between voxels, it guarantees that supervoxels cannot flow across boundaries which are disjoint in 3D space.

Once we obtained the oversegmentation from supervoxels extraction, we implemented the non-parametric clustering detailed in [12] to find the shape of the object hypotheses based on the set of supervoxels.

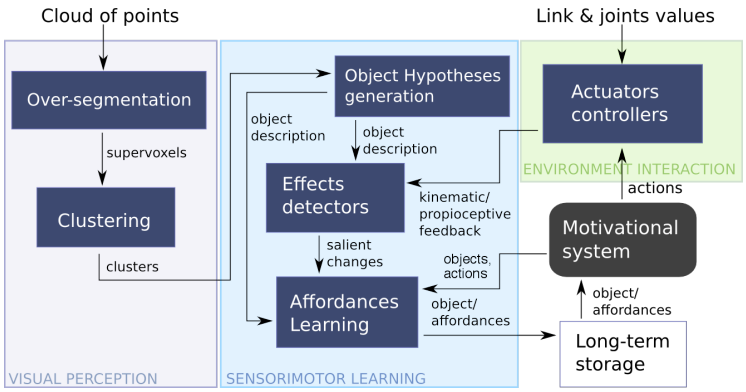


Fig. 1. Architecture of the proposed sensorimotor approach for affordance learning.

3.2 Affordances for sensorimotor representation

We consider an agent (robot) endowed with a set of innate actions A, and a set of innate feature extractors $\mathcal{P}$, that can be augmented through learning. In addition, O is the set of all the objects in the environment, and E is the set of all the possible observable effects. When the agent applies action $a \in A$ to an entity (object) $o \in O$ in the environment, a salient change (effect) $e \in E$ is generated, we call this *acquired* relation an *affordance* [2].

From the agent's perspective, a resulting affordance is defined as follows:

$$\alpha^{\text{agent}} = (e, (o, a)), \text{ for } e \in E, \ o \in O, \text{ and } a \in A \tag{1}$$

more generally, this agent will gradually build a set of affordances *Aff* composed of the affordances α_i:

$$\alpha_i^{\text{agent}} = (e_j, (o_k, a_l)), \text{ for } e_j \in E, \ o_k \in O, \text{ and } a_l \in A \tag{2}$$

An object o_k is defined as the set of values for the n innate properties extractors $\rho \in \mathcal{P}$:

$$o_k = \{\rho_1(cluster), \rho_2(cluster), ...\rho_n(cluster)\}, \tag{3}$$

where *cluster* represents the object hypothesis obtained by the visual perception module.

Actions are a set of motor capabilities $A = \{a_1, ..., a_m\}$, defined as:

$$a_k(V^*, \gamma, \sigma_{a_k}), \tag{4}$$

being V^* the desired value for the robot control variables V, γ its proprioceptive feedback and σ_{a_k} the particular action parameters.

Effects are a set of salient changes in the world ω detected by robot's innate detectors e:

$$E = \{e_1(\omega), e_2(\omega), ..., e_q(\omega)\} \tag{5}$$

which means that effects can be related to objects and agents, allowing to detect exteroceptive and proprioceptive changes.

3.3 Affordance learning

Considering the statistical nature of acquiring affordances through environment exploration, elements E, O and A in (1) can be represented as random variables in a Bayesian Network (BN) $\mathcal{G}$. Through the cycle of perception-interaction we obtain instances of these variables generating a data set $\mathcal{D}$. The problem of discovering the relations between E, O and A can be seen as finding dependencies between the variables in $\mathcal{G}$, i.e., learning the structure of the corresponding BN from data $\mathcal{D}$. Using the BN framework we are capable of displaying relationships between variables. The directed nature of its structure allows us to represent cause-effects relationships and to combine the action and perception components in a stochastic sensorimotor representation.

We implement a score-based maximization approach for finding the BN structure from $\mathcal{D}$ [13]. The score of a BN structure $\mathcal{G}$ is defined as the posterior probability given the data $\mathcal{D}$, i.e. $\mathcal{S}(\mathcal{G}, \mathcal{D}) = P(\mathcal{G}|\mathcal{D})$, we define $\mathcal{S}$ as the compression rate of the data $\mathcal{D}$ with an optimal code induced by the BN. As the number of independent and identical distributed random variables tends to infinity, no compression of data is possible for a rate less than the Shannon entropy [18].

