Autonomous Detection and Experimental Validation of Affordances

Peter Kaiser and Tamim Asfour

***Abstract*—We propose a computational formalization of affordances, which is able to consistently combine affordance-related evidence resulting from different observations. We represent affordances as Dempster-Shafer belief functions defined over the space of end-effector poses, which can be combined using uncertain logic in order to allow their hierarchical organization. The primary source of affordance-related evidence is visual affordance detection, which first simplifies the perceived environment into geometric primitives and then evaluates a hierarchical set of affordance definitions based on the available visual information. The resulting belief functions are used as initial affordance hypotheses, which are subject to further investigation and validation. As pure visual affordance detection can fail to properly estimate important preconditions, e. g. the stability of environmental structures, validation experiments are conducted in order to incrementally improve the system belief and the reliability of detected affordances. The proposed formalism is implemented and evaluated in the context of loco-manipulation affordances for humanoid robots using the simulated robots ARMAR-III and ARMAR-4.**

***Index Terms*—Humanoid Robots, Perception for Grasping and Manipulation, Semantic Scene Understanding**

I. INTRODUCTION

AUTONOMOUS humanoid robots are designed to assist humans in performing large varieties of tedious, exhausting or dangerous tasks. Such tasks include for example household work or elderly care in domestic environments, production, assembly or maintenance in workshops and factories as well as search and rescue in destructed or contaminated buildings. The environments that humanoid robots will be deployed in are human-centered and unstructured, such that they cannot be entirely known to the robot in advance. One key prerequisite for robots to be able to act in such unknown environments is their capability to reason about interaction possibilities with the environment. The psychological theory of *affordances* [1] attempts to explain the process of action possibility perception in humans and animals. It defines affordances as action possibilities latent in the environment, which arise depending on properties of perceived objects and capabilities of the perceiving agent. The theory of affordances is generally accepted as a promising foundation for autonomous robots acting in unknown environments. Moreover, the concept of affordances provides a powerful mechanism for high-level robot control, where affordance-based scene representations are used to operate a humanoid robot by a human pilot. Such *shared autonomous* control modes would greatly benefit if affordance-based perception would reduce the large space of possible actions to a limited set of actions, which are reasonable and applicable in the current context. In this work we aim at developing the foundation for a robotic perceptive-cognitive system that addresses locomotion and manipulation affordances in both, autonomous and shared autonomous control modes.

Manuscript received: September, 10, 2017; Revised December, 7, 2017; Accepted January, 29, 2018.

This paper was recommended for publication by Editor Nikos Tsagarakis upon evaluation of the Associate Editor and Reviewers' comments. The research leading to these results has received funding from the European Union Seventh Framework Programme under grant agreement no 611832 (WALK-MAN).

The authors are with the Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Karlsruhe, Germany. {peter.kaiser,asfour}@kit.edu

Digital Object Identifier (DOI): see top of this page.

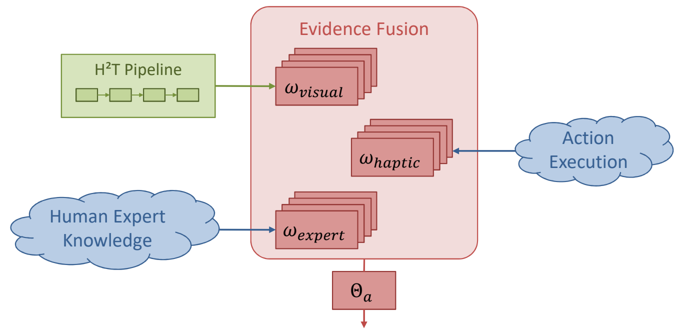


Figure 1: Affordance-related evidence from different sources, such as our perceptual pipeline, human expert knowledge and action executions (green and blue blocks), is aggregated into a set of observations (red blocks). Evidence fusion describes the process of deriving a joint system belief Θ_a from the available observations.

A. Previous Work

In our previous works [2], [3], [4] and [5], we developed and evaluated a computational approach to the detection of affordances in unknown environments. The environment is perceived using an RGB-D camera and subsequently simplified into a representation in terms of geometric primitives such as planes, cylinders and spheres (see Fig. 2). The formalization of affordances employed in our previous work, particularly [4], is based on two fundamental observations:

1) *Affordances are hierarchical*, i. e. the existence of sophisticated affordances is based on the existence of more primitive affordances.
2) *End-effector contact is crucial* for the considered actions and therefore serves as fundamental affordances in our approach.

The foundation of the affordance detection system as proposed in our previous work is the notion of *affordance*

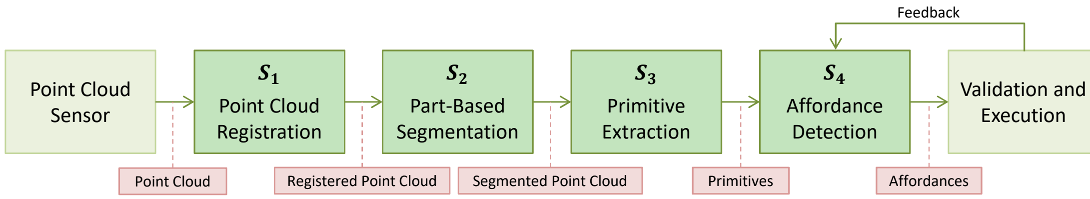


Figure 2: The process of affordance detection (step $\boldsymbol{S_4}$) and validation as a fusion of affordance-related evidence. Evidence results from visual scene information (steps $\boldsymbol{S_1}$, $\boldsymbol{S_2}$ and $\boldsymbol{S_3}$) and feedback of executed actions. While the visual detection of affordances has been employed in our previous works [2], [3], [4] and [5], the understanding of affordance detection as a process of multimodal evidence fusion is a contribution of this paper.

certainty functions Θ_a, which assess the existence certainty of affordances a for given end-effector poses $\boldsymbol{x} \in SE(3)$. Higher-level affordance certainty functions may include certainty estimations from lower-level affordances. *Supportability* is for example composed from *platform graspability* in combination with a horizontal primitive orientation and the primitive's assumed immobility (see Fig. 3). Both of these properties are estimated based on fundamental geometric attributes of the primitive. The framework for affordance detection based on hierarchical affordance certainty functions has been successfully evaluated in multiple scenarios in simulation and on the real humanoid robots ARMAR-III [4] and WALK-MAN [5]. However, we identified two principle problems with our previous approach which are addressed in this work:

1) Affordance-based evidence from sources other than the visual pipeline cannot be considered.
2) Affordance certainty functions do not naturally allow logic operations suitable for their hierarchical composition.

