Affordance Research in Developmental Robotics: A Survey

Huaqing Min, Chang'an Yi, Ronghua Luo, Jinhui Zhu, and Sheng Bi

***Abstract*—Affordances capture the relationships between a robot and the environment in terms of the actions that the robot is able to perform. The notable characteristic of affordance-based perception is that an object is perceived by what it affords (e.g., graspable and rollable), instead of identities (e.g., name, color, and shape). Affordances play an important role in basic robot capabilities such as recognition, planning, and prediction. The key challenges in affordance research are: 1) how to automatically discover the distinctive features that specify an affordance in an online and incremental manner and 2) how to generalize these features to novel environments. This survey provides an entry point for interested researchers, including: 1) a general overview; 2) classification and critical analysis of existing work; 3) discussion of how affordances are useful in developmental robotics; 4) some open questions about how to use the affordance concept; and 5) a few promising research directions.**

***Index Terms*—Affordance, autonomous mental development (AMD), developmental robotics, embodied intelligence (EI), infant learning, intrinsic motivation.**

I. Introduction

HUMANS are good at incrementally constructing the action relationships between objects in the world and their own capabilities and using these relationships to change their environment and complete tasks. For example, a new football player knows how to kick a ball and this skill will become increasingly proficient if he keeps practicing. Humans can also cooperate with partners or select tools to extend their strength and scope. They then use these tools to build complex relationships based on their own capabilities. Building robotic systems with human level intelligence is the ultimate goal in developmental robotics [1].

Manuscript received August 20, 2015; revised February 15, 2016, June 6, 2016, and September 13, 2016; accepted September 27, 2016. Date of publication October 4, 2016; date of current version December 7, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61203310, Grant 61262002, Grant 61300135, and Grant 61372140, and in part by the Key Laboratory of Robotics and Intelligent Software of Guangzhou under Grant 1518007. *(Corresponding author: Chang'an Yi.)*

H. Min and J. Zhu are with the School of Software Engineering, South China University of Technology, Guangzhou 510006, China (e-mail: hqmin@scut.edu.cn; csjhzhu@scut.edu.cn).

C. Yi is with the School of Electronic and Information Engineering, Foshan University, Foshan 528000, China (e-mail: yichangan.ok@gmail.com).

R. Luo and S. Bi are with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China (e-mail: rhluo@scut.edu.cn; picy@scut.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCDS.2016.2614992

Robots need to develop skills, such as navigation and manipulation, in order to work and live with human beings. However, they still perform poorly in dynamic, complex, or real-time environments. Introducing biology-inspired principles is a good way to endow a robot with developmental capabilities [2]. This research methodology is also useful in understanding how humans develop [3].

In 1977, the American psychologist Gibson [4] proposed the affordance concept to explain how inherent *meanings* and *values* for things in the environment can be perceived directly. The core issue is:

"The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill."

The affordance concept is a kind of perception based on actions. Its essence is that an object's category is jointly determined by the action and its associated effect. Gibson [4] does not discuss how an artifact builds its knowledge about affordances and yet there is no scientific theory that fills this gap properly. The affordance theory has gained widespread influence in different fields, ranging from product design [5], human–computer interaction [6], to developmental robotics [7]. For example, the EU-MACS project focuses on designing and implementing affordance-based goal-directed mobile robots that act in a dynamic environment [8]. Some workshops were held to discuss affordance research in robotics and computer vision [9], [10], with a special issue on affordance computational models for cognitive robots to be published in an international journal [11].

Affordance research in developmental robotics is becoming increasingly influential, but different researchers use it in different ways, so a summary work is necessary to provide a roadmap in this field. The survey from Horton *et al.* [12] only discusses the influence of affordance theory on the design of robotic agents at the conceptual level, without mentioning the action relationships between the robot and environment, or mentioning the learning mechanisms. Jamone *et al.*'s [13] survey reports on the most significant evidence of affordances from psychology and neuroscience, and reviews the related work on affordances in robotics. The focus of [13] is to offer a structured and comprehensive affordance work report from these different fields without giving a critical analysis of existing affordance work in robotics or presenting the open issues and future directions in detail. In contrast, this paper aims to position the current state of affordance-based research in developmental robotics, make critical analysis of the formalizations and implementations, and discuss some questions to provide an entry point in

this field. The contributions of this survey are summarized as follows:

1) it provides some new insights into the research field, such as a definition, a developmental path, questions to further use and understand affordances, what remains to be solved and how they might be solved;
2) it gives a critical discussion of existing work such as the formalizations and makes cross comparisons of leading approaches that implement affordances;
3) it discusses affordance research in the context of life science and presents its advantages when applied in classical fields such as planning, recognition and control.

The focus of this paper is on articles that explicitly use the word *affordance*, although in the robotics literature many other works relevant to affordance research have explored similar questions (e.g., object manipulation, social interaction, or navigation) without using the word *affordance*. For example, Klank *et al.* [14] proposed a novel method to finish a manipulation task using different objects and scenarios, where the robot could choose perception models automatically and correctly. In the work of [15], contextual information that is about the type of object being acted on is provided by the manipulation movements, while contextual information about the possible interactions with an object is provided by the object's category. The works of [16]–[18] provide a functional understanding of the environment and they could also predict an object's attributes based on its functionalities.

The rest of this paper is organized as follows. Section II presents the general idea of affordances. Section III provides the basis for affordance research. Section IV connects the implementation mechanisms of existing work, including the algorithms, representations, prior knowledge, advantages, and disadvantages. Section V discusses the developmental learning associated with affordances. Section VI presents the advantages of affordances compared to classical research. Section VII discusses some questions involving better use and understanding of affordances, in order to bring more inspiration into developmental robotics. Section VIII presents some affordance related topics in developmental robotics that should be solved in the future. Section IX concludes this paper.

II. Glance at Affordances

As a perceptual psychologist, Gibson [4] does not have special interest in development [19] and his definition highlights the direct perception of an organism. However, an organism's physiological functions are different from those of a robot. While respecting the original definition we give a more detailed one from a developmental point of view.

Affordances are represented in terms of two critical ingredients: (1) the potential actions between the robot(s) and environment, i.e., object(s), tool(s), human(s), and other robot(s) and (2) the effects of those potential actions. Such relationships are learned and developed for two reasons: extrinsic motivation and intrinsic motivation. Examples of extrinsic motivations are tasks and rewards, while examples of intrinsic motivations are novelty and curiosity. A robot's sensorimotor capabilities will also develop together with affordances.

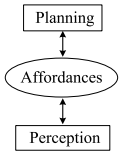

Fig. 1. Role of affordances in robot-environment interactions. The low-level layer processes signals coming directly from sensors and actuators. The high-level layer could reason on symbols. A bi-directional arrow represents that each side can affect the other.

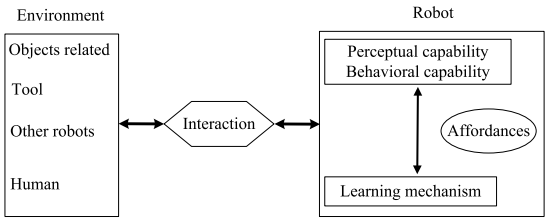

Fig. 2. Affordances in a robot's world. Through interactions with the environment the robot could incrementally and continuously develop its intelligence.

In developmental robotics affordances can be used to bridge the gap between low-level sensory-motor representations that are necessary in the perception and control of a robot, and high-level representations that are required in abstract reasoning and planning, as indicated in Fig. 1. Most work in classical artificial intelligence (AI) focuses on high-level symbolic planning but neglects the low-level sensory-motor information. Building symbolic representations from continuous sensory-motor experience is a prerequisite if a robot is anticipated to perform at levels comparable to humans. This research has gained a great deal of attention in the last decade such as [20]–[22].

Affordances are quite useful because they represent the essential properties of the environment and objects through the actions that the robot would perform [23]. For example, a pen might be graspable while a box might be pushable for the same robot. These action-based relationships are gained and maintained from the embodied interaction processes with the environment and they exist independent of being perceived or not. The robot should first be equipped with some behavioral and perceptual capabilities, which are simple but enough to drive itself to learn increasingly complex affordances [23].

According to [4], an organism's knowledge about the environment is learned and stored in the form of affordances. The relationship between the affordances (the action relationships), robot and environment is described in Fig. 2, where three points should be stressed:

1) an affordance can be stored, used, predefined, learned, or updated and can also be searched from the Internet;
2) the interaction targets in the environment could be objects, tools, robots, and human beings;
3) the learning mechanisms are mutually developed with the perceptual and behavioral capabilities.

Gibson [4] believed that some affordances are learned when an infant plays with objects. Infants first study the affordances of objects using motor activity and then they start

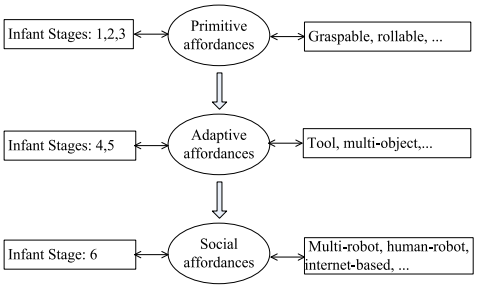

Fig. 3. Developmental path of affordances. The six stages of infant development are: simple reflexes (birth—one month old), primary circular reactions (1–4 months old), secondary circular reactions (4–8 months old), coordination of secondary circular reactions (8–12 months old), tertiary circular reactions (12–18 months old), and internalization of schemes (18–24 months old) [24].

to recognize their properties [3]. A truly autonomous *infant robot* is believed to be a key criterion in the construction of developmental robots [2]. As a result, affordances are anticipated to develop in a way similar to that in an infant, i.e., online, incremental, continuous, generalizable to novel environments, capable of self-discovery, and multistep planning. Each existing work could only meet a few requirements.

Based on the six sensorimotor stages of infant development from birth to two-year-old, as proposed by Piaget [24], the developmental pathway for affordances is indicated in Fig. 3. For example, a human infant develops some primitive affordances such as graspable and droppable for a single-object by nine months. The insertable affordance between multiple objects will be developed by 13 months [25]. Most current works are about primitive affordances, typical works for adaptive and social affordances are [26]–[29], respectively. Social affordances represent affordances whose existence requires the presence of humans and can be used for planning [29]. A robot that could learn complex affordances could also learn simpler affordances. Primitive affordances can be developed into complex systems and there should be interfaces that allow humans to interact with the robot naturally and efficiently [30]. The development of affordance knowledge should not strictly follow the "simple to complex" manner. Affordances in different stages should be overlapping and the affordances learned later might be helpful to strengthen or even correct the ones learned earlier. For instance, an infant's hands develop during the first ten months [31], but they would become more experienced in later life.

III. Basis of Affordance Research

A developmental robot should be able to interact with its surrounding physical world and manipulate objects in an adaptive, productive, and efficient manner. Affordance-based perception provides a basis for the cognitive capabilities of a robot, because it gives the robot early awareness of opportunities for interactions and leads to a revolution in object perception by shifting the focus from visual attributes to functional attributes [32]. Affordances should be used in predicting the effects of actions as well as planning to finish tasks through multiple steps. When carrying out affordance research in developmental robotics the researcher should first become familiar with some basic knowledge, as listed below.

A. Characteristics

An affordance relationship has several commonly associated characteristics, some of which are adapted from [33].

1) Potential to Happen: There might be many possible actions for a robot in a goal-free manner, for instance, a cup is both graspable and pourable. However, the robot should select the most effective action based on the current goal. Take housework as an example, a cup might first be reachable, then graspable and finally pourable.

2) Economical in Perception: Affordance learning discovers the distinctive features and invariant properties related to a robot action [34]. The perception cost would be reduced because the robot does not need to analyze all of the qualities in a given environment.

3) Co-Determined to Exist: An affordance is jointly determined by the robot and object, as a result, the potential action would be different if any of them changes [35]. For example, if a cylinder is too thick to be grasped by one robot, it might be graspable for another robot with a larger gripper. This characteristic is more notable in the domain of multirobot cooperation, where an obstacle is moveable only when two robots work together to move it toward the same direction.

4) Generalizable to the Application: Affordances are often described as general relations such as traversability and climbability. For instance, based on the training result with simple objects in simulation, a real robot could perceive traversable affordances in more complex environments [36]. An affordance can be represented in a general triple form [effect, (object, behavior)] which will be discussed in detail in Section III-C.

B. Affordance Perception

The affordance concept is proposed based on the direct perception in organisms where vision is the main information source. However, in order to interact with the world in depth, the robot needs to perceive the environment in different ways to gain information as completely as possible.