Information scoring functions for structure learning are based on compression [18]. The score of a BN $\mathcal{G}$ is related to the compression that can be achieved over the data $\mathcal{D}$ with an optimal code induced by $\mathcal{G}$. The quality of a BN can be computed by:

$$\mathcal{S}(\mathcal{G}|\mathcal{D}) = \mathcal{S}_{log-l}(\mathcal{G}|\mathcal{D}) - f(N)|\mathcal{G}| \tag{6}$$

where the log-likelihood score $\mathcal{S}_{log-l}$ tendd to favor complete network structures, without providing reliable independence assumptions for the learned network [14]. $|\mathcal{G}|$ denotes the network complexity, i.e. the number of parameters in the network. $f(N)$ is a non-negative penalization function. If $f(N) = 1$, (6) becomes the AIC (Akaike Information Criterion) score; if $f(N) = \frac{log(N)}{2}$, (6) represents the BIC (Bayesian Information Criterion) score [14].

Bayesian inference in our discrete BN provides the probability that an affordance α_i is present. However, it does not provide a mechanism to quantify the affordance w.r.t. the specific environment situations that triggered it.

We believe that by preserving the continuous aspect of the elements in the affordance (1), we also maintain the necessary information for an affordance quantifying approach, i.e., Bayesian inference over a Gaussian BN (GBN). Relations in (2) can be represented as a multivariate normal distribution of continuous random variables, i.e., the affordance's elements. Continuous variables are modeled as linear regressions in a Gaussian BN, where the relevant parameters of each local distribution are the regression coefficients (for each variable *parent*) and the standard deviation of the residuals.

Structure learning is performed by identifying vanishing regression coefficients using two assumptions, event equivalence and parameter modularity, that allow the construction of prior distributions for our multivariate normal parameters [15]. As in learning of discrete BN strucures, we implemented a score function to evaluate the quality of the continous BN. To do so, we used a score equivalent

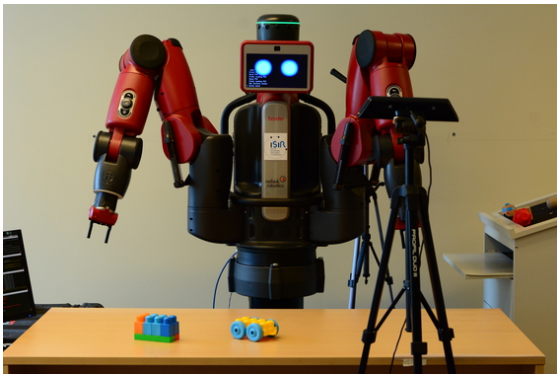


Fig. 2. The Baxter robotics platform used for our experiments. The RGB-D camera used for perception is visible in the foreground.

to the Gaussian posterior density, which follows a Wishart distribution and is at the core of the belief networks framework [17].

3.4 Sensorimotor learning results

Our Baxter experimental platform (Fig. 2) is equipped with 2 arms with 7 degrees of freedom. One electrical gripper is attached to each arm. For visual perception, we use a Microsoft Kinect sensor that captures RGB-D data. For environment interaction we use the left arm and its respective gripper.

In order to evaluate the generalization capabilities of the bayesian model learned with our architecture, we implemented a discrete structure learning for an experiment composed of: several objects represented by dominant color (ρ_{color}), size (ρ_{size}) and shape (ρ_{shape}); four innate actions: *poke* (a_{poke}), *push* (a_{poke}), *open gripper* (a_{open_g}) and *close gripper* (a_{close_g}); and three types of effect detectors: feedback force in the end effector (e_f), distance between the gripper fingers (e_{gr_d}), and one movement detector for each detected object ($e_{mv_{o_i}}$).

We developed a hill-climbing based algorithm to search the optimal structure [13]. Two score functions were implemented: BIC and AIC as described in section 3.3. In both cases, the learned BN structure is increasing the quality of representation for the data when more interactions are made by the agent. In addition, the resulting BN can better generalize the learned knowledge for future interactions. Although the *log-likelihood* loss for both information-based scores seems to be similar, the learned network from AIC score is less complex than the one from BIC score, which influences the performance of BN inference afterwards.

Applying inference over the learned BN, the robot can estimate probability distributions for effect prediction $P(E|O, A)$, feedback in action selection $P(A|O, E)$ or object recognition given its behavioral description $P(O|A, E)$.