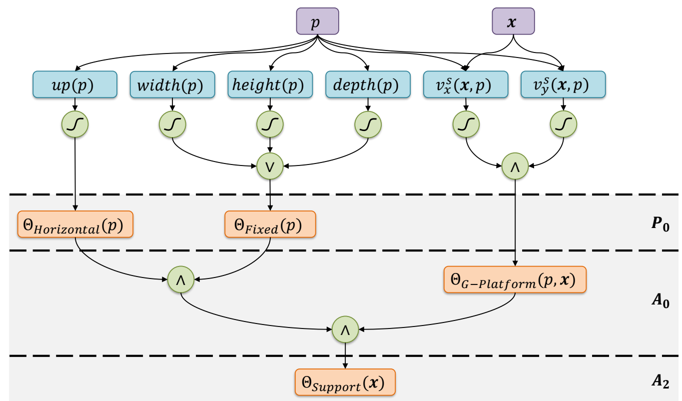


Figure 3: The belief function for a *supportability* affordance $\Theta_{\text{Support}}(\boldsymbol{x})$ is composed of a *platform graspability* affordance $\Theta_{\text{G-Platform}}(\boldsymbol{x})$ in combination with belief functions for the *horizontality* and the *mobility* of the primitive p, which are estimated based on geometric properties of p. See [4] for further details.

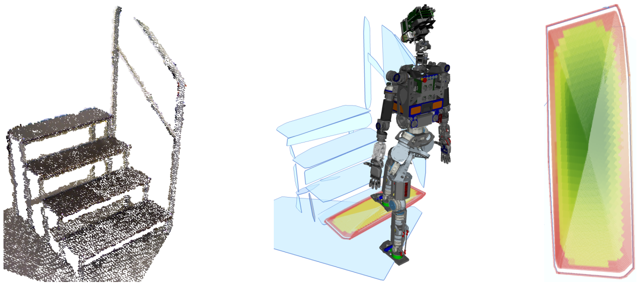
B. Contribution

In this work we extend the previously proposed concept of affordance certainty functions to *affordance belief functions*, which allow the fusion of evidence from various, possibly contradicting, sources and formally support hierarchical organization. We approach this goal by formalizing affordances as *Dempster-Shafer belief* over the space of end-effector poses. In contrast to conventional probability theory, Dempster-Shafer theory (*DST*) [6] allows the consistent combination of evidence from different sources with different assigned degrees of belief. An extensive survey on DST and its various extensions can be found in [7]. We evaluate the principles of affordance-related evidence fusion in two ways: 1) by simulating large amounts of observations drawn from randomized ground-truth affordance belief functions, and 2) by letting the simulated robot ARMAR-III autonomously perform affordance validation experiments in a dynamically simulated kitchen environment (see Fig. 5).

The remainder of the paper is structured as follows: After discussing related work in Section II, we formalize the affordance concept in Section III and the principles of evidence fusion and hierarchy in Section IV and Section V. In Section VI we evaluate the proposed concept with the humanoid robot ARMAR-III, before concluding in Section VII.

II. Related Work

The concept of affordances has inspired many approaches to describing and formalizing the link between perception and action in robotics, resulting in a large variety of affordance-based methods. These approaches differ in their use-cases and their interpretation of the affordance concept. Extensive surveys on the application of affordances in robotics can be found in [8], [9], [10] and [11]. Several approaches attempt to learn the relations between visual features in depth images and affordances, e. g. using selected geometric features [12], CNNs [13] or spatial relations between objects and human body parts [14]. In contrast to such learning-based approaches, we aim at detecting and validating affordances based on primitive geometries and agent embodiments, without requiring significant amounts of labeled training data. The authors in [15] and [16], describe how to infer object affordances by learning object categories, action effects or action parameters for ultimately performing symbolic planning based on learned object affordances. The authors of [17] propose a framework for symbolic inference of affordances under uncertainty using DST. Our concept differs from the above work by concentrating on the lower-level perceptive-cognitive process of the robot, not assuming any prior information about objects and their recognition. Furthermore, some



(a) (b) (c)

Figure 4: An exemplary situation of affordance validation based on real sensor data: The validation of *supportability* Θ_{Support} in a stepping scenario. The uncertainty of the initial visual affordance detection is defined to be relatively high ($\eta = 0.5$), indicated by the yellow visualization of the associated belief function (see Fig. 6). The validation experiment (b) emphasizes the existence of the affordance and therefore increases belief in a local neighborhood around the end-effector pose used for the validation experiment (c).

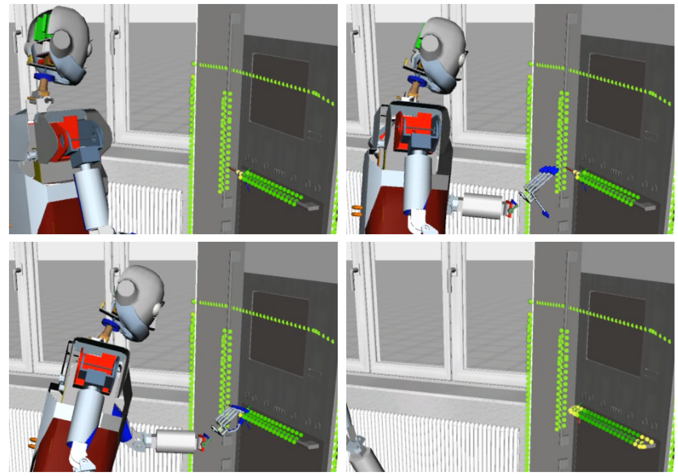


Figure 5: Validation of *prismatic graspability* in a simulated kitchen environment. The evidence gained from the successful haptic validation experiment is fused with existing evidence from visual affordance detection, producing an updated joint affordance belief function $\Theta_{\text{G-Prismatic}}$. The affordance belief function is visualized according to the color map in Fig. 6.

works attempt to jointly learn visual affordance perception and symbolic or subsymbolic information about actions and their effects: e. g. based on denoising autoencoders [18] or Bayesian Networks [19]. More pragmatic approaches define affordances as templates, containing 3D representations of objects and conceptual descriptions of their utilization. Due to the embedded object description, such templates can be easily recognized in captured point cloud data, eventually allowing to reason about available action possibilities. Prominent examples have been demonstrated at the DARPA Robotics Challenge, including [20], [21] and [22]. Template-based approaches work well in environments, where principle structures are known with small variations tolerated. In contrast to template-based approaches, we assume perceived environments to be entirely unknown to the robot. A conceptually related approach for the detection of loco-manipulation affordances was used in [23] for locomotion planning in unknown environments. While this approach is particularly focused on the direct detection of affordances for hand and foot contacts, we aim at improving the robustness of affordance detection by integrating multiple sensory experiments.