1) Some Background Knowledge is Necessary: An affordance might be activated by the combination of current perception and existing knowledge [37]. For example, a pourable cup is put inside a container and only a logo is printed outside the container and now we also know that the object inside is pourable, a texture representation combined with the material are necessary to perceive openable affordances [38].

2) Gain More Visual Information Through Manipulation: An affordance is sometimes hidden because of orientation. As we can select what to look at through hand manipulation, objects near the hands would have priority in visual perception [39]. Vincze *et al.* proposed that a robot can manipulate to modify the object configuration until the hidden affordance appears, and the robot's manipulation capability is also improved during this process [40]. Under the use of Kinect and

TABLE I
TYPICAL FORMALISMS FOR AFFORDANCES

Formalism	Study
A potential which only depends on the environment	Turvey *et al*. [41]
The properties of the animal-environment system	Stoffregen [42]
Relations between the abilities of organisms and features of the environment	Chemero *et al*. [43]
A formalism for reasoning about causal relations over events, using Linear Dynamic Event Calculus	Steedman *et al*. [44]
(*effect*, (*object, behavior*)) in a deterministic manner	Sahin *et al*. [45]
(*action, object, effect*) in a probabilistic manner	Montesano *et al*. [23]
O→AO', object O suggests action A and transforms under this action into object O'	Worgotter *et al*. [21, 46]
(*subtask*, *precondition*, *action*, *post condition*)	Yi *et al*. [47, 48]
$M(a_i) \equiv Pr(f \mid r, a_i)$, captures the likelihood of achieving reward r for taking action a_i after observing the environmental context f	Hart *et al*. [49,50]

manipulation behaviors, the robot could detect hidden affordances in real scenes through object recognition and 6 DOF pose estimation [40].

C. Formalization

A number of studies, from the psychology and robotics fields, have attempted to make a unified understanding of affordances. Typical examples are described in Table I. Our focus is the latter five examples that are robot-related.

The most widely cited formalization, proposed by Şahin *et al.* [45], is used mainly as the basis for autonomous robot control. It states that a potential *action* exists that could generate the *effect*, if applied to a specific *object* in the environment. A single interaction will produce an instance of this triplet (*object*, *action*, *effect*). Multiple interactions could develop a planning tree through forward chains and then make multistep predictions to achieve complex goals. With these affordance instances the robot could also predict the *effect* of an action, choose a number of pairs (*object, action*) to obtain the anticipated goal, and select an object to gain a specific effect.

The formalism, called "object-action complexes (OACs)" and proposed by Krüger *et al.* [21] and Worgotter *et al.* [46], is grounded in the robot's sensor data and used in higher-level planning such as predicting the moveable affordance based on an object's visual appearance. The OAC concept states that the execution of an action should not be separated from the object that the action involves. This coupled viewpoint of objects and actions is associated with the affordance theory, and an instance of affordances could be considered as the precondition of OAC instantiations.

The formalisms proposed by Şahin *et al*. [45] and Worgotter *et al.* [46] are roughly equivalent in four aspects:

1) they both learn casual relationships between objects and actions;
2) they often define a priori knowledge which could reduce the problem dimensionality;
3) they formulate robot actions as preprogrammed activities that require little or no sensory feedback to check whether an action has been successfully executed;
4) an affordance only points to a reactive or primitive action and the task is completed after that action has been executed.

Nevertheless, a complex task might be decomposed into several subtasks, where an object's affordance will influence the efficiency of a subtask. Min *et al.* [47], [48] proposed a novel formalization to describe affordances in a dynamic environment, where the robot could filter out some irrelevant affordances to reduce the state space. The influence of an affordance is then promoted to the subtask level. The formalization proposed by Montesano *et al.* [23] is very similar to that of [45], and their difference is that Montesano *et al.* [23] used the formalization to represent uncertain interactions.

The formalizations proposed by Krüger *et al*. [21], Montesano *et al*. [23], Şahin *et al*. [45], Worgotter *et al*. [46], and Min *et al*. [47], [48] use symbolic or textual analysis to describe the knowledge learning process and lack the sensorimotor signals that are measurable to the robot, thus the robot might have difficulty finishing a task in the real world [51]. In contrast, Hart *et al.* [49], [50] contributed a novel formalization in which any behavioral affordance was explicitly grounded in the robot's dynamic sensorimotor interactions with its environment, the closed-loop feedback control framework could be used to track time-varying reference signals to determine the robot's movements. The implementation mechanism and advantages of [49] and [50] will be discussed in detail in Section IV.

D. Representation Form

Affordances are learned and updated during the interaction process so the researchers should first designate a structure to represent the affordance relationships. The representation form allows the researchers to take advantage of methods proposed in the machine learning community for learning, inference and planning. Some algorithm might be better than another for a given task, for example, a Markov random field (MRF) structure performs better than a simple feature-based support vector machine (SVM) classification when capturing the context [52].

As most interactions are uncertain, the representation form could be deterministic or probabilistic. The deterministic forms are usually represented as transitional rules and stored in a table. Most of the rules are based on the formalism proposed by Krüger *et al.* [21], Şahin *et al.* [45], Worgotter *et al.* [46], or Min *et al.* [47], [48]. The probabilistic forms could model the uncertainty of the interactions and one typical example is described in Fig. 4 which will be discussed further in Section IV.

E. Interaction Style

The interaction style indicates the way a potential action happens. Generally, there are three styles: 1) hand-to-grasp; 2) object-to-surface; and 3) object-to-object.

1) Hand-to-Grasp: The purpose is to find the graspable point of the object for later manipulation, inspection or other complex behaviors. Graspable affordance is the most popular, and most of them should be combined with different methodologies such as vision [53] and experience [54], [55].

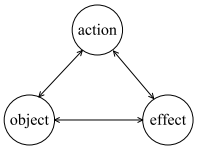

Fig. 4. Pioneering work that considers the uncertainty of the real world represents affordances in a probabilistic form [23]. It describes the statistical relationships of the actions, objects, and observed effects. A probabilistic graphical model known as a Bayesian network (BN) is used to represent such dependencies. Note: this figure is adapted from [23].

2) Object-to-Surface: The typical purpose is to operate in the best direction, e.g., moveable [56], pourable [57], pullable [58], and traversable [59], [60]. The interaction target is the surface of the object and the robot does not necessarily have to touch an object, for example, traversable affordances [33] are to freely navigate in an environment cluttered with objects.

3) Object-to-Object: A robot often needs to interact with multiple objects, such as stackable which should make the objects stable [28], sortable which considers the spatial relations [27], and tool-object which considers extending the functionality of the robot [26].

The first two interaction styles consider the object as a whole. The third style is interested in the object's parts, e.g., a hollow surface would enable a robot to put something inside. These three styles can also be roughly divided into two categories: 1) locomotion such as traversable affordances and 2) manipulation.

F. Robotic Platform

This paper is carried out in simulation or using real robots. A robot's simulator software plays a complementary role because it allows the researchers to take infinite and systematic experimentation [61], and it can also protect the robot's hardware from too many interactions with the real environment. If the robot only needs to study the strategy based on Markov decision process (MDP) without perceiving the complex environment, a simulation environment is enough [62]. However, in most tasks real-world experience is irreplaceable for a robot although it is costly to gain.

Robots with different capabilities are suitable for solving different problems, for instance, equipping the robot with a grip or finger is necessary for graspable tasks, a wheeled mobile robot is preferred in traversable prediction because its movement is more precise than a biped robot. To perceive the natural environment as much as possible, sensors such as Kinect are necessary. For example, the wrist angle of the hand in graspable affordance is decided by the axis of the object, soft skin with tactile sensors is necessary in gaining touchable affordance. Platform improvement might lead to more analysis on the functional relationships between a robot and an organism [3].

G. Classification of Affordance Learning

Because the real world is so changeable and current robotic platforms are equipped with different capabilities, it is too difficult to establish a unified affordance learning algorithm that works well in different robots and environments.

Based on the learning algorithm to learn affordances, this paper can be classified into three categories: 1) supervised; 2) un-supervised; and 3) reinforcement learning (RL). Existing algorithms are selected from classical AI, and as far as we are concerned no algorithm is invented specifically for affordance research. Based on the robot's interaction target, this paper can be classified into four types: 1) manipulative [23]; 2) traversable [33]; 3) human–robot environment [29], [52], [63]; and 4) tool-use [26]. Through the methods used to gain affordances, existing work could be classified into four categories: 1) by observing the effects of exploratory behaviors [64], [65]; 2) by visual cues of the object [40], [60], [66]; 3) by human-object interaction [52], [63]; and 4) by tool-use [26].

An affordance is generally intended as a property emerging from the interaction process between the robot and environment, and it is discovered through algorithms in the machine learning community. Generally, affordance research consists of four steps:

1) define and represent the affordance relationships, which is also called an affordance model or computational model;
2) make clear which parameters should be learned;
3) select a learning algorithm to learn the relationships;
4) develop affordances incrementally and continuously during task execution.

IV. Existing Implementation Mechanisms

An affordance encodes the relationship between a robot and environment in terms of an action that the robot is able to perform [23]. The taxonomy of the implementation mechanisms will depend on the representation style, i.e., probabilistic or deterministic. Existing implementations are briefly summarized in Table II, where the rows match the order of the sub-subsections in this section.

A. Probabilistic

In the real world an action exists probabilistically, depending on many properties of the run-time environment. For instance, a box is moveable but the difficulty level associated with this affordance depends on the box size and the friction of the ground. The typical probabilistic representations include Markov and BN structures. Each of them could be a chain or network, thus they are good at capturing the relationships between any elements.

1) Markov Structure: A markov-based affordance model can be used to understand the human activities and object properties in 3-D environments [52], indirectly capture the communications between humans and robots based on the name of an action or an object [63], model the human body based on the behaviors [51], etc.

a) 3-D environment understanding: A robot needs to understand human activities and object affordances in order to provide better service in a human environment. In the human activity detection task an object's affordance, which also describes how the object could be used, would provide more information compared to its name or other features [52]. In [52], the robot could detect human activities and object

TABLE II
BRIEF SUMMARY OF EXISTING IMPLEMENTATION MECHANISMS

Affordance representation	Learning method	Application scene	Robot platform	Papers
Probabilistic	Markov random field	Personal robot	PR2	[52, 67, 68]
	Conditional random field	Personal robot	PR2	[69]
	Mixed-order Markov chain	Human-robot communication	WorkPartner	[63, 70]
	Markov random field based on And-Or graph	Manipulation	Dexter	[51, 71]
	Generative Bayesian network	Manipulation	Baltazar	[23, 72]
	Statistical relational learning	Manipulation	iCub	[27, 73, 74]
	Generative Bayesian network	Navigation	LAGR	[60, 75]
	Discriminative Bayesian network	Navigation	LAGR	[60, 75]
	Reinforcement learning based on intrinsic motivation	Manipulation	Dexter	[49, 50, 76, 77]
Deterministic	Hierarchical clustering followed by SVM	Manipulation	Gifu hand	[78]
	SVM based on intrinsic motivation	Manipulation	Kuka	[25]
	SVM	Navigation	KURT3D	[33, 59]
			Pioneer	[66]
	Ontology represented by table	Personal robot	Aibo	[79, 80]
	Reinforcement learning	Tool using	CRS+A251	[26, 81]
	Reinforcement learning based on genetic algorithm	Manipulation	NAO	[82]
	K-means	Manipulation	PR2	[83]

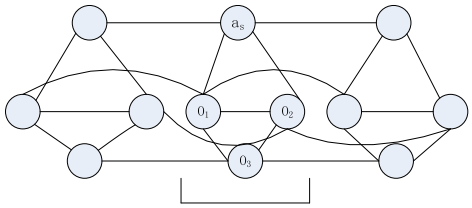

Fig. 5. Affordance model based on MRF [52]. Each of the three temporal segments has one activity node (a_i) and three object nodes (o_i). Note: this figure is adapted from [52].

affordances from RGB-D videos, and could also actively figure out how to interact with objects and plan actions. The affordance model, based on MRF, as shown in Fig. 5, consists of semantic, 3-D spatial and temporal trajectory components [52]. The model is defined based on two kinds of nodes: 1) subactivity such as reaching and opening and 2) object. The edges represent the relationship between object affordances and also their evolution over time and their relations with subactivities. The human activities and object affordances are jointly represented in the affordance model which could integrate all of the information available in human interaction activities. In the learning algorithm, the temporal segments are labeled as latent variables. In later work, they used a predefined color to emphasize the most likely affordances in a heat map generated for each affordance by scoring the points in the 3-D space using the potential function. Each score represents the degree of the specific affordance at that location [84].