Using our proposed architecture (Fig.1), we defined our experiment on the continuous framework as follows. The relevant object properties are the dominant

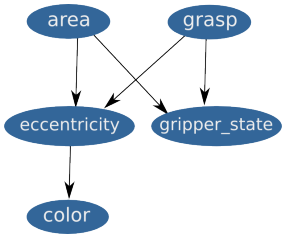


Fig. 3. Learned Gaussian BN. All nodes are defined as continuous random variables.

color, visible area as *size* and object elliptical eccentricity as *shape*. Action *grasp* is defined as $grasp(pos)$, where *pos* is a parameter describing the position (w.r.t. to object's longitudinal component) where *grasp* is performed. The relevant effect *gripper_state* varies over the distance between gripper's fingers.

Fig. 3 shows the Gaussian BN learned by our architecture. Interactions were done on two different objects: a blue bar and a red bat of baseball. The learned GBN models *grasp-ability* as present in both objects, but only the blue bar is graspable for every configuration of the action. *Grasp-ability* on the red bat disappears after passing the half of its length. A video demonstration can be found at `https://cloud.isir.upmc.fr/owncloud/index.php/s/9EKUq05D58Wiyfo`.

4 Experiments with affordances of composite objects

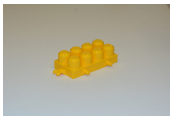
The goal of our experiments is to identify a formalism that could infer the affordances of composite objects, based on prior knowledge about the affordances of the primary objects that constitute them. We state that Bayesian networks, through structure learning, can not only discover affordances, but also capture their quantitative aspect, by employing continuous variables in the representation of actions and effects, that are represented as continuous variables. The experiments will help us demonstrate this.

Our experimental procedure is composed of four steps: (1) performing a certain action with a set of objects (separate and composite) and observing the effects, (2) defining the random variables corresponding to the observed objects, actions, and effects inside the Bayesian network, (3) feeding the interaction data to the structure learning algorithm of the Bayesian network, and (4) interpreting the structure of the Bayesian network that best fits the recorded data according to calculations.

First, we consider the discovery of affordances. From our experiments, we can interpret the model learned by the discrete Bayesian Network as a qualitative aspect of an affordance, regarding the presence or absence of a relation between the elements of an affordance (e.g. an object is *push-able*, i.e. it goes a certain distance from its original location).

We further attempt to attach a quantitative dimension to the learned affordance by representing its elements as continuous random variables. This allows

Table 1. The objects used in the experiments, and the employed composition order.

	Experiment 1	**Experiment 2**	**Experiment 3**
top object			
bottom object			
composite object			
observed effect for the bottom object	affordance acquisition	affordance maintenance	affordance loss

not only to predict that the affordance is present, but also to infer the parameter values of its elements that influence this affordance (e.g. the effect of the *push* action on the object is a function of the action's input parameters).

Three experiments are analysed in this section, all related to the inference of affordances of composite objects: (1) affordance acquisition, (2) affordance maintenance, and (3) affordance loss. Since our experiments focused on the composition of objects, we performed them on objects specifically designed for that: toys that can assemble and disassemble. These experiments are detailed in the following sections, and are illustrated in Table 1, which shows the objects employed, as well as their composition method.

4.1 Affordance acquisition

Following the experiment description from Table 1, column Experiment 1, each object is described by two elements: one describing the number of atomic perceptual properties that forms it, and the other the position of the atomic property inside it (top or bottom). These properties allow to represent atomic objects (with only one property) and possible composite objects. In this scenario, we have two atomic perceptual properties *wheel* and $cartFrame$ and together they can combine to form a *cart*.

The robot performed random interactions with the action a_{poke} and the atomic objects *wheel*, $cartFrame$ and with the composite object *cart* (50 interactions with each object). The effect detector developed was based on the *distance* that an object moves after the action is executed.

We use Gaussian random variables to represent the perceptual properties and the distance effect. We use nominal variables to represent the action undertaken (poke, no action), and the objects employed. The object composition was represented using 2 variables: $objectBottom$ and $objectTop$, representing respectively the atomic object at the bottom of the composite object, and the one on top.

Figure 5 shows the resulting network after the learning process. We can notice that the parameters influencing the distance, over which an object travels after an interaction, are correctly inferred. The action *poke* influences this distance, while the action *noAction* does not. The object at the bottom also influences this distance: *wheels* roll further than the *cartFrame* after poking. The object at the top is also linked to the distance variable, since the distance travelled by the *cart* (i.e. *wheels* with *cartFrame* on top) differs from the one travelled by the individual wheels.

Let us use the relationships learned in this example, to infer the affordances of a similar composite object.