Although not explicitly denoted in the following formalization, affordance belief functions as proposed in this work establish a link between geometric primitives p and affordances a. In our general framework introduced in [4], affordances are further linked with combined representations of action execution skills, symbolic preconditions and effects. Hence, this work can be seen in the tradition of popular approaches which understand affordances as relations between objects, actions and action effects (e. g. [24]).

III. Affordance Formalization

The central contribution of this paper is a computational formalization of loco-manipulation affordances which satisfies the two prerequisites introduced in Section I:

1. Evidence Fusion: The consistent fusion of affordance-related evidence is important when considering a humanoid robot as an inherently redundant machine, offering a multitude of possible sensor modalities which provide information regarding the existence of affordances. Affordance-related evidence could further result from expert knowledge or the robot's own learned experience. The possible availability of affordance-related evidence from different sources with different attributed reliabilities necessitates a consistent formalism for evidence fusion.

2. Hierarchical Representation: The requirement of a hierarchical organization of affordances becomes self-evident by the observation that many loco-manipulation actions establish power grasping contact with environmental structures. Hence, *graspability* affordances can be considered prerequisites for higher-level affordances like *pushability* or *pullability*, which themselves serve as prerequisites for affordances such as *bimanual pushability*. The hierarchical representation of affordances proposed in this work ensures that evidence of lower-level affordances is appropriately propagated to higher-level affordances and finally allows the hierarchical organization of loco-manipulation affordances as introduced in [4].

A. Affordance Belief Functions

Based on the observation that end-effector contact is crucial for loco-manipulation actions, an affordance a is defined to exist with respect to a geometric primitive p and end-effector poses $\boldsymbol{x} \in \mathrm{SE}(3)$. Belief in the existence of a is expressed by *affordance belief functions* $\Theta_a(\boldsymbol{x})$, mapping end-effector poses $\boldsymbol{x} \in \mathrm{SE}(3)$ to Dempster-Shafer belief expressions $d \in \mathcal{D}$:

$$\Theta_a \colon \mathrm{SE}(3) \to \mathcal{D}. \tag{1}$$

The above belief function is suitable for representing unimanual affordances. While this focus on the unimanual case will prevail throughout this paper, the formalism can be extended to consider an arbitrary number of N end-effectors by extending the definition space of Θ_a:

$$\Theta_a\colon \underbrace{\mathrm{SE}(3) \times \cdots \times \mathrm{SE}(3)}_{N \text{ times}} \to \mathcal{D}. \tag{2}$$

For expressing the system belief $\Theta_a(\boldsymbol{x})$ in the existence of an affordance a with respect to an end-effector pose $\boldsymbol{x} \in \mathrm{SE}(3)$, two fundamental hypotheses are defined: a^+ representing the assumption that a exists and a^- representing the assumption that a does not exist. It is the inherent task of the affordance detection and validation system to obtain certainty about which of the two hypotheses is true by combining and evaluating available evidence. The set $\mathcal{X}_a = \{a^+, a^-\}$ is called *hypothesis space* and the set of possible combinations of hypotheses results in:

$$2^{\mathcal{X}_a} = \big\{\emptyset, \{a^+\}, \{a^-\}, \mathcal{X}_a\big\}. \tag{3}$$

In the interest of simplicity the notations are abbreviated to $a^+ := \{a^+\}$ and $a^- := \{a^-\}$. A hypothesis space which is suitable for DST must be *complete* and *mutually exclusive*, i. e. it must contain the true hypothesis and no two hypotheses can be true at the same time. As affordances can only either exist or not exist, and as both possibilities are reflected in $\mathcal{X}_a$ as distinct hypotheses, the hypothesis space is complete. As the complements a^+ and a^- are the only contained hypotheses, the hypothesis space is further mutually exclusive. These two properties justify the applicability of DST to the hypothesis space $\mathcal{X}_a$. The space $\mathcal{X}_a$ contains two complementary hypotheses, and therefore constitutes the simplest non-degenerated case of a hypothesis space. Such *binary* spaces would more classically be denoted as $\mathcal{X}_a = \{a, \neg a\}$. The simplicity of $\mathcal{X}_a$ will play an important role for the formalization and for the computational feasibility of the approach. In the DST, belief is formally expressed by attributing probability mass to each element of the power set $2^{\mathcal{X}_a}$. Such probability mass assignments $m\colon 2^{\mathcal{X}_a} \to [0,1]$ are called *basic belief assignments* if $m(\emptyset) = 0$ and $\sum_{A \in 2^{\mathcal{X}_a}} m(A) = 1$. Distributed probability mass can be intuitively interpreted as follows:

- $m(a^+)$ expresses belief in the existence of a
- $m(a^-)$ expresses belief in the non-existence of a
- $m(\mathcal{X}_a)$ expresses uncertainty about the existence of a.