The model in Fig. 5 was then extended to model the uncertainty of the grounded affordances [68], as indicated in (1). An affordance is described as a triplet proposed by Şahin *et al.* [45]. $\mathcal{L}, \mathcal{S}, \mathcal{T}$, respectively, denote semantic affordance, spatial distribution of the affordance, and the motion trajectory. H is the robot's capability, I represents the robot's intention, and E is the environment which might include many objects. The goal of this model is to infer the probability of a particular $<\mathcal{L}, \mathcal{S}, \mathcal{T}>$ given the context (H, E, I). This model has three advantages:

1) it could achieve high accuracy in labeling affordances, subactivity, and high-level activity from RGB-D videos;
2) as the human-object relationships are modeled as object affordances, it provides a more compact way to model the contextual relationships than to model the object-object relationships [69];
3) it could make a robot perform better in assistive tasks [84].

Nevertheless, the limitations of their work are that the robot could only respond with preprogrammed actions and the activity anticipation accuracy falls rapidly over long-time periods [67]

$$\varsigma(<\text{affordance}>|H, E, I) = <\mathcal{L}, \mathcal{S}, \mathcal{T}>. \quad (1)$$

b) Indirect human–robot communication: Known object properties could constrain the possible tasks that we can perform with the object. For example, when we ask another person to drink tea we only need to say a word *drink*. Heikkilä *et al.* [63], [70] formulated a new affordance-based method for face-to-face astronaut-robot task communication, in which the words used are as few as possible. The astronauts are able to communicate using only the task-related action or target object name, without remembering full task

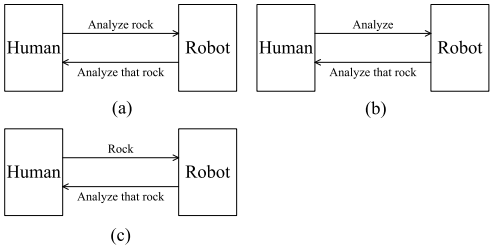

Fig. 6. Indirect human–robot communication based on affordances [63]. Parameters above the arrows are affordance related, and the ones below the arrows represent the full request understood by the robot. Take (b) for example, when the human says *analyze* to the robot, the robot understands that *rock is analyzable* and it should *analyze that rock*. Note: this figure is adapted from [63]. (a) To communicate explicitly all the parameters are required. (c) Indirect affordance-based communication where only the object's name is required in the task.

request utterances. Three examples used in the experiment are illustrated in Fig. 6, where a mixed-order Markov model-based sequence prediction algorithm is used. The subjective human workload as well as the total test round execution times are both decreased, and this affordance relationship allows the robot to understand action possibilities in a human-like manner. However, a robot in this case could only work with known objects and this limitation would prevent the robot from working in general environments.

c) Human modeling based on behavior: The human body has many parts, and one part such as a hand may have some subparts which can afford actions individually or cooperatively. Based on the work of Hart *et al.*, [76] which will be discussed later in this section, Ou *et al.* [51], [71] proposed a behavioral approach for human–robot communication and they modeled a human as a collection of behavioral affordances that are discovered by the robot. The relational constraints between affordances are formulated in an "AND–OR graph" and modeled as an MRF. The principle under that approach is similar to the observations of mirror neurons from neuroscience and psychology literature [85]. In the experiment a robot could develop increasingly complex affordance models of humans through action exploration. However, the development of verbal related affordances is not addressed in their work, although this capability is also very important in communication.

2) Bayesian Structure: We can use the bayesian structure to build affordance models based on the elements from a triplet (object, action, effect) [23] or the categories of objects [60].

a) Bi-directional mapping: Bayesian methods are extremely powerful in affordance learning because they take into account the uncertainty of the physical interactions between objects and effectors and also consider multiple action possibilities provided by objects to complex behaviors [86]. Montesano *et al.* [23] and Lopes *et al.* [72] considered affordance learning as a structure learning problem, where affordances represent the probabilistic relations between actions, objects, and effects. The graph for the network (Fig. 4) is inferred directly from the proprioceptive and exteroceptive measurements, without assuming any prior knowledge of the probabilistic relations. The robot first studies its sensory-motor capabilities and then interacts with the environment to learn the objects' affordances. The BN could implicitly represent affordances as mappings from actions to effects, and these mapping are mediated by the objects' visual features [87]. The learned affordance network could then be used in a bi-directional way and its functionality is indicated in Table III. In the imitating learning experiment, the tappable, graspable and touchable affordances are learned from both self-observation and imitation. The model could also be used to perform goal-directed imitation, which has gained much attention in developmental robotics [1].

TABLE III
ADVANTAGE OF LEARNED AFFORDANCES: BI-DIRECTIONAL INFERENCE [23]

Input	Output	Functionality
(object, action)	effect	Predict effect
(object, effect)	action	Action planning & recognition
(action, effect)	object	Object selection & recognition

Montesano *et al.* [23] put affordances at the center of developmental process, containing three stages: 1) sensory-motor coordination; 2) world interaction; and 3) imitation. The drawback of this approach includes: 1) one individual network should be trained for each object and 2) only one-step planning could be made. Based on their model, Moldovan *et al.* [27], [73], [74] employed recent advances in statistical relational learning to propose an affordance model that can generalize to any number of objects and also effectively deal with uncertainty based on probabilistic logic-based techniques. Krunic *et al.* [88] extended the affordance model to incorporate verbal descriptions of a task to produce links between speech utterances and the involved objects, actions, and outcomes. Although the robot could develop a basic understanding of speech through its own experience, the human–robot interaction is very simple and the language between the instructor and the robot is not flexible.

b) Category-affordance perception: Navigability is a challenge in RoboCup Rescue competitions where part of the task is to predict the traversable affordances of obstacles in a ragged and cluttered environment. Sun *et al.* [60] proposed a probabilistic graphical model that utilizes visual object categorization, which is a latent variable to connect object appearance and affordances, for affordance learning. The core idea is that object categorization can be used as an intermediate representation that makes affordance learning and prediction more tractable. For example, the category of *stone* or *chair* is closer to the essential features compared to the appearance and we prefer to detect an obstacle's moveable or un-moveable affordance through its category. The category-affordance perception model is shown in Fig. 7 and the experiment is conducted using a real robot [60], [75]. The key contribution of their work is to provide a computational model to integrate object categorization and affordance prediction. As a result a robot could predict the affordance before actually performing the action, instead of learning the affordance after the experiment is completed. Such prediction capability is more useful in environment understanding

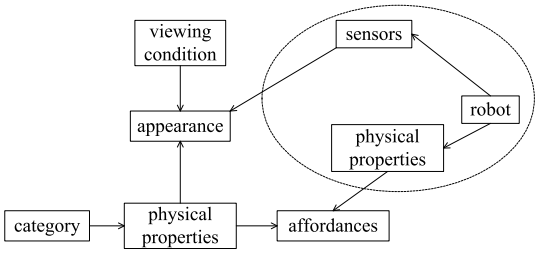

Fig. 7. Category-based affordance model [60]. It assumes that an affordance is jointly defined by the physical properties of the robot and object. The object's appearance perceived by the robot is a probabilistic function of its physical property, viewing conditions, and the robot's sensors. Note: this figure is adapted from [60].

and task planning. Nevertheless, the images are segmented manually and it is questionable that using global maps is orthogonal to Gibson's view of direct perception. On the other hand, they do not consider object affordances in the human context and a robot is difficult to work with well when the category or feature changes.

3) Probability From Sensorimotor and Environment Information to Reward:

a) Sensorimotor interaction and intrinsic motivation: RL could guide the development through self exploration [64], and it can be used to learn predictive visual cues [89] or hierarchical programs [50]. According to how to define the reward, RL can be divided into two categories: 1) intrinsic motivation or 2) not.

Hart [76] proposed a closed-loop control quadruple $c \equiv \phi|_{\tau}^{\sigma}$ to describe affordances, where C, ϕ, σ, and τ represent the control operator, potential function, sensory signal, and effector signal, respectively. Affordances are discovered by an intrinsic reward and the robot could integrate closed-loop strategies for interactive behavior [77]. From the control standpoint the discovery of an affordance is measured by the convergence event

$$\left(c(\phi, \sigma, \tau)_i^{t-1} \neq 1\right) \wedge \left(c(\phi, \sigma, \tau)_i^{t} = 1\right) \tag{2}$$

where $c(\phi, \sigma, \tau) = 1$ occurs when either a controller reaches an attractor state in its potential or when a lack of progress along the gradient of that potential is observed. Both of these conditions are predefined by the user. i and t represent the number of quadruples and steps, respectively. Their affordance catalogs could be extended to model multiobject assemblies in a controlled manner, where the affordance relationships are acquired in terms of both visual and force domain. In the bimanual tasks, a robot develops hierarchical control programs for affordances searchable, trackable, reachable and touchable based on the feedback from sensors. In later work, they extended the same motivator to create probabilistic models concerning the condition in which the environment affords these strategies [90].

The advantages of that methodology are as follows:

1) generalizable to new contexts by abstracting sensory and effector information into environment description;
2) a single intrinsic motivation function to discover affordances could guide long-term learning, and behaviors could be derived through a staged developmental process;
3) adaptable to act in unstructured environments because a behavior is discovered by combining the sensorimotor resources.

However, that approach also has some disadvantages:

1) the intrinsic reward is task-specific, the robot motor skills as well as the control programs are preprogrammed, so it is difficult to apply the proposed techniques to different robots with different morphologies;
2) the robot does not perform any low-level motor learning, as a result the resulting policies cannot be improved if the controllers are sequenced [91].

B. Deterministic

This kind of representation mainly uses binary assertions, either exist or not, to describe affordances.

1) SVM Structure: An SVM structure can be used to deterministically describe the affordance relationships in the form of a triplet (object, action, effect) [22], or predict the potential actions from visual attributes [66].

a) Object-action-effect: Ugur *et al.* [30], [36] are representative researchers in affordance research. Their work lasts from the beginning formalization work [45], traversability [33], and single-object manipulation [78] to the recent stackability work for multiple objects [28] and staged development [30], from a simulation environment to anthropomorphic or mobile real robots, from goal-free exploration to goal-directed planning, and from open-loop nonparametric or parametric to closed-loop behaviors [59]. In their earlier work, a breadth-first plan tree could be constructed based on the learned affordance relations [59], [92]. In later work, they manually designed hierarchical structures where prelearned basic affordances can be used as an input to drive the learning of complex affordances [30]. Their affordance model is based mainly on the SVM structure which can be trained to predict and select an action deterministically, as well as training for each action to predict the affordance label. Three notable characteristics of their models are as follows:

1) use low-level features, which are extracted from stereo vision, range image or Kinect, and these features are used in affordance learning and prediction;
2) choose a set of features, such as position, shape and visibility which are computed mostly from visual perception instead of explicitly linked to the robot's actions, to detect the effects of actions that have been performed on objects [93];
3) a robot learns affordances from continuous manipulative exploration and uses them for symbolic planning.

Their method has some advantages:

1) provides good generalization through the *initial*→*effect* mapping;
2) encodes the effects and objects in the same feature space [78] which results in multistep planning;
3) both flat and hierarchical prediction (Fig. 8);
4) the prediction mechanism could be generalized to actions that involve more than one object, through combining all object features and affordances as the input attributes of the predictors.

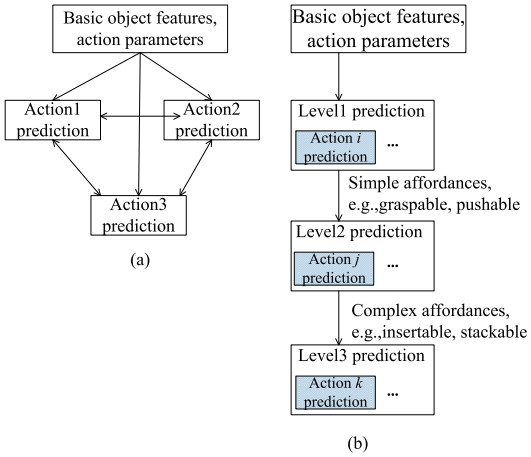

Fig. 8. Prediction system. (a) Flat affordance learning structure, where affordances are predicted based on low-level object features, action parameters, and all other perceived affordances. (b) Simple hierarchical structure where simple affordance predictions can be used to detect complex affordances [25]. Note: this figure is adapted from [25].