4.2 Affordance maintenance and loss

The second and third experiments consist in learning the correct structure of the Bayesian network, so as to correctly predict the maintenance or loss of affordances of atomic objects that form the composite object. In this example, we will consider two new objects: the *cart*, and the *blockLoad* that we can put on or under the cart (see Experiments 2 and 3 in Table 1). We feed the BN the data obtained in the interactions with these new atomic objects (50 interactions with each object), but not for their composition, and obtain the BN seen in Figure 4.

We stated the acquired nature of an affordance in eq. 2, and for this reason, the inferred affordance will be considered an estimation until the robot, by interaction, validates it. If we represent the composite object $obj_{composite} = \{objectBottom = cart,\ objectTop = blockload\}$, we can obtain an estimation of its affordance by calculating $P(distance|objectBottom = cart, objectTop = blockload)$ from the learned BN. In our experiment, the probability distribution for this calculation is similar to $P(distance|objectBottom = cart)$ showing experimentally that the estimated affordance *movable* of the composite object is similar to the affordance of one of its elements.

In the BN represented in Figure 4, if the value of the *objectBottom* variable is known, the variables *distance* and the *objectTop* are conditionally indepen-

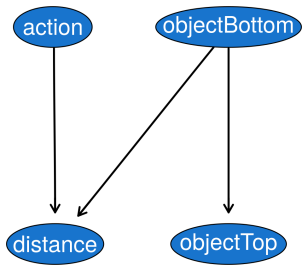


Fig. 4. The Bayesian Network obtained after feeding the interaction data with the atomic objects *cart* and *blockLoad*.

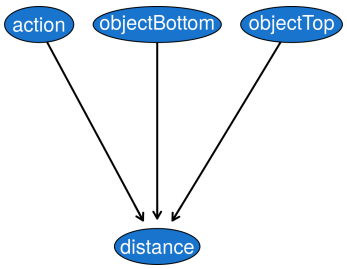


Fig. 5. Conditional linear Gaussian network obtained after learning process.

dent. This means that the upper part of a composite object does not influence the distance that this composite object traverses after a poke action. This can be interpreted as an *affordance loss*. On the other hand, the bottom part of a composite object (i.e. *objectBottom*) does impact the distance it traverses after a poke, suggesting that its *affordance is maintained*. This is confirmed experimentally: after a poke, the atomic objects *cart* and *blockLoad* travel an average distance of 45 centimeters and 9 centimeters, respectively. The composite object with the *cart* at the bottom travels an average distance of 28.4 centimeters, while the one with the *blockLoad* at the bottom travels an average distance of only 3.8 centimeters.

4.3 Discussion on the results

The goal of our ongoing experiments is to infer the relations that exist between affordances, which would allow to refine the definition and formalization of an affordance.

By decomposing an object offering a specific affordance into its constituent parts, we may wonder what are the affordances of the obtained parts. Answering this question requires us to introduce a mathematical operator, which would be able to estimate the affordances of an object obtained through the decomposition of an object, or through the composition of objects. It is yet unclear if this mathematical operator would apply to the objects and their properties (identifying their affordances as a consequence), or if it would apply to the entire affordance relation $(E, (O, A))$. This opens a whole new domain of inquiry about the relations between affordances.

We designed three experiments in order to test our hypothesis. Following (2), we can define a particular affordance for an object o_i as $\alpha_i = (e_k, (o_i, a_l))$. Then, we can *decompose* o_i into two *new* objects o'_i and o''_i, by dissecting its property set in 2 complementary parts $(\varrho_1, \varrho_2 \subset o_i,\ \varrho_1 \cup \varrho_2 = o_i,\ \varrho_1 \cap \varrho_2 = \emptyset)$, one for each object, and padding them with null values:

$$\begin{aligned} o'_i &= \{\rho_x | \rho_x \in \varrho_1\} \cup \{\rho_y = null | \rho_y \in \varrho_2\}, \\ o''_i &= \{\rho_x | \rho_x \in \varrho_2\} \cup \{\rho_y = null | \rho_y \in \varrho_1\}. \end{aligned} \tag{7}$$

Using the learned model from our proposed architecture, we can infer:

$$\alpha'_i = (e_k, (o'_i, a_l)),\ \alpha''_i = (e_k, (o''_i, a_l)). \tag{8}$$

If the removal of a property does not influence the affordance of an object $(\alpha'_i \equiv \alpha_i)$, then this property can be considered as non salient for this particular affordance. In addition, if we can rewrite α_i as:

$$\alpha_i = (e_1, (o'_1 \otimes o''_1, a_1)), \tag{9}$$

the computation defined by the operator $\otimes$ suggests the existence of a combination of affordances. Experiments can help to discover the properties of this *composition* operator.