The space of belief expressions $\mathcal{D}$ introduced in Eq. 1 can now formally be defined as the space of possible basic belief assignments:

$$\mathcal{D} := \left\{ m\colon 2^{\mathcal{X}_a} \to [0,1] \;\middle|\; m(\emptyset) = 0, \sum_{A \in 2^{\mathcal{X}_a}} m(A) = 1 \right\}.$$

In order to simplify the following formalizations, the evaluation of affordance belief functions Θ_a for end-effector poses $\boldsymbol{x} \in SE(3)$ and hypotheses $A \in 2^{\mathcal{X}_a}$ is abbreviated as:

$$\Theta_a(\boldsymbol{x}, A) := \big(\Theta_a(\boldsymbol{x})\big)(A). \tag{4}$$

The DST defines two fundamental functions: *belief* $\mathrm{bel}(A)$, describing the confidence that $A \in 2^{\mathcal{X}_a}$ contains the true hypothesis, and *plausibility* $\mathrm{pl}(A)$, describing the confidence that the true hypothesis does not contradict A. Belief and plausibility can be expressed for affordance belief functions Θ_a, end-effector poses $\boldsymbol{x} \in SE(3)$ and hypotheses $A \in 2^{\mathcal{X}_a}$ as follows:

$$\begin{aligned} \mathrm{bel}_a(\boldsymbol{x}, A) &= \sum_{B \subseteq A} \Theta_a(\boldsymbol{x}, B) \in [0,1] \\ \mathrm{pl}_a(\boldsymbol{x}, A) &= \sum_{B \cap A \neq \emptyset} \Theta_a(\boldsymbol{x}, B) \in [0,1]. \end{aligned} \tag{5}$$

The set-theoretic definitions of belief and plausibility can become computationally hard for large hypothesis spaces. However, by exploiting the simplicity of $\mathcal{X}_a$, the equations become pleasantly simple. The (classical) probability $p(A)$ of a hypothesis $A \in 2^{\mathcal{X}_a}$ can be interpreted to lie in the interval $[\mathrm{pl}(A), \mathrm{bel}(A)]$. Based on this relation, the *expected probability* $E_a(\boldsymbol{x}, A)$ as described in [25] is used in cases when belief expressions are compacted into single real numbers:

$$E_a(\boldsymbol{x}, A) = \mathrm{bel}_a(\boldsymbol{x}, A) + \frac{1}{2}\big(\mathrm{pl}_a(\boldsymbol{x}, A) - \mathrm{bel}_a(\boldsymbol{x}, A)\big). \tag{6}$$

Note that the definition of expected probability used here is the special case for binary hypothesis spaces. The formalization of affordances as belief functions Θ_a over the space of end-effector poses constitutes the core of our proposed affordance formalization. In the following section, the initial requirements of evidence fusion and hierarchy will be properly formalized.

IV. Evidence Fusion

For formalizing the process of evidence fusion, let $\Omega = \{\omega_1, \ldots, \omega_N\}$ be a sequence of *observations*. Each observation $\omega \in \Omega$ is defined as an affordance belief function $\omega : SE(3) \to \mathcal{D}$. Hence, observations express affordance-related evidence over the space of end-effector poses. Fig. 1 shows an overview of the concept of evidence fusion. The DST defines *Dempster's rule of combination* as an associative operator $\oplus$ for combining *compatible* basic belief assignments. Two basic belief assignments are compatible if they are defined over the same hypothesis space. Hence, in the context of affordance belief functions, two observations are compatible if they express evidence related to the same affordance. For $A \in 2^{\mathcal{X}_a}, A \neq \emptyset$, the fusion of compatible observations $\omega_1, \ldots, \omega_N$ is formalized as:

$$\left(\bigoplus_{i=1}^{N} \omega_i\right)(\boldsymbol{x}, A) = \frac{1}{1 - K(\boldsymbol{x})} \sum_{\left(\bigcap_{j=1}^{N} A_j\right) = A} \prod_{k=1}^{N} \omega_k(\boldsymbol{x}, A_k)$$

with the *conflict factor* $K(\boldsymbol{x})$:

$$K(\boldsymbol{x}) = \sum_{\left(\bigcap_{j=1}^{N} A_j\right) = \emptyset} \prod_{k=1}^{N} \omega_k(\boldsymbol{x}, A_k). \tag{7}$$

Incremental evidence fusion is possible due to the associativity of the combination rule. The simplicity of $\mathcal{X}_a$ allows the simplification of the combination rule into efficiently computable equations.

A. Spatial Generalization of Selective Observations

We consider two fundamental categories of observations, which can occur depending on the utilized sensors and the experimental setup:

1) *Extensive Observations:* These observations inherently provide spatially distributed evidence, as the employed sensor and the experimental setup evaluate affordances for whole ranges of possible end-effector poses. Extensive observations are for example produced by visual affordance detection.
2) *Selective Observations:* These observations provide evidence for specific end-effector poses only. Selective observations are for example produced by haptic affordance validation.

In order to allow effective reasoning about affordances in a larger scale, selective observations need to be spatially generalized, which in accordance to [26] is performed by combining two distribution functions

$$\begin{aligned} n(\boldsymbol{x}_{\text{ref}}, \boldsymbol{x}) &\propto \mathcal{N}\left(\mathfrak{t}\left(\boldsymbol{x}_{\text{ref}}\right), \sigma^2_{\text{pos}}\right) \\ m(\boldsymbol{x}_{\text{ref}}, \boldsymbol{x}) &\propto \mathcal{M}\left(\mathfrak{R}\left(\boldsymbol{x}_{\text{ref}}\right), \sigma^2_{\text{rot}}\right), \end{aligned} \tag{8}$$

for the translational component $\mathfrak{t}(\boldsymbol{x}_{\text{ref}})$ and the rotational component $\mathfrak{R}(\boldsymbol{x}_{\text{ref}})$ of the end-effector pose $\boldsymbol{x}_{\text{ref}}$, respectively. The distribution function n is proportional to a normal distribution $\mathcal{N}$, while m is proportional to a von Mises-Fisher distribution $\mathcal{M}$, both normalized to a maximum value of 1. Using the combined distribution function $\delta(\boldsymbol{x}_{\text{ref}}, \boldsymbol{x}) = n(\boldsymbol{x}_{\text{ref}}, \boldsymbol{x}) \cdot m(\boldsymbol{x}_{\text{ref}}, \boldsymbol{x})$, the spatial generalization of selective observations ω with associated observation certainty $\eta \in [0, 1]$ is modeled as:

$$\omega_\eta(\boldsymbol{x}, A) = \begin{cases} \eta \cdot \delta(\boldsymbol{x}_{\text{ref}}, \boldsymbol{x}) \cdot \omega(\boldsymbol{x}, a^+), & \text{if } A = a^+ \\ \eta \cdot \delta(\boldsymbol{x}_{\text{ref}}, \boldsymbol{x}) \cdot \omega(\boldsymbol{x}, a^-), & \text{if } A = a^- \\ 1 - \omega(\boldsymbol{x}, a^+) - \omega(\boldsymbol{x}, a^-), & \text{if } A = \mathcal{X}_a. \end{cases}$$