They investigated many approaches to implement the learning process and each one has its own characteristics or even shortcomings. For example:

1) when manipulating objects (not stacking), it is based on an assumption that only one object is affected by the action;
2) it has difficulty capturing uncertainty or making predictions in partially known environments, and the prediction is only single-way [36];
3) the plan-tree cannot be used to find the behavior parameters given the current and anticipated next states, although it could be used to predict the next perceptual state given the current perceptual state and behavior parameters.

b) Visual-attribute-affordance: Hermans *et al.* [66] believed that visual and physical attributes support information sharing between objects and affordances. They also used an attribute-based affordance model to describe the mappings from sensor data to visual object categories, and finally to object affordances. Although the advantage of their methodology is validated compared with direct and category-based approaches, the robot would not perform well when it comes to novel objects. Besides predicting opportunities for interaction with an object using only visual cues, their work also learns affordances by observing the effects of exploratory actions [65].

2) Affordance-Based Ontology: Affordance-based ontology could be used to represent an object based on its potential actions, as a result, one object can be substituted by another with similar functionalities [79].

a) Selection and manipulation: A robot should be able to understand the objects and their context in order to complete a mission. Affordance-based ontology (ABO) is proposed to facilitate robots dealing with substantive and nonsubstantive objects. An ABO is to represent objects and their relationships by what it is related to and how it is related in the robot-understandable manner, and a robot can understand the representation of a substantive object and its relation with other

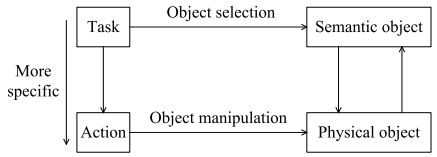

Fig. 9. ABO that works in semantic and physical layers [79].

nonsubstantive objects. When a substantive object is not available, the robot can select a nonsubstantive object based on their functionalities in order to complete the mission. For example, when the user plans to go outside on a rainy day, the robot is able to pick up a protective raincoat or hat if the umbrella is absent. Hidayat *et al.* [79], [80] treated objects and tools in both semantic and physical layers, which are used for selection and manipulation, respectively, as shown in Fig. 9. In their work, physical landmark information is attached to the objects as perceptual source, logical relations between the robot and objects in terms of actions are created, and a query engine is used to obtain the appropriate action.

The work of Koppula *et al.* [52], [67], [68], [84] and Hidayat *et al.* [79], [80] both used semantic affordances. However, the work of Koppula *et al.* [52], [67], [68], [84] could be applied in human–robot interactions in complex environments, while that of Hidayat *et al.* [79], [80] is suitable only in fixed situations.

AfNet is an open affordance network which could provide generic visual knowledge ontology based on affordance recognition [9]. AfRob, an extension to AfNet, offers inference mechanisms to process the semantic information of the visual world and identify semantic affordance features. However, their applicability is limited to man-made objects in a domestic environment.

3) Look-Up Tables: A tabular form is good at expressing the affordance relationships in detail because a table can contain attributes as many as necessary [26].

a) Tool use: The ability to use tools is necessary for an organism to overcome its limitations, and many animals are able to make multiple-step plans in complex environments [94]. For example, a monkey could stack some boxes to pick a banana that is out of its reach. Like infants, an animal also uses its existing exploratory behaviors that are predetermined by its specific species when it confronts a new object [95]. Tool-use research has potential applications in rescue, medical and other fields, where a robot could operate tools to extend its sensor range and affect the scope, and some robots can even operate a complex tool together. However, this research has received relatively less attention from both ecological psychology and developmental robotics.

The key characteristic of tool affordances is that a robot holds one object in its hand to produce a desired effect on another object [96]. The typical work is from [26], [57], and [81], which formulated a behavior-grounded computational model of tool affordances based on the robot's behavioral repertoire, and the robot learns affordances through active trial and error. The tool's affordances are represented in terms of the actions and their effects, as indicated in Table IV [26], where *Os* and *Oe* represent the features observed before and after the action, respectively,

TABLE IV
TABLE TO DESCRIBE TOOL AFFORDANCES [26]

Binding behavior	Binding parameter	Explore Behavior	Explore parameter	O^s	O^e	Times used	Times succeed

"times used" and "times succeed" represent the total times of actions and successful times, respectively. The robot learns the tool's affordances to search tool-behavior pairs that could produce the desired effects, and the learned affordances can be used to solve tool-using tasks by sequencing the behaviors. This representation makes the tool adaptable in different conditions. For example, if a small part of a known tool breaks off, the tool will become a new one and the robot could test the affordances by taking the previously executed actions, then, any inconsistent result would be used to update the tool's representation.

The work of [26] has some drawbacks. First, any object is kept fixed and its affordances are not learned, and color is the only feature to identify a tool. As a result, although efficient in constrained environments, their approach is not generalizable because a developmental robot is sure to encounter novel objects in human environments [26]. Later, Sinapov [97], the members in the research group of Stoytchev [26], proposed a behavior-grounded approach which could individuate a new object successfully. Second, the tool is always assumed to be grasped in a correct and static manner, without considering that the way to grasp a tool will also greatly affect its functionality [98].

b) Table-rule: Wang *et al.* [82] combined RL and genetic algorithms into an affordance framework to implement online learning and use. Active learning is also introduced so that the robot could decide by itself whether it has collected sufficient data to learn the underlying object affordances, but the state space is very small which does not need to be trained for a long-time even if trained offline.

C. Summary

After the classification and analysis of typical works, now we will briefly summarize the existing research.

1) General Analysis: In most of the existing works only a single object is involved, while in a few works multiple-objects [27] or tool-objects [26] are considered. The implementation mechanisms of affordances have continued to develop in recent decades. At the beginning, the earliest differentiation in affordances lay in simple features such as color [26], [99]. Later, some works such as [50] transferred the learned affordances into novel objects and environments, multimodal which means that robot has at least two sensory modalities such as audio and visual were introduced [100], [101], concepts which could combine the learned affordances and past experiences [102] were introduced because they might consist of a sequence of affordances whose execution details could be neglected [103], different parts of an object are described to present affordances to different degrees [104] and a typical example is that the top part of a bottle is more graspable than its bottom part. Recently, functional descriptors were proposed and proved to perform better than the classical appearance-based method [105]. Due to the different representations of affordances, there is no strong connection among current implementations, but the connection within the same research group is stronger because they use similar methodologies to define affordances, i.e., [52] and [69] are both Markov-based.

With the development of hardware and the changes in robotic applications, the focus of affordance research also shifts a little. For example, affordance detection from 3-D image [83], [106] and videos [52] would be more popular because of the availability of inexpensive RGB-D sensors. The detection of human activity, tool, and multiobject affordances will attract more attention, because more and more work that used to be performed by human labor will be the task of the next generation personal robot. The learned affordances could also provide information for other remote robots based on cloud-robot connections [107].

2) Learning Algorithms: Current affordance models are based mainly on SVM, MDP, BN, RL, etc. Most existing implementation mechanisms have four common features:

1) only needs to find the relationships in a given task or learn the limited parameters;
2) only learn what it is preplanned to learn and a single learning algorithm is used;
3) only learns for one subject at one time;
4) the information is gained mainly from visual perception.

There is no algorithm that works the best in existing affordance research because algorithm selection depends on the affordance relationships, and it seems that current algorithms are enough to learn the relationships. However, the state-of-the-art algorithms might perform poorly if more human-like affordance relationships are constructed. Different kinds of learning algorithms, i.e., supervised and unsupervised learning, might have to be used together to finish a complex task because a single algorithm has obvious shortcomings. Take supervised learning for example, the trainer would provide the preferred answer and a single data-set to the robot. Thus, there is no ambiguity about the action choices and environment uncertainty [108].

When it comes to developmental learning, the work of Hart *et al*. [76], [90] seems the closest to full autonomous skill learning for two reasons:

1) the robot is equipped with an intrinsic reward function, and a set of controllers are combined to achieve intrinsically rewarding events;
2) the learned controllers are added to the robot's behavioral repertoire and the learned skills are then used as primitive ingredients in later tasks.

Erez and Smart [109] believed that there are at least six different dimensions that can be explored in shaping, but current RL methods cover only one dimension. The six dimensions are: 1) modifying the reward function; 2) modifying the dynamics; 3) modifying internal parameters; 4) modifying the initial state; 5) modifying the action space; and 6) extending the time horizon [109]. Multidimensional hierarchical RL might be necessary, because the human brain-spine system is a hierarchical structure [3] and infant learning has six dimensions [110].

V. Developmental Perspectives to Consider Affordances

Affordances are believed to be the key building blocks of the most sophisticated natural intelligent systems. As a result, they might also serve as a basis to develop architectures and algorithmic principles within artificial cognitive systems [23], [111]. Originally proposed by a psychologist, the affordance concept is tightly connected to life science, infant learning, developmental learning, and developmental robotics research.

A. Affordances in the Context of Life Science

The affordance phenomenon has been proven in many experiments related to life science. Seeing objects could activate affordances from past experiences and the observed behavior could be simulated inside the vision-action neural systems [112]. Graspable, placeable, and manipulative affordances have a higher chance to activate mirror neurons [113]. Norman believed that affordance detection is the main activity of the dorsal system [114]. The anterior intraparietal area of the parietal cortex uses visual input to extract affordances highlighting the features of the graspable object [115], and object identification information is also sent to the parietal cortex [116]. Each time a graspable object is presented to a monkey, its premotor cortex will be activated regardless whether it will grasp it or not [117]. The observation of tool affordances would activate the left dorsal premotor cortex [94].

B. Affordances in the Context of Infant Learning

Infants use their innate exploratory activities, such as shaking and grasping, to obtain the perceptual data and learn affordances [34]. Piaget [24] argued that infants start to discriminate means-ends relations, which could be called affordances at 4–8 months, and begin to use these affordances for goal-directed tasks at 12 months. There are two key aspects, intrinsic motivation and long-time interaction, that influence affordance learning in infants. Berlyne [119] believed that infant exploration is guided by intrinsic motivations such as curiosity and surprise, and a similar phenomenon is also found in infant chimpanzees [119]. An infant needs a long time, usually several months, to practice a single action in different environments and gain complete knowledge about the action's applicability and expected effects [120]. Law *et al.* [31] designed a detailed infant development timeline and they also illustrated how it could be used in the autonomous development in an iCub humanoid robot.

C. Affordances in the Context of Developmental Learning

A developmental robot is required to develop its intelligence like that of human beings [121] who learn actively, intrinsically motivated and life-long lasting.

1) Active Learning: In robotics the true state is often partially observable and filled with noise. As a result, a robot is necessary to interact with the environment and collect the desired data actively. Active learning is possible to improve the speed and accuracy of learning with fewer training labels, because a robot is allowed to choose the data through which it learns [122], [123]. For instance, a robot actively creates uncertain situations or asks a human teacher to reduce the amount of training data in symbolic concepts learning. Hart [76] proposed active learning of controllable environmental affordances for object manipulation, despite the motor skills are preprogrammed and high-level control programs are given. Ivaldi *et al.* [93] suggested that active learning could cause an object's appearance to change more rapidly, thus the objects' physical properties or affordances will be discovered more quickly. Through active learning the robot could study object categories based on its own sensorimotor experiences and the learned perceptual model is generalizable to new categories [106]. Wang *et al.* [124] proposed a learning approach in which the robot could decide to collect the training data based on its own observation of the environment without human intervention, meanwhile, the prediction error is served as intrinsic reward to update the action exploration policy. However, the learned policies could not be transferred to novel objects and the convergence speed of the learning process should be improved.

2) Intrinsically Motivated Learning: Intrinsic motivation mechanisms might be the core part for developmental robotics and have been studied in psychology, robotics, and RL communities in recent years [125]. Stout *et al.* [126] and Chentanez *et al.* [127] suggested that a good way to obtain general and reusable skills is intrinsic motivation. Stringer *et al.* [128] believed that hierarchical RL is more biologically plausible for learning complex tasks and the intrinsic reward function should be designed to drive a robot to acquire new skills autonomously. Under the affordance framework proposed by Hart and Grupen [90], a robot is intrinsically motivated to control the interactions with its environment, and it could learn new control programs as well as predict their outcomes in a large variety of runtime situations. An organism would gain the highest reward when the level of novelty is in the *familiar* and completely *new* scope [118], and an intrinsically motivated strategy could reduce the exploration time [129]. Prediction errors can be used as the intrinsic reward to optimize the learning process in intrinsic motivation-based affordance work [25], [124].