Let us represent the set of salient features from objects o_x and o_y for one of their affordances as $salient_{\alpha_i}(o_x)$ and $salient_{\alpha_j}(o_y)$ respectively, where

$$\alpha_i = (e_{kx}, (o_x, a_{lx})),\ \alpha_j = (e_{ky}, (o_y, a_{ly})). \tag{10}$$

If o_x and o_y do not share salient features, $salient_{\alpha_i}(o_x) \cap salient_{\alpha_j}(o_y) = \emptyset$, and $|salient_{\alpha_i}(o_x)| + |salient_{\alpha_j}(o_y)| = n$, we can construct a new object o_{xy} by *selectively* combining the salient properties of o_x and o_y,

$$o_{xy} = salient_{\alpha_i}(o_x) \cup salient_{\alpha_j}(o_y), \tag{11}$$

which by definition should retain affordances α_i and α_j. We can empirically discover the properties of affordances of this new object o_{xy} w.r.t. the properties of o_x and o_y.

Through these experiments (decomposition, composition and selective composition) we will be able to estimate the affordances of combined or de-composed objects, and verify this estimation empirically, shedding light on the nature of these *affordance operators*.

5 Conclusion and future work

We introduced a Bayesian architecture for learning sensorimotor representations from the interaction between objects, robot actions, and the generated effects. In particular, it employs Gaussian random variables that capture the quantitative aspect of actions and effects.

We introduce the concept of *primary objects* to capture prior knowledge on their affordances. We also introduce the concept of *composite objects*, for which we want to identify a relationship between the objects they are composed of, and the way they are assembled, in order automatically infer their affordances. We performed experiments to infer the affordances of composite objects, based on prior knowledge about the affordances of the primary objects that constitute them. Results form the learned Bayesian network showed information regarding the acquisition, maintenance, and loss of affordances by the employed primary objects, depending on their position in the composite object. The obtained results suggest that it may be possible to define an operator acting on the elements of affordances, which could predict the affordances of new objects, obtained through the combination of known objects.

In our future work, we plan to identify the salient features of objects that endow these objects with specific affordances. These salient features can be identified while performing object composition. In this case, a gained affordance would be related to features acquired after object composition. Salient features can also be identified during object decomposition. In this case, a lost affordance would be related to features lost after object decomposition into constituent parts.

Although our approach is a statistically based learning technique, it would be interesting to analyse other approaches that could provide statistically similar results or improvement with fewer interactions. It would be interesting to employ algorithms that can identify causal relationships between actions, object features and effects with as few observations as possible (one or two).

Acknowledgments. This work has been funded by a DGA (French National Defense Agency) scholarship (ER), and by French Agence Nationale de la Recherche ROBOERGOSUM project under reference ANR-12-CORD-0030.