B. Examples

For exemplary visualization of 2D affordance belief functions, the system belief is projected into the HSL color space by computing the expected probabilities of the hypotheses a^+ and a^- and computing the *decision value* $v_a(\boldsymbol{x}) \in [0, 1]$:

$$v_a(\boldsymbol{x}) = \frac{1}{2}\big(E_a(\boldsymbol{x}, a^+) - E_a(\boldsymbol{x}, a^-) + 1\big). \tag{9}$$

This value is transformed into a hue interval ranging from green to red, while the lightness value represents uncertainty (see Fig. 6). Two-dimensional affordance belief functions constructed from eight consecutive selective observations are shown in Fig. 7. Confirming observations emphasize belief in the existence of the investigated affordance, resulting in dark green areas of high belief $\text{bel}(a^+)$ (Fig. 7a). Contradicting observations emphasize belief in the absence of the investigated affordance, resulting in dark red color in areas of high belief $\text{bel}(a^-)$ (Fig. 7c). The principle of evidence fusion based on Dempster-Shafer belief functions is visualized in Fig. 4 for the exemplary scenario of *stepping* based on real sensor data. The affordance detection system successfully

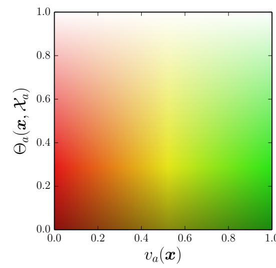
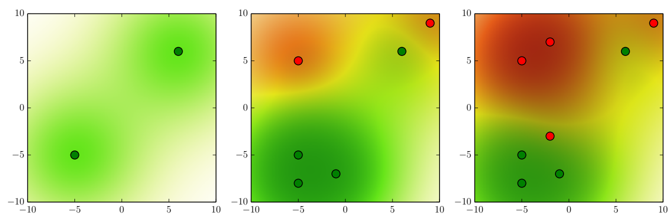


Figure 6: Affordance belief functions are visualized by projection to the HSL color space: The decision value $v_a(\boldsymbol{x})$ is represented by the hue value, ranging from red to green, while red indicates predominant belief in a^- and green indicates predominant belief in a^+. Uncertainty $\Theta_a(\boldsymbol{x}, \mathcal{X}_a)$ is represented by the lightness value.



(a) (b) (c)

Figure 7: Visualization of an affordance belief function, iteratively composed from eight consecutive selective observations, visualized as red and green dots, for 2D end-effector positions $\boldsymbol{x} \in [-10, 10] \times [-10, 10]$.

evaluates the belief function Θ_{Support} for *supportability* for each of the steps. A *supportability* affordance is derived from a *platform graspability* affordance in combination with a horizontal orientation of the primitive surface (see Fig. 3). The visual detection of support surfaces is particularly critical as the physical stability of the surface is hard to estimate purely based on visual information. The conduction of affordance validation experiments in addition to visual perception is therefore a promising strategy for reliable legged locomotion in unknown environments. The initial belief function obtained through visual perception exposes a large degree of uncertainty, which is reduced using a validation experiment.

V. Hierarchy

As Dempster's rule of combination is only defined for belief assignments that share the same hypothesis space, the combination of belief functions Θ_{a_1} and Θ_{a_2} for different affordances a_1 and a_2 is not possible. In [4] we proposed to compose affordances based on hierarchical derivation rules which requires the capability of combining affordance belief functions in the sense of logic operations. As a simple example, consider a hierarchical rule for the existence of *supportability* affordances, for which *platform graspability* and the horizontal primitive orientation are given as belief functions $\Theta_{\text{G-Platform}}$ and $\Theta_{\text{Horizontal}}$:

$$\frac{\Theta_{\text{G-Platform}}(\boldsymbol{x}) \wedge \Theta_{\text{Horizontal}}(p)}{\Theta_{\text{Support}}(\boldsymbol{x})}. \tag{10}$$

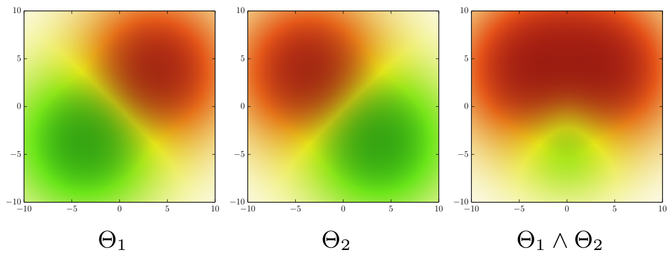


Figure 8: DS-theoretic conjunction $(\Theta_1 \wedge \Theta_2)(\boldsymbol{x})$ applied to two assumed affordance belief functions $\Theta_1(\boldsymbol{x})$ and $\Theta_2(\boldsymbol{x})$ for 2D end-effector positions $\boldsymbol{x} \in [-10, 10] \times [-10, 10]$.

In this case, Θ_{Support} is called the *higher-level affordance* as its existence depends on the *lower-level affordance* $\Theta_{\text{G-Platform}}$. The *Theory of Subjective Logic* [25] provides the theoretical means for applying logic operations to Dempster-Shafer belief values. Let a and b be distinct affordances with respective hypothesis spaces $2^{\mathcal{X}_a}$ and $2^{\mathcal{X}_b}$ and $\boldsymbol{x} \in SE(3)$ be an end-effector pose. Further, let $A \in 2^{\mathcal{X}_a}$ and $B \in 2^{\mathcal{X}_b}$ be affordance hypotheses. Then the subjective logic operations $A \wedge B$, $A \vee B$ and $\neg A$ are defined as follows:

$$\begin{aligned} \text{bel}_{a\wedge b}(\boldsymbol{x}, A \wedge B) &= \text{bel}_a(\boldsymbol{x}, A) \cdot \text{bel}_b(\boldsymbol{x}, B) \\ \text{bel}_{a\vee b}(\boldsymbol{x}, A \vee B) &= \text{bel}_a(\boldsymbol{x}, A) + \text{bel}_b(\boldsymbol{x}, B) \\ &\quad - \text{bel}_a(\boldsymbol{x}, A) \cdot \text{bel}_b(\boldsymbol{x}, B) \\ \text{bel}_a(\boldsymbol{x}, \neg A) &= 1 - \text{pl}_a(\boldsymbol{x}, A). \end{aligned} \tag{11}$$