3) Life-Long Learning: Life-long learning is based on the assumption that many tasks should be learned and their learning mechanisms might be related. The learner should transfer knowledge across different tasks and generalize accurately, especially when there is not enough training data [130]. When the learning process continues, the learner can become *more generalizable* and *more experienced*. Ugur *et al.* [30] proposed a staged developmental skill acquisition framework that could transform the affordance learning capacity from a simple robot to a complex one. Hart and Grupen [50] contributed a longitudinal development, where the behavioral affordances are explicitly connected to the robot's dynamic sensorimotor information and the learning is guided by an intrinsic rewards mechanism.

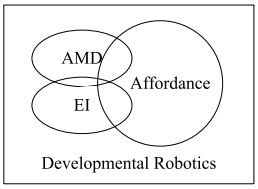

Fig. 10. Relationship between AMD, EI, and affordance. They are all branches of developmental robotics.

D. Affordances in the Context of Developmental Robotics

In developmental robotics a robot uses learning mechanisms and embodiment interactions to learn its environment with little expert knowledge. It can acquire new skills incrementally and continuously as well as change its shape or hardware to adjust to the environment. Autonomous mental development (AMD) [131], embodied intelligence (EI) [132], and affordance are three research methodologies in developmental robotics, as shown in Fig. 10. AMD is related to perception capability that predicts affordances, while EI represents physical capability that executes affordances [70]. If a robot is able to construct complex affordances, both its mental and EI are developed. If a robot finds that it needs a gripper to generate the graspable affordance, it can ask the user to add this hardware and this adaptable capability can be regarded as EI.

Intelligence is closely and inseparably linked with behavioral interactions [39]. AMD and EI also have intersections for the following reasons:

1) embodiment plays a central role in the cognitive processes of any learning agent (robot) [39];
2) the embodiment and autonomous mental system should be combined within a relational worldview [133];
3) the developmental process emerges from the extended brain-body-behavior networks, which means that the brain and body are mutually affected through behaviors [134];
4) embodied actions could generate multimodal information which is necessary for mental development [134].

VI. Affordances' Advantages in Developmental Robotics

It is difficult to answer this question completely because affordance research is still active. Affordance methodology is used as a source of inspiration in developmental robotics [7] and has been used to improve the efficiency of object recognition, robot control, and planning. As a result, object affordances should be a prerequisite for planning and prediction in a developmental robot.

A. Planning

Classical planning approaches have difficulty in preventing a robot from considering actions that are obviously irrelevant to the current problem [135]. Under the affordance concept we first consider affordance relationships as state-action pairs that could change the environment and then make plans based on these pairs [136]. Maron *et al.* [62] formalized affordances as knowledge added to an MDP that described the actions in state- and reward- general way. This method makes a robot act near optimally because the number of state-action pairs that the robot needs to evaluate are greatly reduced [62]. Awaad *et al.* [35] introduced affordances into some reasoning phases to make plans, and the robot could easily adapt the plans in a dynamic environment. Lörken and Hertzberg [137] incorporated affordances into an ordinary planning method and proved that the resulting planning approach is more flexible and robust. Piyathilaka and Kodagoda [138] converted a dense 3-D point cloud into an *affordance-map* that contains much semantic information and consists of virtual human models. The path planner could then produce a better human robot interaction.

B. Control

To precisely execute an action the affordances should be described in detail. For instance, the methods used to perform an action for *graspable* affordance upon a round cereal bowl and a cup with handle might differ. Hermans *et al.* [65] decomposed an affordance-based behavior into three components: a control policy, a primitive behavior and a perceptual proxy. A control policy describes how to generate a desired change in object state through changing the robot's state, a primitive behavior specifies how the controller output is translated to the robot's movements. A perceptual proxy defines the object representation that is the input for the controller during execution. This factored representation could enable the robot to explore different combinations of movements and learn which behavior works best for a given object. This approach could also provide a way to transform the continuous actuator control space into a discrete selection process in a specific task [87].

C. Recognition

Most works consider the recognition of objects and their affordances as a classification problem and train separate classifiers for them. However, affordance recognition based on objects and environments has gained more attention in recent years. Zhu *et al.* [139] proposed a knowledge-based representation to reason about them during human-object interactions, and that approach could perform better than classical classification methodologies. Castellini *et al.* [141] introduced a graspable affordance into an object model to improve its recognition, where the visual features of an object are mapped into the kinematic features of a hand when the affordance is performed. Koppula *et al.* [52] believed that object affordance is more useful than its category when it comes to activity recognition. They also used an affordance model to increase activity recognition from RGB-D videos [52]. Kjellström *et al.* [141] provided a functional understanding of the environment by predicting affordance-based object attributes. Jiang [142] proved that human and object affordances can better model the environment. To use and manipulate objects based on their affordances is extremely useful in uncertain environments such as search-and-rescue [143].

D. Transferability

We human beings are good at transferring knowledge and can easily capture the essential features of the environment.

The learned affordances could be easily combined with symbol and language grounding to make the robot better transfer to other objects that have a similar shape, and the transferable knowledge could be conveniently extended to different robotic applications [144]. Maron *et al.* [62] designed a fully learning process under which a robot could autonomously learn useful affordances that might be used across different task types, reward functions, and state space.

E. Programming style

The affordance concept is separate from the mechanism that implements it. The researcher could use the most recent algorithms or propose a new one. Affordances lead to a conceptual change in the way to consider robotic programming as well as develop new knowledge, because the robot would actively explore the environment through its behaviors. There is some difference between knowing how to detect an object and how to use it. The former could be solved by computer vision, while the latter is about robotic manipulation. However, an affordance representation could combine these two steps into one. For instance, if a robot could detect a container then it already knows how to use it [106].

VII. Discussion

As discussed in Section IV, existing work mainly uses the affordance concept in the form of a triplet (*object*, *action*, *effect*) deterministically or probabilistically. A robot under this framework develops in a predefined manner which is not fully developmental like humans. To make affordance learning more human-like, in this section we will discuss some questions about: 1) the use of affordances originated from Gibson's definition and 2) the perception of affordances as well as the effect of such perception.

A. Can Affordance be Decomposed?

In fact, the affordance concept proposed by Gibson [4] is abstract and can be called *macro-affordance*. In cognitive psychology, Ellis and Tucker [145] described the effects of seen objects on the components of an action as *micro-affordances*, which are regarded as the intentional states of the viewer's nervous system. Thill *et al.* [146] stated that micro-affordances are concerned with specific action components, instead of the whole action. For instance, a moveable affordance might activate two different components such as direction (e.g., left or right) and location (e.g., with the hand put on the top or bottom part). As a result, micro-affordances are independent of the agent's goal.

Borghi and Riggio [147] believed that *stable* and *variable* affordances can be used to further discriminate affordance representation and selection. A stable affordance represents object features that remain constant in different experiences (e.g., size and shape), while a variable affordance represents the object features that would change in different experiences (e.g., orientation) and such information should be gained online. The distinction between stable and variable affordances is related to the distinction between the ventral and dorsal neural pathways [148].

However, these two kinds of decompositions are not utilized in robotics community.

B. Is Affordance Extendable?

The original concept of affordance depends on the actor's goals, values, plans, beliefs, and past experiences [68]. In a complex task or social situation the environment should be taken into account. For example, a ball is first kickable, but the kickable affordance might not exist if the ball has been kicked into a corner. Currently, most works do not consider the context information which allows predicting the effects in different situations properly. Kammer *et al.* [149] proposed the concept of situated affordance but they did not carry out work in robotic platforms. The difference between the *context* mentioned here and *stable* mentioned in Section VII-A is that *context* represents the environment excluding the current object and robot while *stable* relates to part of the features of current object.

C. Do We Need to Investigate the Difference of Affordance Perception in Organisms and in Robotics?

Although a number of approaches could develop affordances, each of them has its own limitations and assumptions that are not in accordance with the findings in neuroscience. An organism perceives affordances using its brain's neurons, and it also has instincts which are very important for later life [31]. However, a robot perceives affordances based on the electronic equipment, and it is programmed to have some basic capabilities. To create an affordance model in a robot similar to the ones in an organism, as least three problems should be solved.

1) How to fill the gap between robotic models and psychology models?
2) How to simulate the neuron computing in terms of robotic hardware?
3) In which aspects (e.g., visual, tactile, and audition) and in which degree to endow a robot with instinctive capabilities?

Solving these problems might be useful to provide mechanisms and principles that guide affordance exploration and learning strategies in an organism-level developmental manner.

D. What is the Effect of Affordance Perception in Robotics?

In Section V, we discussed that the affordance phenomenon has been verified in the life science community. For instance, seeing the handle of a cup, our motor schema that is related to grasping such category of handles will be activated subconsciously even before we recognize the cup's name [13]. However, what is the output of affordance perception in robotics? The potential action(s), or the effects of these action(s), or both? Moreover, is there a unified model to represent affordance perception? For example, the effects of moveable affordances might be salient changes in the environment, but how to represent the effect of the sitability of a sofa or chair?

VIII. Future Trends

Some questions that are related to affordance use and exploration were discussed in the previous section. Regardless whether these questions have been solved, the ultimate goal is to apply affordances in developmental robotics and make a robot develop as much as possible like an organism. Based on the analysis and discussions presented in this survey, now we will highlight several promising topics in affordance research in developmental robotics that have received limited attention, ranging from continuous representation to the benchmarks.

A. From Discrete to Continuous Affordances

Current work focuses on affordances with binary *true/false* values, without considering the transitional process. A robot designed for outdoor [60] or indoor [33] navigation could predict the traversable affordance from image features, but there is no intermediate or gradual transition in the traversability. Sometimes the possibility of an affordance is a value that can change gradually, for instance, the degree of walkability of a sand surface is a value that can change gradually. As a result, this *continuous* value of an affordance should be determined automatically by the robotic system. It is a classification problem in the prediction of discrete value affordance, and it is a regression problem in the prediction of continuous value affordance [75].

B. Affordances and the Transferability for Multirobot Tasks

A multirobot system is necessary in tasks such as rescue and exploration, where the robots could cooperate both passively and actively [150]. Yi *et al.* [151] described a robot in terms of its capabilities, thus affordances could be created based on the shared information and existing affordances. However, that principle has only been testified in a specific task in a simple environment. Several problems should be solved in a multirobot environment. For instance, if one robot has learned some affordances in current environment and later another robot with different hardware equipments is brought to work here, then at least three questions should be answered [152].

1) Which part of the learned affordances should be transferred?
2) How to transfer the learned affordances, i.e., which algorithm should be used?
3) In which conditions the affordances should be transferred and in which conditions they should not be transferred.

C. Integrating Deterministic and Probabilistic Methods

A robot needs to plan actions based on object affordances in a dynamic environment. Deterministic methods such as [36] could work well in planning based on logic rules, but it performs poorly in uncertainty or partially known environments. Probabilistic methods such as [23] could solve specific tasks in a robust and automatic manner, but it is not good at multistep planning. As a result, one possible improvement might be to integrate the preciseness of deterministic and robustness of probabilistic methods.

D. Generalizable Perceptual Representation Methodology in Different Environments and Tasks

Existing methodologies use hand selected features to perceive affordances. For example, Montesano *et al.* [23] and Ugur *et al.* [78] used different and object dependent features in manipulation tasks. However, to perceive traversable affordances the robot does not need to detect objects [33]. As a result, it is necessary to find a generalizable or even unified perceptual representation mechanism that can automatically discover the most relevant features in different affordances such as graspable, sittable, and traversable. Deep learning [153] as a bottom-up method might be used to extract features automatically in complex scenarios.

E. Online Affordance Learning

The learning and using process should be integrated or the robot will have difficulty when the environment changes [82]. For example, when the robot has learned that a cube is moveable, it will constantly try to push an un-movable object which looks like the cube. Current methods are mainly about how to perceive affordances instead of using them to promote its performance in tasks, and this online learning should be similar to infant learning.

F. Intertwined Multimodal Affordance Learning

Infants simultaneously develop many capabilities such as the visual, tactile, and audio. When growing up, they use different strategies such as imitation, exploration, and observation to learn affordances. A developmental robot should incrementally build its knowledge based on low-level multimodal sensory information [93], and the modalities should be linked in an intertwined manner [154].