References

1. Gibson, J.: The theory of affordances. Perceiving, acting, and knowing: Toward an ecological psychology, pp. 67–82 (1977)
2. Sahin, E., Cakmak, M., Dogar, M., et al.: To Afford or Not to Afford: A New Formalization of Affordances Toward Affordance-Based Robot Control. Adaptive Behavior, vol. 15, pp. 447-472 (2007)
3. Montesano, Luis, et al. "Learning object affordances: From sensory–motor coordination to imitation." IEEE Transactions on Robotics 24.1 (2008): 15-26.
4. Stoytchev, A. (2008). Learning the affordances of tools using a behavior-grounded approach. Lecture Notes in Computer Science, 4760 LNAI, 140158.
5. Ugur, E., Sahin, E. Traversability: A Case Study for Learning and Perceiving Affordances in Robots. Adaptive Behavior, 18(3-4), pp.258–284 (2010)
6. Hermans, T., Rehg, J.M., Bobick, A.: "Affordance prediction via learned object attributes." IEEE ICRA: Workshop on Semantic Perception, Mapping, and Exploration. 2011.
7. Jain, R., Inamura, T.: Bayesian learning of tool affordances based on generalization of functional feature to estimate effects of unseen tools. Artificial Life and Robotics, vol. 18, pp. 95–103, Springer (2013)
8. Moldovan, B., Moreno, P., van Otterlo, M.: On the use of probabilistic relational affordance models for sequential manipulation tasks in robotics. IEEE ICRA, pp. 1290–1295 (2013)
9. Zhu, Y., Yibiao Z., Song C. Z.: Understanding tools: Task-oriented object modeling, learning and recognition. Proc. IEEE CVPR, (2015).
10. Ugur, E., Nagai, Y., Sahin, E., Oztop, E.: Staged Development of Robot Skills: Behavior Formation, Affordance Learning and Imitation with Motionese. IEEE Transactions on Autonomous Mental Development, vol. 7, pp. 119–139 (2015)
11. Papon, J., Abramov, A., Schoeler, M., et al.: Voxel cloud connectivity segmentation - Supervoxels for point clouds. Proc. IEEE CVPR, pp. 2027–2034 (2013)
12. Comaniciu, D., Meer, P., Member, S.: Mean Shift: A Robust Approach Toward Feature Space Analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 24(5), 603619 (2002).
13. Tsamardinos, I., Brown, L. E., Aliferis, C. F.: The max-min hill-climbing Bayesian network structure learning algorithm. Machine Learning, 65(1), pp. 31–78 (2006)
14. Koski, T. J. T., Noble, J. M.: A Review of Bayesian Networks and Structure Learning. Mathematica Applicanda, 40(1), pp. 53–103 (2012)
15. Geiger, D., Heckerman, D.: Learning Gaussian Networks. In Proc. of the Tenth Conf. on Uncertainty in Artificial Intelligence, pp. 235243 (1994).
16. Dehban A., Jamone L., Kampff A. R., Santos-Victor J.: Denoising auto-encoders for learning of objects and tools affordances in continuous space. IEEE ICRA (2016)
17. Geiger, D., Heckerman, D. Learning Gaussian Networks. In Proc. of the Tenth Conf. on Uncertainty in Artificial Intelligence. pp. 235–243 (1994)
18. Chiu, E., Lin, J., Mcferron, B., Petigara, N., Seshasai, S. : Mathematical Theory of Claude Shannon. (2001)
19. Harnad, Stevan. "The symbol grounding problem." Physica D: Nonlinear Phenomena 42.1-3 (1990): 335-346.

From Perception and Manipulation to Affordance Formalization

R. Omar Chavez-Garcia, Mihai Andries, Pierre Luce-Vayrac and Raja Chatila

***Abstract*— Reasoning jointly on perception and action processes requires to interpret the scene in terms of the agent's own potential capabilities. We propose a Bayesian approach and an architecture for learning sensorimotor representations from the interaction between *perception*, *action*, and the *salient changes* generated by *action*. These representations connect the three elements in a common formalism: *affordances*. We formalize both qualitative and quantitative aspects of affordances based on empirical evidence. In addition, our formalization motivates several experiments for exploring hypothetical combinations between learned affordances.**

I. Introduction

Scene *understanding* has commonly been addressed as a process of observation. Research in developmental robotics and cognitive science have however shown that there is a strong relationship between perception and action.

We study and develop sensorimotor representations and scene interpretation processes based on visual and proprioceptive inputs when the robot physically interacts with objects. These processes build models of objects based on perceptual clues and effects of robot actions on them, which relate to the notion of affordances [1].

Several efforts have followed developmental ideas for exploring and learning robots' environment. Their background approach is performed via a cycle of exploration-manipulation which is initialized with a collection of minimal knowledge and innate capabilities. They study how to discover meaningful discrete motion primitives [2] or sequences thereof [3] with stochastic and deterministic [4] approaches, which allow the robot to learn object affordances and predictors that anticipate the effects of these primitives.

The approach we propose is depicted in Fig. 1. Data from sensors and actuators generate a sensorimotor representation that encodes, through the learning of affordances, effects, objects and actions in the same formalism [5]. Although we assume that the agent has a minimal innate set of sensors and motor capabilities, the architecture allows for learning these capabilities as well.

The existing literature lacks a formalization of the constituent elements of an affordance [6], [7]. This may explain the absence of experiments regarding relationships between learned affordances. We design and perform experiments to discover these relations, aiming at expanding the definition and formalization of affordances.