Note that, with the exception of the negation $\neg A$, resulting belief is expressed over the new hypothesis spaces $2^{\mathcal{X}_{a\wedge b}}$ and $2^{\mathcal{X}_{a\vee b}}$. The initially stated affordance inference rule from Eq. 10 can now be written as:

$$\Theta_{\text{Support}}(\boldsymbol{x}) = \Theta_{\text{G-Platform}}(\boldsymbol{x}) \wedge \Theta_{\text{Horizontal}}(p). \tag{12}$$

Fig. 8 shows a visualization of the DS-theoretic logic AND operator applied to exemplary affordance belief functions Θ_1 and Θ_2. The images show that subjective logic operations applied to affordance belief functions with areas of different belief and uncertainty, produce intuitive results.

VI. Evaluation

The hierarchical formulation of loco-manipulation affordances as functions over the end-effector pose space has been evaluated in various realistic scenarios on real humanoid robot platforms, including ARMAR-III and WALK-MAN in our previous work. The performed experiments demonstrate the feasibility of the concept and its applicability to realistic scenarios with real sensor data (see e. g. Fig. 4). The central contribution of this work is the computational formalization of the affordance concept in terms of *affordance belief functions* suitable for the hierarchical arrangement of whole-body affordances. Hence, the presented experiments concentrate on the evaluation of the proposed evidence fusion mechanism in synthetic scenes (Section VI-A) and in the system-wide application of multimodal affordance detection and validation in a dynamic simulation environment (Section VI-B).

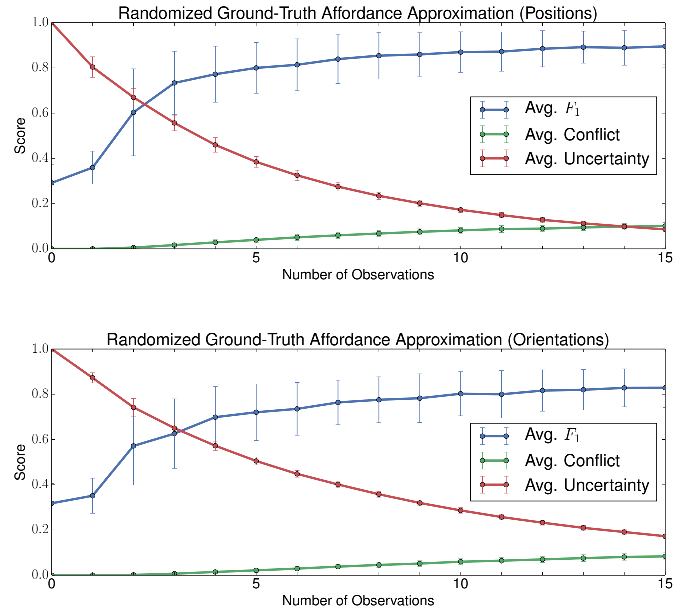


Figure 9: Averaged F_1 scores, conflict and uncertainty of joint affordance belief functions obtained via consecutive fusion of increasing numbers of observations drawn from ground-truth affordance distributions. The figures have been averaged over 400 ground-truth affordances each with $\sigma_{\text{pos}} = 0.6$, $\sigma_{\text{rot}} = \frac{\pi}{8}$ and $\eta = 0.6$.

A. Affordance Belief Functions

We first investigate if the proposed mechanisms for evidence fusion are able to produce consistent joint belief functions via the fusion of consecutive observations, that converge against assumed ground-truth affordances. More formally, we assume ground-truth affordance distributions $\mathcal{G}\colon \text{SE}(2) \rightarrow \{0, 1\}$ which are generated as intersections of randomly sampled half-spaces in 2D. Sequences of observations $\omega_1, \ldots, \omega_N$ are then sampled from the ground-truth affordances. In analogy to the plots in Fig. 7, affordances are defined in 2D, either representing the positional or orientational components of $\text{SE}(2)$. Fig. 9 shows averaged F_1 scores of the obtained joint belief functions for end-effector positions (*top*) and orientations (*bottom*) over increasing numbers of fused observations. The results indicate that affordance belief functions are able to accurately resemble ground-truth affordance distributions via the fusion of spatially distributed observations ω. While the resulting belief functions do not *exactly* represent the respective ground-truth affordances, an F_1 score greater than 0.8 for both, positions and orientations, obtained from few observations is suitable for the intended application. This is particularly true as extensive observations such as visual affordance detection provide prior belief in real applications. The results further show that the average uncertainty $\Theta_a(\boldsymbol{x}, \mathcal{X}_a)$ decreases with the growing number of observations, indicating that the system belief converges against a state of high certainty. Further, the average conflict $\Theta_a(\boldsymbol{x}, a^+) \cdot \Theta_a(\boldsymbol{x}, a^-)$ moderately increases with the number of observations, which is expected as the fusion of contradicting evidence causes conflict.

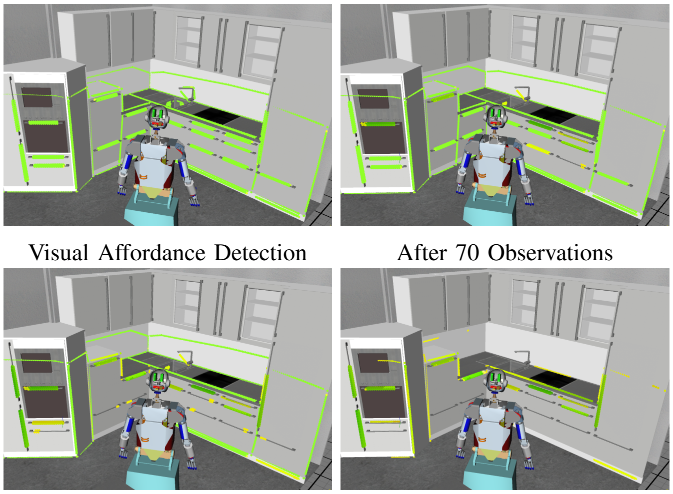


Figure 10: Visualization of the affordance belief function for *prismatic graspability* in different stages of the evaluation. Validation experiments predominantly reduce the amount of false positives, e. g. the handles of low drawers, which are unreachable for the robot.