A new affordance learning architecture is proposed in Fig. 11. *Basic instinct* is predefined by the researchers, *human guidance* is important for innovative learning because it acts like the role of parents for infant learning, *multimodal fusion* is used to integrate all of the information from each modality in an intertwined way. *Affordances*, which consist of different modalities at the first level, represent the whole intelligence of a robot. A single modality consists of the hierarchical affordance relationships described from the second to the last level. The levels from top to down generally represent affordances from simple to complex. For example, the vision-based affordances of graspable, insertable and two-robot-graspable become much more complex. In different tasks the working modalities might differ. The main differences between existing methods and this method include three aspects:

1) this method could integrate all of the modalities;
2) in each modality affordances are described in a hierarchical form;
3) *mental simulation* is used to execute the actions in simulator software before actually executing them.

This learning architecture also integrates staged developmental learning [30], because any affordance learned in any stage could be used to develop later affordances.

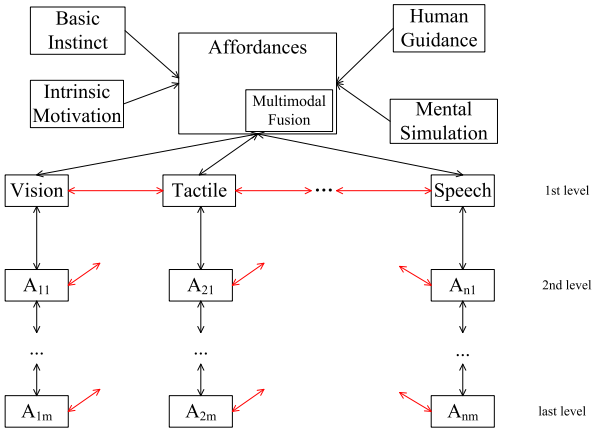

Fig. 11. Multimodal affordance learning in real time. For A_{ij}, A is affordance for short, i is the modality number, and j is the hierarchical number in a given modality. Human guidance could provide labels. The arrows in red mean that they can correlate with any element in any level. The bi-directional arrows between the modalities at the first level and affordances represent that they will develop mutually.

G. Basic Principles for Affordance Research

In the community of developmental robotics, Stoytchev [155] formulated five basic principles for autonomous tool-use and Cangelosi *et al.* [86] provided milestones which acted as a research roadmap. As the influence of affordance research becomes increasingly profound, it needs some benchmarks to unify this research and thus the researchers would be able to compare the results from different methods. For instance, for a given task which learning algorithm provides clear advantages compared to others? which indexes should be used to examine the efficiency and effect of the learning algorithm? The basic principles can be used to guide affordance research although they are not laws, and they will be modified or rejected when they are found to be inadequate.

Besides the above topics, new achievements in related fields such as material science and AI should also be applied in affordance research. For instance, material science would make a robot easier to produce and more adaptable in a changing environment, algorithms inspired from AlphaGo [156] which combines deep neural network and tree search could make affordance learning more effective.

Despite attempting to cover the full range of future topics, we might have missed many important ones. Even though there are still a number of issues that remain to be executed, affordance research is a promising approach in developmental robotics. The challenges of affordance research could also provide inspiration, validation and impact for human development research.

IX. Conclusion

Affordance research provides a biology-inspired way to perceive and interact with the environment. In this survey, the first three sections provided a general overview of affordances in developmental robotics. These sections also answered many questions such as: the role and developmental path of affordances, the relationship between affordance learning and infant learning, the methods used to perceive affordances, etc. Section IV classified the existing work according to the representation form, identified each typical work clearly about what is achieved and what is missed in general, and described the history of affordance research ranging from the initial simple approaches to the recent descriptors beyond the appearance. The algorithms developed to implement affordance relationships were compared. Section V indicates that this concept has a solid foundation in biology as well as good potential to promote the cognitive level of robots. Section VI focuses on the question "are affordances really useful in developmental robotics," and it gives some examples to illustrate its usefulness especially in planning, control and recognition. Because existing affordance knowledge develops in a predefined manner instead of fully autonomously, Section VII discusses whether the classical use of Gibson's affordance or even the concept itself could be improved so as to bring some innovative ideas to apply affordances into developmental robotics. Section VIII points out some important and practical research issues that need to be addressed when applying affordances in developmental robotics.

In conclusion, a developmental robot's knowledge should be grounded in multimodal affordances it discovers in its environment. Behavioral affordances should be explicitly grounded in the robot's dynamic sensory-motor interactions with its environment, and the robot must be equipped with an intrinsic motivation system to permanently seek out affordances and the conditions in which they occur.

Acknowledgment

The authors would like to thank the Associate Editor and the anonymous reviewers for their invaluable comments and suggestions.

References

[1] V. Tikhanoff, A. Cangelosi, and G. Metta, "Integration of speech and action in humanoid robots: iCub simulation experiments," *IEEE Trans. Auton. Mental Develop.*, vol. 3, no. 1, pp. 17–29, Mar. 2011.

[2] M. Lungarella, G. Metta, R. Pfeifer, and G. Sandini, "Developmental robotics: A survey," *Connect. Sci.*, vol. 15, no. 4, pp. 151–190, 2003.

[3] M. Asada *et al.*, "Cognitive developmental robotics: A survey," *IEEE Trans. Auton. Mental Develop.*, vol. 1, no. 1, pp. 12–34, May 2009.

[4] J. J. Gibson, "The theory of affordances," in *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, R. Shaw and J. Bransford, Eds. Englewood Cliffs, NJ, USA: Lawrence Erlbaum, 1977, pp. 67–82.

[5] S. W. Hsiao, C. F. Hsu, and Y. T. Lee, "An online affordance evaluation model for product design," *Design Stud.*, vol. 22, no. 2, pp. 126–159, 2012.

[6] C. Schneider and J. Valacich, "Enhancing the motivational affordance of human–computer interfaces in a cross-cultural setting," in *Information Technology and Innovation Trends in Organizations*. Heidelberg, Germany: Springer, 2011, pp. 271–278.

[7] E. Rome, J. Hertzberg, and G. Dorffner, "Towards affordance-based robot control," in *Proc. Int. Seminar*, Wadern, Germany, Jun. 2006.

[8] MACS, (2015). *A Research Project for Exploring and Exploiting the Concept of Affordances for Robot Control*. [Online]. Available: http://www.macs-eu.org/

[9] (2014). *AfNet 2.0: The Affordance Network*. [Online]. Available: http://affordances.info/workshops

[10] (2015). *Workshops and Tutorials*. [Online]. Available: http://www.iros2015.org/index.php/program/workshops-and-tutorials

[11] (2016). *Call for Papers Special Issue on Computational Models of Affordances for Cognitive Robots*. [Online]. Available: http://cis.ieee.org/ieee-transactions-on-cognitive-and-developmental-systems.html

[12] T. E. Horton, A. Chakraborty, and R. S. Amant, "Affordances for robots: A brief survey," *AVANT. Pismo Awangardy Filozoficzno Naukowej*, vol. 3, no. 2, pp. 70–84, 2012.

[13] L. Jamone *et al.*, "Affordances in psychology, neuroscience and robotics: A survey, " *IEEE Trans. Cogn. Develop. Syst.*, to be published, doi: 10.1109/TCDS.2016.2594134.

[14] U. Klank, L. Mosenlechner, A. Maldonado, and M. Beetz, "Robots that validate learned perceptual models," in *Proc. IEEE Int. Conf. Robot. Autom.*, St. Paul, MN, USA, 2012, pp. 4456–4462.

[15] A. Gupta and L. S. Davis, "Objects in action: An approach for combining action understanding and object perception," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, USA, 2007, pp. 1–8.

[16] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth, "Describing objects by their attributes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, 2009, pp. 1778–1785.

[17] J. Liu, B. Kuipers, and S. Savarese, "Recognizing human actions by attributes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Colorado Springs, CO, USA, 2011, pp. 3337–3344.

[18] O. Russakovsky and F. F. Li, "Attribute learning in large-scale datasets," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 1–14.

[19] A. Szokolszky, "An interview with Eleanor Gibson," *Ecol. Psychol.*, vol. 15, no. 4, pp. 271–281, 2003.

[20] K. Welke *et al.*, "Grounded spatial symbols for task planning based on experience," in *Proc. IEEE Int. Conf. Humanoid Robots*, Atlanta, GA, USA, 2013, pp. 484–491.

[21] N. Krüger *et al.*, "Object–action complexes: Grounded abstractions of sensory–motor processes," *Robot. Auton. Syst.*, vol. 59, no. 10, pp. 740–757, 2011.

[22] E. Ugur, E. Sahin, and E. Oztop, "Unsupervised learning of object affordances for planning in a mobile manipulation platform," in *Proc. IEEE Int. Conf. Robot. Autom.*, Shanghai, China, 2011, pp. 4312–4317.

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

[24] J. Piaget, *The Origins of Intelligence in Children.* New York, NY, USA: Int. Univ. Press, 1952.

[25] E. Ugur and J. Piater, "Emergent structuring of interdependent affordance learning tasks," in *Proc. IEEE Int. Conf. Develop. Learn. Epigenet. Robot.*, Genoa, Italy, 2014, pp. 489–494.

[26] A. Stoytchev, "Behavior-grounded representation of tool affordances," in *Proc. IEEE Int. Conf. Robot. Autom.*, Barcelona, Spain, 2005, pp. 3060–3065.

[27] A. Moldovan, P. Moreno, M. V. Otterlo, J. Santos-Victor, and L. D. Raedt, "Learning relational affordance models for robots in multi-object manipulation tasks," in *Proc. IEEE Int. Conf. Robot. Autom.*, St. Paul, MN, USA, 2012, pp. 4373–4378.

[28] E. Ugur and J. Piater, "Bottom-up learning of object categories, action effects and logical rules: From continuous manipulative exploration to symbolic planning," in *Proc. IEEE Int. Conf. Robot. Autom.*, Seattle, WA, USA, 2015, pp. 2627–2633.

[29] K. F. Uyanik *et al.*, "Learning social affordances and using them for planning," in *Proc. 35th Annu. Meeting Cogn. Sci. Soc.*, 2013, pp. 3604–3609.

[30] E. Ugur, Y. Nagai, E. Sahin, and E. Oztop, "Staged development of robot skills: Behavior formation, affordance learning and imitation," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 2, pp. 119–139, Jun. 2015.

[31] J. Law, M. Lee, M. Hülse, and A. Tomassetti, "The infant development timeline and its application to robot shaping," *Adapt. Behav.*, vol. 19, no. 5, pp. 335–358, 2011.

[32] L. Paletta, G. Fritz, F. Kintzler, J. Irran, and G. Dorffner, "Learning to perceive affordances in a framework of developmental embodied cognition," in *Proc. IEEE Int. Conf. Develop. Learn.*, London, U.K., 2007, pp. 110–115.

[33] E. Uğur and E. Şahin, "Traversability: A case study for learning and perceiving affordances in robots," *Adapt. Behav.*, vol. 18, nos. 3–4, pp. 258–284, 2010.

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

[35] I. Awaad, G. K. Kraetzschmar, and J. Hertzberg, "Affordance-based reasoning in robot task planning," in *Proc. Plan. Robot. Workshop Int. Conf. Autom. Plan. Scheduling*, Rome, Italy, 2013.

[36] E. Ugur, "A developmental framework for learning affordances," Ph.D. dissertation, Dept. Comput. Eng., Middle East Tech. Univ., Ankara, Turkey, 2010.

[37] D. Pecher, "The role of motor affordances in visual working memory," *Baltic Int. Yearbook Cogn. Logic Commun.*, vol. 9, no. 1, pp. 1–18, 2014.

[38] K. M. Varadarajan and M. Vincze, "AfRob: The affordance network ontology for robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Vilamoura, Portugal, 2012, pp. 1343–1350.

[39] M. Wilson, "Six views of embodied cognition," *Psychonom. Bull. Rev.*, vol. 9, no. 4, pp. 625–636, 2002.

[40] A. Aldoma, F. Tombari, and M. Vincze, "Supervised learning of hidden and non-hidden 0-order affordances and detection in real scenes," in *Proc. IEEE Int. Conf. Robot. Autom.*, St. Paul, MN, USA, 2012, pp. 1732–1739.

[41] M. T. Turvey, "Affordances and prospective control: An outline of the ontology," *Ecol. Psychol.*, vol. 4, no. 3, pp. 173–187, 1992.

[42] T. A. Stoffregen, "Affordances as properties of the animal-environment system," *Ecol. Psychol.*, vol. 15, no. 2, pp. 115–134, 2003.