R. Omar Chavez-Garcia, Mihai Andries, Pierre Luce-Vayrac and Raja Chatila are with the Institut des Systèmes Intelligents et de Robotique (ISIR), Sorbonne Universités, UPMC Univ Paris 06, CNRS, 75005 Paris, France. {chavez, andries, luce-vayrac, raja.chatila}@isir.upmc.fr

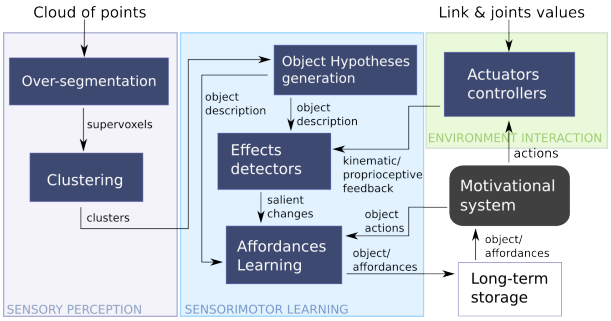


Fig. 1. Architecture of the proposed sensorimotor approach for affordance learning. A set of object hypotheses and action commands are combined through affordance learning into a sensorimotor representation.

II. Proposed Approach for Affordance Learning

When an agent g applies an action $a \in A$ over an object $o \in O$ in the environment, a salient change (effect) $e \in E$ is generated. From the agent's perspective, the i_{th} affordance is defined as follows:

$$\alpha_i^{agent} = ((o_k, a_l), e_j), \text{ for } o_k \in O, a_l \in A \text{ and } e_j \in E \quad (1)$$

where the set of actions A is innate to the agent, as are its capabilities to represent objects O and detect salient changes called effects E. The agent perception module segments 3D input images into clusters. We assume that the agent has a set of n innate feature detectors ρ (e.g., shape, size, color, ...). We define an object o_k as the set of feature values provided by those detectors:

$$o_k = \{\rho_1(c_k), \rho_2(c_k), ...\rho_n(c_k)\}, \quad (2)$$

where c_k represents a given cluster (object hypothesis). Actions are a set of motor capabilities $A = \{a_1, ..., a_m\}$. An action a_l is defined with respect to their control variables in joint space: $a_l : \{Q, \dot{Q}, \ddot{Q}\}_\tau$ where Q are the joint parameters of the robot used in action a_l, and τ the duration of this action. Effects are a set of q salient changes in the world detected by robot's innate detectors ξ, $E = \{\xi_1, \xi_2, ..., \xi_q\}$.

Considering the statistical nature of acquiring affordances through environment exploration, elements of E, O and A in (1) can be represented as random variables in a Bayesian Network (BN) $\mathcal{G}$. Through the cycle of perception-interaction we obtain instances of these variables generating a data set $\mathcal{D}$. The problem of discovering the relations between E, O and A can be seen as learning the structure of the corresponding BN from data $\mathcal{D}$. We implement an information-based score maximization approach for finding the BN structure

from $\mathcal{D}$. The score of the BN structure $\mathcal{G}$ is defined as the posterior probability given data $\mathcal{D}$.

We implemented a hill-climbing algorithm for structure learning with an information-based selection criterion [8]. Applying inference over the learned BN, the robot can estimate probability distributions for effect prediction $P(E|O, A)$, feedback in action selection $P(A|O, E)$ or object recognition given its behavioral description $P(O|A, E)$.

Bayesian inference in our discrete BN provides the probability that an affordance α_i is present or not. However, it does not provide a mechanism for quantifying the affordance w.r.t. the variables of the specific situation in which it was discovered. For example, if we consider the push-ability affordance of an object on a table, it can exhibit this affordance with a certain degree depending on the table surface, its weight and the force of the action.

We therefore quantify the elements defining the affordance in (1) and maintain the necessary information of affordance quantization in a Gaussian BN (GBN). Relations in (1) can then be represented as a multivariate normal distribution of continuous random variables, i.e. the affordance elements.

In addition, we can discover quantitative relationships between the objects, actions, and effects involved in an affordance by employing continuous values for the random variables. We also propose to study the notion of affordance combination. The underlying principle is to discover the affordance of an assembled object from the affordances of its components.

III. Experimental results

The goals of our ongoing experiments are to discover the affordances of objects, to attach a quantitative aspect to them, to generalize affordances over similar objects, and to predict the affordances of compound objects, based on the affordances of their components.

We use the Baxter experimental platform equipped with two 7 d.o.f. arms. One electrical gripper is attached to each arm. For visual perception we use a Microsoft Kinect that captures RGB-D data.

We performed experiments to infer the affordances of objects created by composition of other objects, depending on the way they are assembled (see Table I). As an example, the system learned that a composite object *cart*, composed of the objects *wheels* and the object *body* will maintain the affordance *roll-ability* of *wheels* in terms of traveled distance after action *poke*, only in the case where *body* is on top of *wheels* (Table I column 1). Combining the new object *cart* with another object *block* on top maintains the *roll-ability* affordance of *cart* (Table I column 2), while this affordance is lost if *cart* is put on top of *block* (Table I column 3).