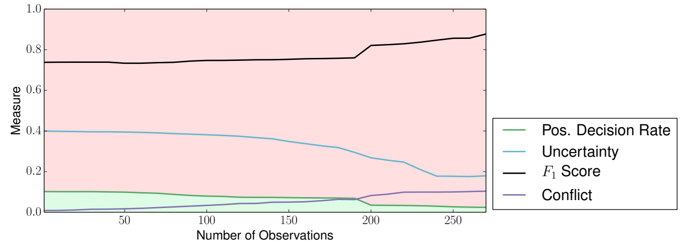


Figure 11: Validation of visually detected *prismatic graspability* affordances through consecutive validation experiments in a simulated kitchen environment.

B. Kitchen Evaluation Scenario

In the second evaluation scenario, we simulate the iterative, haptic validation of *prismatic graspability* affordances in a dynamic simulation environment using the humanoid robot ARMAR-III. With this evaluation scenario, we aim at demonstrating the feasibility of the fusion of multimodal affordance-related evidence in the context of humanoid robotics. In this scenario, visual scene perception is simulated by passing a manually segmented point cloud of the kitchen environment to the perception system which extracts geometric primitives and evaluates the loco-manipulation affordance hierarchy as described in Fig. 2. The robot then successively selects end-effector poses based on a measure $C_a(\boldsymbol{x})$ that incorporates uncertainty and conflict:

$$C_a(\boldsymbol{x}) = \underbrace{\Theta_a(\boldsymbol{x}, \mathcal{X}_a)}_{\text{Uncertainty}} + \underbrace{\Theta_a(\boldsymbol{x}, a^+) \cdot \Theta_a(\boldsymbol{x}, a^-)}_{\text{Conflict}} . \tag{13}$$

This measure is exemplarily implemented in order to derive end-effector poses $\boldsymbol{x}$ that seem interesting for affordance validation as the existing belief contains large portions of uncertainty or conflict. Different approaches to assessing potential validation poses are possible and a thorough review is beyond the scope of this work. The general structure of the experiment is as follows:

1) Simulated perception of the full environment and visual affordance detection
2) Selection of a geometric primitive p to inspect. Primitives p are validated in the ascending order of their surface area and primitives are considered validated if the maximum uncertainty of the $\Theta_{\text{G-Prismatic}}$ is below 0.3.
3) Selection of the end-effector pose $\boldsymbol{x}$ for validation based on the measure C_a, i. e. $\boldsymbol{x} \leftarrow \arg\max_{\boldsymbol{x} \in SE(3)} C_a(\boldsymbol{x})$.
4) Execution of a validation action for $\boldsymbol{x}$ and simulation of the achieved effects.
5) Generation of an observation ω based on the result of the validation action and fusion of ω with the existing belief $\Theta_{\text{G-Prismatic}}$.
6) Go to step 2.

The employed validation action for *prismatic graspability* first determines a suitable robot pose for action execution and subsequently attempts a prismatic grasp at the chosen end-effector pose $\boldsymbol{x}$ (see Fig. 5). The success of the grasping attempt is determined based on the hand joint angles. The results depicted in Fig. 10 and Fig. 11 show that the total amount of uncertainty in the affordance belief functions decreases with the number of performed validation experiments. The results further demonstrate that the robot can gradually improve the initial belief from visual affordance detection, which is already relatively accurate ($F_1 > 0.6$), by performing consecutive haptic validation experiments. In order to achieve reproducible results, the evaluation experiment was structured in the sense that the robot attempted to validate geometric primitives in the order of their size (defined by their surface area). Because early validation actions are applied to smaller primitives, the effects of these validation actions are less visible in Fig. 11 than those of later validation actions. Multiple validation experiments can be identified that lead to large gains in the F_1 score by significantly eliminating false positives. Note that the excessive amount of validation experiments carried out in this evaluation scenario, although providing a suitable validation of the evidence fusion formalism, does not qualify as a general strategy for affordance-based autonomy. In real applications, autonomous and shared autonomous humanoid robots are intended to perform individual validation experiments in cases of high risk or uncertainty, possibly demanded by a human pilot.

C. Performance

Fig. 12 displays runtime measurements for the computation of eight affordance belief functions in the exemplary scenes depicted in Fig. 4 and Fig. 10 and highlights the time spent in the individual pipeline steps $\boldsymbol{S_2}$, $\boldsymbol{S_3}$ and $\boldsymbol{S_4}$ (see Fig. 2), as well as general system overhead. Fig. 12 further subdivides the runtime of the affordance detection step $\boldsymbol{S_4}$ into the times spent for computing individual affordance belief functions. The runtimes have been generated on a standard desktop PC

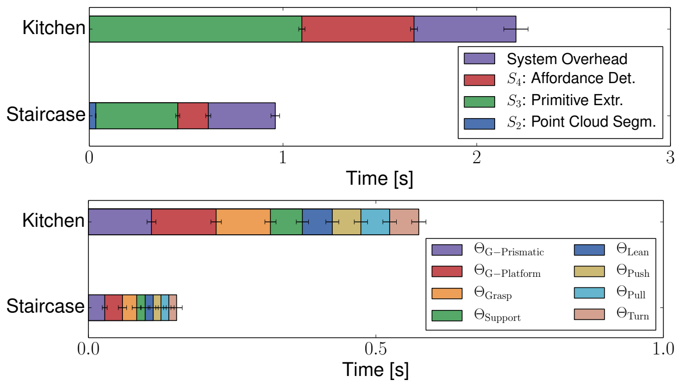


Figure 12: Runtime measurements for the evaluation of eight different affordance belief functions in the evaluation scenarios shown in Fig. 4 and Fig. 10. *Top:* Runtime measurements of the individual system components shown in Fig. 2. *Bottom:* Runtime measurements of the evaluation of the individual affordance belief functions.

using a spatial sampling distance of 2.5 cm and 16 sampled orientations per position, averaged over 100 measurements.