[43] A. Chemero, "An outline of a theory of affordances," *Ecol. Psychol.*, vol. 15, no. 2, pp. 181–195, 2003.

[44] M. Steedman, "Plans, affordances, and combinatory grammar," *Linguist. Philosophy*, vol. 25, nos. 5–6, pp. 723–753, 2002.

[45] E. Şahin, M. Çakmak, M. R. Doğar, E. Ugur, and G. Ucoluk, "To afford or not to afford: A new formalization of affordances toward affordance-based robot control," *Adapt. Behav.*, vol. 15, no. 4, pp. 447–472, 2007.

[46] F. Worgotter, A. Agostini, N. Krüger, N. Shylo, and B. Porr, "Cognitive agents—A procedural perspective relying on the predictability of object-action-complexes (OACs)," *Robot. Auton. Syst.*, vol. 57, no. 4, pp. 420–432, 2009.

[47] H. Q. Min *et al.*, "Affordance learning based on subtask's optimal strategy," *Int. J. Adv. Robot. Syst.*, vol. 12, no. 8, p. 111, 2015.

[48] H. Q. Min, C. A. Yi, R. H. Luo, and J. H. Zhu, "Goal-directed affordance prediction at the subtask level," *Ind. Robot Int. J.*, vol. 43, no. 1, pp. 48–57, 2016.

[49] S. Hart, "An intrinsic reward for affordance exploration," in *Proc. IEEE Int. Conf. Develop. Learn.*, Shanghai, China, 2009, pp. 1–6.

[50] S. Hart and R. Grupen, "Learning generalizable control programs," *IEEE Trans. Auton. Mental Develop.*, vol. 3, no. 3, pp. 216–231, Sep. 2011.

[51] S. Ou, "A behavioral approach to human–robot communication," Ph.D. dissertation, Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, 2010.

[52] H. S. Koppula, R. Gupta, and A. Saxena, "Learning human activities and object affordances from RGB-D videos," *Int. J. Robot. Res.*, vol. 32, no. 8, pp. 951–970, 2013.

[53] A. Saxena, J. Driemeyer, and A. Ng, "Robotic grasping of novel objects using vision," *Int. J. Robot. Res.*, vol. 27, no. 2, pp. 157–173, 2008.

[54] R. Detry, D. Kraft, A. G. Buch, N. Kruger, and J. Piater, "Refining grasp affordance models by experience," in *Proc. IEEE Int. Conf. Robot. Autom.*, Anchorage, AK, USA, 2010, pp. 2287–2293.

[55] R. Detry *et al.*, "Learning grasp affordance densities," *J. Behav. Robot.*, vol. 2, no. 1, pp. 1–17, 2011.

[56] G. H. Liang and M. Yan, "Distributed reinforcement learning for coordinate multi-robot foraging," *Intell. Robot. Syst.*, vol. 60, nos. 3–4, pp. 531–551, 2010.

[57] S. Griffith, V. Sukhoy, T. Wegter, and A. Stoytchev, "Object categorization in the sink: Learning behavior–grounded object categories with water," in *Proc. Workshop Semantic Percept. Mapping Explor. Int. Conf. Robot. Autom.*, 2012.

[58] V. Tikhanoff, U. Pattacini, L. Natale, and G. Metta, "Exploring affordances and tool use on the iCub," in *Proc. IEEE RAS Int. Conf. Humanoid Robots*, Atlanta, GA, USA, 2013, pp. 130–137.

[59] E. Ugur, M. R. Dogar, M. Cakmak, and E. Sahin, "The learning and use of traversability affordance using range images on a mobile robot," in *Proc. IEEE Int. Conf. Robot. Autom.*, Rome, Italy, 2007, pp. 1721–1726.

[60] J. Sun, J. L. Moore, A. Bobick, and J. M. Rehg, "Learning visual object categories for robot affordance prediction," *Int. J. Robot. Res.*, vol. 29, nos. 2–3, pp. 174–197, 2010.

[61] T. Ziemke, "On the role of robot simulations in embodied cognitive science," *AISB J.*, vol. 1, no. 4, pp. 389–399, 2003.

[62] G. B. Maron, D. Abel, J. MacGlashan, and S. Tellex, "Affordances as transferable knowledge for planning agents," in *Proc. Assoc. Adv. Artif. Intell. Fall Symp. Series*, 2014.

[63] S. S. Heikkilä, A. Halme, and A. Schiele, "Affordance-based indirect task communication for astronaut-robot cooperation," *J. Field Robot.*, vol. 29, no. 4, pp. 576–600, 2012.

[64] L. Paletta and G. Fritz, "Reinforcement learning of predictive features in affordance perception," in *Towards Affordance-Based Robot Control*. Heidelberg, Germany: Springer, 2008, pp. 77–90.

[65] T. Hermans, J. M. Rehg, and A. F. Bobick, "Decoupling behavior, perception, and control for autonomous learning of affordances," in *Proc. IEEE Int. Conf. Robot. Autom.*, Karlsruhe, Germany, 2013, pp. 4989–4996.

[66] T. Hermans, J. M. Rehg, and A. Bobick, "Affordance prediction via learned object attributes," in *Proc. IEEE Int. Conf. Robot. Autom. Workshop Semantic Percept. Mapping Exploration*, 2011.

[67] H. S. Koppula and A. Saxena, "Anticipating human activities using object affordances for reactive robotic response," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 14–29, Jan. 2016.

[68] H. S. Koppula and A. Saxena, "Physically grounded spatio-temporal object affordances," in *Proc. Eur. Conf. Comput. Vis.*, Zürich, Switzerland, 2014, pp. 831–847.

[69] Y. Jiang, H. Koppula, and A. Saxena, "Hallucinated humans as the hidden context for labeling 3D scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, 2013, pp. 2993–3000.

[70] S. S. Heikkilä, "Affordance-based task communication methods for astronaut-robot cooperation," Ph.D. dissertation, Dept. Autom. Syst. Technol., Aalto Univ., Espoo, Finland, 2011.

[71] S. Ou and R. Grupen, "From manipulation to communicative gesture," in *Proc. ACM/IEEE Int. Conf. Human Robot Interact.*, Osaka, Japan, 2010, pp. 325–332.

[72] M. Lopes, F. S. Melo, and L. Montesano, "Affordance-based imitation learning in robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, San Diego, CA, USA, 2007, pp. 1015–1021.

[73] B. Moldovan, M. V. Otterlo, and P. M. Lopez, "Statistical relational learning of object affordances for robotic manipulation," in *Latest Advances in Inductive Logic Programming*. London, U.K.: Imperial College Press, 2012.

[74] B. Moldovan, P. Moreno, and M. V. Otterlo, "On the use of probabilistic relational affordance models for sequential manipulation tasks in robotics," in *Proc. IEEE Int. Conf. Robot. Autom.*, Karlsruhe, Germany, 2013, pp. 1290–1295.

[75] J. Sun, "Object categorization for affordance prediction," Ph.D. dissertation, College Comput., Georgia Inst. Technol., Atlanta, GA, USA, 2008.

[76] S. Hart, "The development of hierarchical knowledge in robot systems," Ph.D. dissertation, Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, 2009.

[77] S. Hart, S. Sen, and R. Grupen, "Intrinsically motivated hierarchical manipulation," in *Proc. IEEE Int. Conf. Robot. Autom.*, Pasadena, CA, USA, 2008, pp. 3814–3819.

[78] E. Ugur, E. Oztop, and E. Sahin, "Goal emulation and planning in perceptual space using learned affordances," *Robot. Auton. Syst.*, vol. 59, nos. 7–8, pp. 580–595, 2011.

[79] S. S. Hidayat, B. K. Kim, and K. Ohba, "Learning affordance for semantic robots using ontology approach," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nice, France, 2008, pp. 2630–2636.

[80] S. S. Hidayat, B. K. Kim, and K. Ohba, "An approach for robots to deal with objects," *Int. J. Comput. Sci. Inf. Technol.*, vol. 4. no. 1, pp. 19–32, 2012.

[81] A. Stoytchev, "Robot tool behavior: A developmental approach to autonomous tool use," Ph. D. dissertation, Dept. Comput. Sci., Georgia Inst. Technol., Atlanta, GA, USA, 2007.

[82] C. Wang, K. V. Hindriks, and R. Babuska, "Robot learning and use of affordances in goal-directed tasks," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Tokyo, Japan, 2013, pp. 2288–2294.

[83] D. I. Kim and G. S. Sukhatme, "Semantic labeling of 3D point clouds with object affordance for robot manipulation," in *Proc. IEEE Int. Conf. Robot. Autom.*, Hong Kong, 2014, pp. 5578–5584.

[84] H. S. Koppula, "Understanding people from RGBD data for assistive robots," Ph.D. dissertation, Dept. Comput. Sci., Cornell Univ., Ithaca, NY, USA, 2016.

[85] G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi, "Premotor cortex and the recognition of motor actions," *Cogn. Brain Res.*, vol. 3, no. 2, pp. 131–141, 1996.

[86] A. Cangelosi *et al.*, "Integration of action and language knowledge: A roadmap for developmental robotics," *IEEE Trans. Auton. Mental Develop.*, vol. 2, no. 3, pp. 167–195, Sep. 2010.

[87] T. Hermans, "Representing and learning affordance-based behaviors," Ph.D. dissertation, School Interact. Comput., Georgia Inst. Technol., Atlanta, GA, USA, 2014.

[88] V. Krunic, G. Salvi, A. Bernardino, L. Montesano, and J. Santos-Victor, "Affordance based word-to-meaning association," in *Proc. IEEE Int. Conf. Robot. Autom.*, Kobe, Japan, 2009, pp. 4138–4143.

[89] L. Paletta, G. Fritz, F. Kintzler, J. Irran, and G. Dorffner, "Perception and developmental learning of affordances in autonomous robots," in *Advances in Artificial Intelligence*. Heidelberg, Germany: Springer, 2007, pp. 235–250.

[90] S. Hart and R. Grupen, "Intrinsically motivated affordance discovery and modeling," in *Intrinsically Motivated Learning in Natural and Artificial Systems*. Heidelberg, Germany: Springer, 2013, pp. 279–300.

[91] G. D. Konidaris, "Autonomous robot skill acquisition," Ph.D. dissertation, Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, 2011.

[92] E. Ugur, E. Oztop, and E. Sahin, "Going beyond the perception of affordances: Learning how to actualize them through behavioral parameters," in *Proc. IEEE Int. Conf. Robot. Autom.*, Shanghai, China, 2011, pp. 4768–4773.

[93] S. Ivaldi *et al.*, "Object learning through active exploration," *IEEE Trans. Auton. Mental Develop.*, vol. 6, no. 1, pp. 56–72, Mar. 2014.

[94] B. B. Beck, *Animal Tool Behavior: The Use and Manufacture of Tools by Animals*. New York, NY, USA: Garland STPM Press, 1980.

[95] T. G. Power, *Play and Exploration in Children and Animals*. Hoboken, NJ, USA: Taylor and Francis, 1999.

[96] A. Antunes *et al.*, "Robotic tool use and problem solving based on probabilistic planning and learned affordances," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. Workshop Learn. Object Affordances Fundamental Step to Allow Prediction Plan. Tool use*, 2015.

[97] J. Sinapov, "Behavior-grounded multi-sensory object perception and exploration by a humanoid robot," Ph.D. dissertation, Dept. Comput. Sci. Human Comput. Interact. Program, Iowa State Univ., Ames, IA, USA, 2013.

[98] T. Mar, V. Tikhanoff, G. Metta, and L. Natale, "2D and 3D functional features for tool affordance learning and generalization on humanoid robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. Workshop Learn. Object Affordances Fundamental Step Allow Prediction Plan. Tool Use*, 2015.

[99] P. Fitzpatrick, G. Metta, L. Natale, A. Rao, and G. Sandini, "Learning about objects through action—Initial steps towards artificial cognition," in *Proc. IEEE Int. Conf. Robot. Autom.*, vol. 3. Taipei, Taiwan, 2003, pp. 3140–3145.

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

[101] E. Hyttinen, D. Kragic, and R. Detry, "Learning the tactile signatures of prototypical object parts for robust part-based grasping of novel objects," in *Proc. IEEE Int. Conf. Robot. Autom.*, Seattle, WA, USA, 2015, pp. 4927–4932.

[102] A. M. Borghi, "Object concepts and action," in *Grounding of Cognition: The Role of Perception and Action in Memory, Language, and Thinking*. Cambridge, U.K.: Cambridge Univ. Press, 2004, pp. 8–34.