Fig. 2 shows the Gaussian BN learned by our architecture. If the value of the *objectBottom* variable is known, the variables *distance* and the *objectTop* are conditionally independent. This means that the upper part of a composite object does not influence the distance that this composite object traverses after action *poke*. Only *objectBottom* transmits its affordance. This was confirmed experimentally.

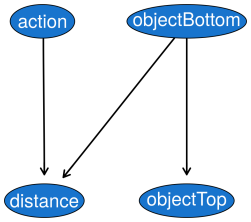
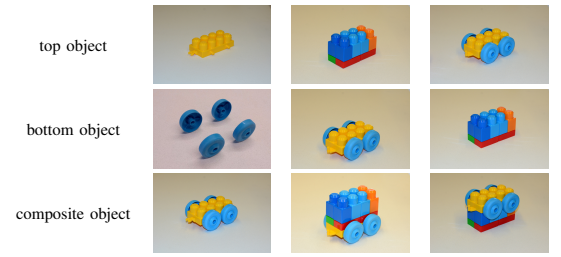


Fig. 2. Learned Gaussian BN for the experiments depicted in Table I.

TABLE I

Objects employed and their composition order.



Conversely, we can disassemble an object, to identify which parts are required for a given affordance.

This inference of affordances of assembled/disassembled objects, together with their empirical verification, opens a whole new research direction about the relations between affordances.

Acknowledgment

This work has been funded by a DGA (French National Defence Agency) scholarship (ER), and by French Agence Nationale de la Recherche ROBOERGOSUM project under reference ANR-12-CORD-0030.

References

[1] Gibson J J, *The theory of affordances, in Perceiving, Acting, and Knowing. Towards an Ecological Psychology.*, S. R. and B. J, Eds. Hoboken, NJ: John Wiley & Sons Inc., 1977.

[2] E. Ugur, Y. Nagai, E. Sahin, and E. Oztop, "Staged Development of Robot Skills: Behavior Formation, Affordance Learning and Imitation with Motionese," *Autonomous Mental Development, IEEE Transactions on*, vol. PP, no. 99, p. 1, 2015.

[3] B. Moldovan, P. Moreno, and M. Van Otterlo, "On the use of probabilistic relational affordance models for sequential manipulation tasks in robotics," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 1290–1295, 2013.

[4] A. Dehban, L. Jamone, A. R. Kampff, and J. Santos-Victor, "Denoising Auto-encoders for Learning of Objects and Tools Affordances in Continuous Space," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 4866 – 4871.

[5] E. Sahin, M. Cakmak, M. R. Dogar, E. Ugur, and G. Ucoluk, "To Afford or Not to Afford: A New Formalization of Affordances Toward Affordance-Based Robot Control," *Adaptive Behavior*, vol. 15, no. 4, pp. 447–472, 2007.

[6] D. Vernon, C. Hofsten, and L. Fadiga, *A Roadmap for Cognitive Development in Humanoid Robots*, 2011, vol. 11.

[7] E. Ugur and E. Sahin, "Traversability: A Case Study for Learning and Perceiving Affordances in Robots," *Adaptive Behavior*, vol. 18, no. 3-4, pp. 258–284, 2010.

[8] I. Tsamardinos, L. E. Brown, and C. F. Aliferis, "The max-min hill-climbing Bayesian network structure learning algorithm," *Machine Learning*, vol. 65, no. 1, pp. 31–78, 2006.

References

[1] N. Lyubova, "Developmental approach of perception for a humanoid robot," Ph.D. dissertation, Ecole Nationale Sup\'erieure de Techniques Avanc\'ees - ENSTA, 2013.

[2] P. Haazebroek, S. van Dantzig, and B. Hommel, "A computational model of perception and action for cognitive robotics." *Cognitive processing*, vol. 12, no. 4, pp. 355–65, nov 2011.

[3] J. Bohg, K. Hausman, B. Sankaran, O. Brock, D. Kragic, S. Schaal, and G. Sukhatme, "Interactive Perception: Leveraging Action in Perception and Perception in Action," *IEEE Transactions on Robotics*, 2016.

[4] D. Hoiem, A. a. Efros, and M. Hebert, "Closing the loop in scene interpretation," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. Ieee, jun 2008, pp. 1–8.

[5] P. Langley, J. E. Laird, and S. Rogers, "Cognitive architectures: Research issues and challenges," *Cognitive Systems Research*, vol. 10, no. 2, pp. 141–160, 2009.