VII. Conclusion and Future Work

In this work we proposed a novel concept for the formalization and experimental validation of loco-manipulation affordances in unknown environments. We formalize affordances as belief functions over the space of end-effector poses which allow their hierarchical organization. Moreover, the formalization allows the consistent fusion of affordance-related evidence from multiple sources and sensorimotor experience with different degrees of certainty. Visual perception in this system is treated as one out of possibly many sources of experimental evidence. The concept of DS-based affordance belief functions was first theoretically introduced and then evaluated in two ways: 1) by comparing joint belief functions aggregated from sequences of observations with randomly generated ground-truth affordances and 2) by performing validation experiments using the humanoid robot ARMAR-III in a simulated kitchen environment. The experiments demonstrate the strengths of the proposed formalism in the consistent fusion of affordance-related evidence from multiple sources into a joint system belief. In our future work, we plan to extend the evaluation scenario towards the simultaneous validation of multiple affordances, demonstrating the propagation of evidence in the affordance hierarchy from [4] both, in simulation and on the real robot ARMAR-III. Furthermore, we aim at improving the formalization of affordance belief functions with respect to the concept of time in order to capture temporally extended observations or environmental state changes.

References

[1] J. J. Gibson, *The Ecological Approach to Visual Perception*. 1978.

[2] P. Kaiser, N. Vahrenkamp, F. Schültje, J. Borràs, and T. Asfour, "Extraction of Whole-Body Affordances for Loco-Manipulation Tasks," *International Journal of Humanoid Robotics*, vol. 12, no. 3, 2015.

[3] P. Kaiser, M. Grotz, E. E. Aksoy, M. Do, N. Vahrenkamp, and T. Asfour, "Validation of Whole-Body Loco-Manipulation Affordances for Pushability and Liftability," in *IEEE-RAS International Conference on Humanoid Robots*, 2015.

[4] P. Kaiser, E. E. Aksoy, M. Grotz, and T. Asfour, "Towards a hierarchy of whole-body loco-manipulation affordances," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2016.

[5] P. Kaiser, D. Kanoulas, M. Grotz, L. Muratore, A. Rocchi, E. M. Hoffman, N. G. Tsagarakis, and T. Asfour, "An affordance-based pilot interface for high-level control of humanoid robots in supervised autonomy," in *IEEE-RAS International Conference on Humanoid Robots*, 2016.

[6] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[7] K. Sentz and S. Ferson, "Combination of evidence in dempster-shafer theory," in *SAND2002-0835, Sandia National Laboratories*, 2002.

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

[9] T. E. Horton, A. Chakraborty, and R. S. Amant, "Affordances for robots: a brief survey," *Avant*, vol. 3, no. 2, pp. 70–84, 2012.

[10] H. Min, C. Yi, R. Luo, J. Zhu, and S. Bi, "Affordance research in developmental robotics: A survey," *IEEE Trans. on Cognitive and Developmental Systems*, 2016.

[11] L. Jamone, E. Ugur, A. Cangelosi, L. Fadiga, A. Bernardino, J. Piater, and J. Santos-Victor, "Affordances in psychology, neuroscience and robotics: a survey," *IEEE Trans. on Cognitive and Developmental Systems*, 2016.

[12] A. Myers, C. L. Teo, C. Fermüller, and Y. Aloimonos, "Affordance detection of tool parts from geometric features," in *IEEE International Conference on Robotics and Automation*, pp. 1374–1381, 2015.

[13] A. Nguyen, D. Kanoulas, D. G. Caldwell, and N. G. Tsagarakis, "Detecting object affordances with convolutional neural networks," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2765–2770, 2016.

[14] H. S. Koppula and A. Saxena, "Anticipating human activities using object affordances for reactive robotic response," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 14–29, 2016.

[15] E. Uğur and J. Piater, "Bottom-up learning of object categories, action effects and logical rules: From continuous manipulative exploration to symbolic planning," pp. 119–139, 2015.

[16] S. Höfer and O. Brock, "Coupled learning of action parameters and forward models for manipulation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3893–3899.

[17] V. Sarathy and M. Scheutz, "Beyond Grasping - Perceiving Affordances Across Various Stages of Cognitive Development," in *IEEE International Conference on Developmental Learning and Epigenetic Robotics*, 2016.

[18] A. Dehban, L. Jamone, A. R. Kampff, and J. Santos-Victor, "Denoising auto-encoders for learning of objects and tools affordances in continuous space," in *IEEE International Conference on Robotics and Automation*.

[19] R. O. Chavez-Garcia, P. Luce-Vayrac, and R. Chatila, "Discovering affordances through perception and manipulation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3959–3964.

[20] A. Romay, S. Kohlbrecher, A. Stumpf, O. von Stryk, S. Maniatopoulos, H. Kress-Gazit, P. Schillinger, and D. C. Conner, "Collaborative autonomy between high-level behaviors and human operators for remote manipulation tasks using different humanoid robots," *Journal of Field Robotics*, vol. 34, no. 2, pp. 333–358, 2017.

[21] S. Hart, P. Dinh, and K. Hambuchen, "The affordance template ROS package for robot task programming," in *IEEE International Conference on Robotics and Automation*, pp. 6227–6234.

[22] M. Fallon, S. Kuindersma, S. Karumanchi, M. Antone, T. Schneider, H. Dai, C. Pérez D'Arpino, R. Deits, M. DiCicco, D. Fourie, T. Koolen, P. Marion, M. Posa, A. Valenzuela, K.-T. Yu, J. Shah, K. Iagnemma, R. Tedrake, and S. Teller, "An architecture for online affordance-based perception and whole-body planning," *Journal of Field Robotics*, vol. 32, no. 2, pp. 229–254, 2015.

[23] W. Pryor, Y.-C. Lin, and D. Berenson, "Integrated affordance detection and humanoid locomotion planning," in *IEEE/RAS International Conference on Humanoid Robots*, pp. 125–131.

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

[25] A. Jøsang, "A logic for uncertain probabilities," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, pp. 279–311.

[26] R. Detry, D. Kraft, O. Kroemer, L. Bodenhagen, J. Peters, N. Krüger, and J. Piater, "Learning Grasp Affordance Densities," *Paladyn, Journal of Behavioral Robotics*, vol. 2, no. 1, pp. 1–17.