[103] S. Kalkan, N. Dağ, O. Yürüten, A. M. Borghi, and E. Şahin, "Verb concepts from affordances," *Interact. Stud.*, vol. 15, no. 1, pp. 1–37, 2014.

[104] A. K. Pandey and R. Alami, "Mightability maps: A perceptual level decisional framework for co-operative and competitive human–robot interaction," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Taipei, Taiwan, 2010, pp. 5842–5848.

[105] A. Pieropan, C. H. Ek, and H. Kjellström, "Functional descriptors for object affordances," *IEEE/RSJ Int. Conf. Intell. Robots Systems, Workshop Learn. Object Affordances Fundam. Step Allow Prediction Plan. Tool Use*, 2015.

[106] S. Griffith, J. Sinapov, M. Miller, and A. Stoytchev, "Toward interactive learning of object categories by a robot: A case study with container and non-container objects," in *Proc. IEEE 8th Int. Conf. Develop. Learn.*, Shanghai, China, 2009, pp. 1–6.

[107] M. Tenorth, A. C. Perzylo, R. Lafrenz, and M. Beetz, "The RoboEarth language: Representing and exchanging knowledge about actions, objects, and environments," in *Proc. IEEE Int. Conf. Robot. Autom.*, St. Paul, MN, USA, 2012, pp. 1284–1289.

[108] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1238–1274, 2013.

[109] T. Erez and W. D. Smart, "What does shaping mean for computational reinforcement learning?" in *Proc. IEEE Int. Conf. Develop. Learn.*, Monterey, CA, USA, 2008, pp. 215–219.

[110] L. Smith and M. Gasser, "The development of embodied cognition: Six lessons from babies," *Artif. Life*, vol. 11, nos. 1–2, pp. 13–29, 2005.

[111] E. Oztop, H. Imamizu, G. Cheng, and M. Kawato, "A computational model of anterior intraparietal (AIP) neurons," *Neurocomputing*, vol. 69, nos. 10–12, pp. 1354–1361, 2006.

[112] S. P. Tipper, M. A. Paul, and A. E. Hayes, "Vision-for-action: The effects of object property discrimination and action state on affordance compatibility effects," *Psychon. Bull. Rev.*, vol. 13, no. 3, pp. 493–498, 2006.

[113] V. Gallese, L. Fadiga, L. Fogassi, and G. Rizzolatti, "Action recognition in the premotor cortex," *Brain*, vol. 119, no. 2, pp. 593–609, 1996.

[114] J. Norman, "Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches," *Behav. Brain Sci.*, vol. 25, no. 1, pp. 73–96, 2002.

[115] M. A. Arbib, A. Billard, M. Iacoboni, and E. Oztop, "Synthetic brain imaging: Grasping, mirror neurons and imitation," *Neural Netw.*, vol. 13, nos. 8–9, pp. 975–997, 2000.

[116] G. Metta and P. Fitzpatrick, "Better vision through manipulation," *Adapt. Behav.*, vol. 11, no. 2, pp. 109–128, 2003.

[117] S. T. Grafton, L. Fadiga, M. A. Arbib, and G. Rizzolatti, "Premotor cortex activation during observation and naming of familiar tools," *Neuroimage*, vol. 6, no. 4, pp. 231–236, 1997.

[118] D. E. Berlyne, *Conflict, Arousal, and Curiosity*. New York, NY, USA: McGraw-Hill, 1960.

[119] M. Hayashi and T. Matsuzawa, "Cognitive development in object manipulation by infant chimpanzees," *Animal Cogn.*, vol. 6, no. 4, pp. 225–233, 2003.

[120] F. Guerin, N. Kruger, and D. Kraft, "A survey of the ontogeny of tool use: From sensorimotor experience to planning," *IEEE Trans. Auton. Mental Develop.*, vol. 5, no. 1, pp. 18–45, Mar. 2013.

[121] J. Weng, "Developmental robotics: Theory and experiments," *Int. J. Humanoid Robot.*, vol. 1, no. 2, pp. 199–236, 2004.

[122] M. Cakmak, C. Chao, and A. L. Thomaz, "Designing interactions for robot active learners," *IEEE Trans. Auton. Mental Develop.*, vol. 2, no. 2, pp. 108–118, Jun. 2010.

[123] B. Settles, "Active learning literature survey," Univ. Wisconsin-Madison, Dept. Comput. Sci., Madison, WI, USA, Comput. Sci. Tech. Rep. 1648, 2010.

[124] C. Wang, K. V. Hindriks, and R. Babuska, "Active learning of affordances for robot use of household objects," in *Proc. IEEE RAS Int. Conf. Humanoid Robots*, Madrid, Spain, 2014, pp. 566–572.

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

[126] A. Stout, G. D. Konidaris, and A. G. Barto, "Intrinsically motivated reinforcement learning: A promising framework for developmental robot learning," in *Proc. Assoc. Adv. Artif. Intell. Spring Symp. Develop. Robot.*, 2005.

[127] N. Chentanez, A. G. Barto, and S. P. Singh, "Intrinsically motivated reinforcement learning," in *Proc. 18th Annu. Conf. Neural lnf. Process. Syst.*, 2004, pp. 1281–1288.

[128] S. M. Stringer, E. T. Rolls, and P. Taylor, "Learning movement sequences with a delayed reward signal in a hierarchical model of motor function," *Neural Netw.*, vol. 20, no. 2, pp. 172–181, 2007.

[129] E. Ugur, M. R. Dogar, M. Cakmak, and E. Sahin, "Curiosity-driven learning of traversability affordance on a mobile robot," in *Proc. IEEE Int. Conf. Develop. Learn.*, 2007, pp. 13–18.

[130] S. Thrun, "Is learning the n-th thing any easier than learning the first?" in *Proc. Adv. Neural Inf. Process. Syst.*, Denver, CO, USA, 1996, pp. 640–646.

[131] J. Weng *et al.*, "Autonomous mental development by robots and animals," *Science*, vol. 291, no. 5504, pp. 599–600, 2001.

[132] R. Pfeifer, J. Bongard, and S. Grand, *How the Body Shapes the Way We Think: A New View of Intelligence*. Cambridge, MA, USA: MIT Press, 2007.

[133] P. J. Marshall, "Beyond different levels: Embodiment and the developmental system," *Front. Psychol.*, vol. 5, pp. 1–5, 2014.

[134] L. Byrge, O. Sporns, and L. B. Smith, "Developmental process emerges from extended brain-body-behavior networks," *Trends Cogn. Sci.*, vol. 18, no. 8, pp. 395–403, 2014.

[135] M. Grounds and D. Kudenko, "Combining reinforcement learning with symbolic planning," in *Adaptive Agents and Multi- Agent Systems III. Adaptation and Multi-Agent Learning*. Heidelberg, Germany: Springer, 2008, pp. 75–86.

[136] J. Barry, K. Hsiao, L. P. Kaelbling, and T. Lozano-Pérez, "Manipulation with multiple action types," in *Experimental Robotics*, vol. 88. Cham, Switzerland: Springer, 2013, pp. 531–545.

[137] C. Lörken and J. Hertzberg, "Grounding planning operators by affordances," in *Proc. Int. Conf. Cogn. Syst.*, Karlsruhe, Germany, 2008, pp. 79–84.

[138] L. Piyathilaka and S. Kodagoda, "Affordance-map: A map for context-aware path planning," in *Proc. Australasian Conf. Robot. Autom.*, Melbourne, VIC, Australia, 2014, pp. 1–8.

[139] Y. Zhu, A. Fathi, and L. Fei-Fei, "Reasoning about object affordances in a knowledge base representation," in *Proc. Eur. Conf. Comput. Vis.*, Zürich, Switzerland, 2014, pp. 408–424.

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

[141] H. Kjellström, J. Romero, and D. Kragiæ, "Visual object-action recognition: inferring object affordances from human demonstration," *Comput. Vis. Image Understand.*, vol. 115, no. 1, pp. 81–90, 2011.

[142] Y. Jiang, "Hallucinated humans: Learning latent factors to model 3D environments," Ph.D. dissertation, Dept. Comput. Sci., Cornell Univ., Ithaca, NY, USA, 2015.

[143] V. Sarathy and M. Scheutz, "Semantic representation of objects and function," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. Workshop Learn. Object Affordances Fundamental Step Allow Prediction Plan. Tool Use*, 2015.

[144] O. Yürüten *et al.*, "Learning adjectives and nouns from affordances on the icub humanoid robot," *From Animals to Animats 12*. Heidelberg, Germany: Springer, 2012, pp. 330–340.

[145] R. Ellis and M. Tucker, "Micro-affordance: The potentiation of components of action by seen objects," *Brit. J. Psychol.*, vol. 91, no. 4, pp. 451–471, 2000.

[146] S. Thill, D. Caligiore, A. M. Borghi, T. Ziemke, and G. Baldassarre, "Theories and computational models of affordance and mirror systems: An integrative review," *Neurosci. Biobehav. Rev.*, vol. 37, no. 3, pp. 491–521, Feb. 2013.

[147] A. M. Borghi and L. Riggio, "Sentence comprehension and simulation of object temporary, canonical and stable affordances," *Brain Res.*, vol. 1253, pp. 117–128, Feb. 2009.

[148] G. Rizzolatti and M. Matelli, "Two different streams form the dorsal visual system: Anatomy and functions," *Exp. Brain Res.*, vol. 153, no. 2, pp. 146–157, 2003.

[149] M. Kammer, T. Schack, M. Tscherepanow, and Y. Nagai, "From affordances to situated affordances in robotics—Why context is important," in *Proc. IEEE Int. Conf. Develop. Learn. Epigenet. Robot.*, vol. 5. 2011, p. 30.

[150] J. Leitner, "Multi-robot cooperation in space: A survey," in *Proc. Adv. Technol. Enhanced Qual. Life*, 2009, pp. 144–151.

[151] C. Yi, H. Q. Min, and R. H. Luo, "Affordance matching from the shared information in multi-robot," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, Zhuhai, China, 2015, pp. 66–71.

[152] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[153] Y. Bengio, "Learning deep architectures for AI," *Found. Trends® Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.

[154] Y. Zhang and J. Weng, "Spatio–temporal multimodal developmental learning," *IEEE Trans. Auton. Mental Develop.*, vol. 2, no. 3, pp. 149–166, Sep. 2010.

[155] A. Stoytchev, "Some basic principles of developmental robotics," *IEEE Trans. Auton. Mental Develop.*, vol. 1, no. 2, pp. 122–130, Aug. 2009.

[156] D. Silver *et al.*, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

Huaqing Min received the Ph.D. degree in computer software and theory from the Huazhong University of Science and Technology, Wuhan, China.

He joined the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China, in 2001, where he is currently a Professor with the School of Software Engineering and the Director of the Key Laboratory of Robotics and Intelligent Software. His current research interests include intelligent robots and software.

Chang'an Yi received the B.S. degree in computer science and technology from Jingzhou Normal University, Jingzhou, China, in 2003, the M.S. degree in computer software and theory from Shantou University, Shantou, China, in 2007, and the Ph.D. degree in computer science and engineering from the South China University of Technology, Guangzhou, China, in 2015.

He is currently an Associate Professor with the School of Electronic and Information Engineering, Foshan University, Foshan, China. His current research interests include developmental and cognitive robotics, biologically-inspired robotic systems, machine learning, and computer vision.

Ronghua Luo received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 1998, 2001, and 2006, respectively.

He is currently an Associate Professor with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His current research interests include intelligent robotics, robot vision, and real time algorithms for semantic segmentation of images in order to make the robot better understand its environment, with the support from the National Natural Science Foundation of China.

Jinhui Zhu received the B.S. degree in automatic control, the M.S. degree in mechanical design and theory, and the Ph.D. degree in computer science from the South China University of Technology, Guangzhou, China, in 1999, 2002, and 2010, respectively.

He has been with the School of Computer Science and Engineering, South China University of Technology, since 2002, where he became an Associate Professor in 2011 and joined the School of Software Engineering, in 2012. His current research interests include robot software architecture, embedded operating system, machine learning, and their applications.

Sheng Bi received the M.S. and Ph.D. degrees from the South China University of Technology, Guangzhou, China, in 2003 and 2010, respectively.

He has been with the School of Computer Science and Engineering, South China University of Technology, since 2003, where he was appointed as an Associate Professor in 2014. His current research interests include intelligent robots, field-programmable gate array fast processing algorithms, embedded intelligent terminal, and smartphone development.