Towards a Hierarchy of Loco-Manipulation Affordances

Peter Kaiser, Eren E. Aksoy, Markus Grotz and Tamim Asfour

***Abstract*— We propose a formalism for the hierarchical representation of affordances. Starting with a perceived model of the environment consisting of geometric primitives like planes or cylinders, we define a hierarchical system for affordance extraction whose foundation are elementary power grasp affordances. Higher-level affordances, e.g. bimanual affordances, result from combining lower-level affordances with additional properties concerning the underlying geometric primitives of the scene. We model affordances as continuous certainty functions taking into account properties of the environmental elements and the perceiving robot's embodiment. The developed formalism is regarded as the basis for the description of whole-body affordances, i.e. affordances associated with whole-body actions. The proposed formalism was implemented and experimentally evaluated in multiple scenarios based on RGB-D camera data. The feasibility of the approach is demonstrated on a real robotic platform.**

I. Introduction

Humanoid robots are intended to operate in unstructured, human-centered environments that are not specifically designed for robot interaction. Robots that are exposed to such environments need to employ flexible perceptual mechanisms for identifying possible ways of interaction with the environment. Such interactions between the robot and environmental objects include whole-body actions like i) stepping on and over obstacles, ii) the manipulation of large objects or iii) the utilization of environmental objects to support balance. With this work we aim at creating the perceptual basis for the detection of whole-body affordances in a loco-manipulation context, i.e. affordances for actions that incorporate the whole body for locomotion and/or manipulation purposes.

The psychological concept of *affordances*, originally introduced by Gibson [1] as an approach to understand the human perceptual process, states that agents perceive action possibilities latent in the environment with respect to their own action capabilities. Affordance-based approaches have been widely applied in robotics, especially in the areas of grasping and manipulation, human-robot interaction, and locomotion and navigation. A comprehensive overview is found in [2]. Krüger et al. proposed the idea of coupling objects and actions to combined representations of sensorimotor experience, termed *Object-Action Complexes* (OACs) [3]. The affordance extraction process proposed in this work could provide the perceptual preconditions for the instantiation of OACs.

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@kit.edu

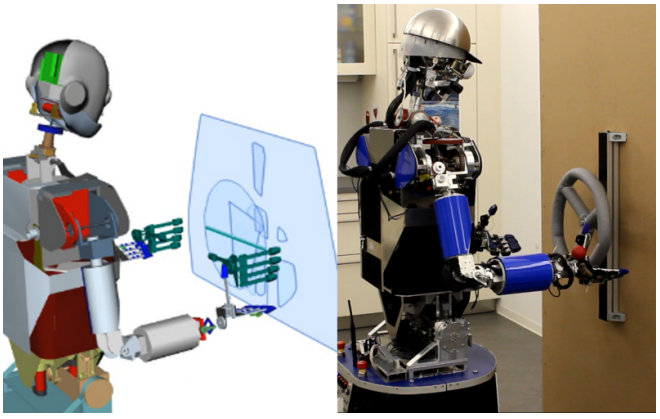

Fig. 1: The humanoid robot ARMAR-III confronted with a valve: It needs to identify environmental objects that afford *turning* and subsequently determine suitable end-effector poses for turning the chosen object. The perceived primitives are visualized in blue on the left side. The extracted bimanual turn affordance is shown in green including the proposed end-effector poses.

Works like [4] indicate that many of the aspired whole-body actions can be differentiated on the lowest level by fundamental grasp affordances. The authors also show that this is not generally true, e.g. in the case of *sitting*. However, we think that the research on whole-body locomotion and manipulation can greatly benefit from a deeper understanding of the perception of actions that establish stabilizing contacts by grasping. Hence, we propose a hierarchical approach to affordance perception based on elementary power grasp affordances.

Many of the teams participating in the DARPA Robotics Challenge (DRC) Finals in 2015 pursued an affordance-driven approach to manipulation that was supervised by a human operator (e.g. [5], [6], [7]). In contrast to these works based on object templates, we assume unknown environments. The methods proposed in this work could help increasing the level of autonomy for tasks similar to the ones from the DRC. Reducing the degree of human intervention necessary to operate a humanoid robot in complex environments will be a valuable step towards the development of fully autonomous humanoid robots. A survey on the degree of autonomy used by the DRC teams can be found in [8].

In our previous works ([9], [10], [11]) we proposed a general perceptual pipeline for the extraction of environmental primitives like planes, cylinders and spheres from RGB-D point clouds. In this work we extend our previous rule-based approach for affordance extraction to a hierarchical

formalism based on a set of fundamental grasp affordances. The affordance extraction pipeline is bootstrapped with a part-based object segmentation method [12] that leads to over-segmentation of the perceived scene. Part-based object segmentation is intensively used in robotics and computer vision for the task of affordance extraction, for instance, from geometric features [13]. Our perceptional pipeline differs from previous works, such as [14], [15], in employing geometric features for categorizing scene segments as planes, cylinders, or spheres in an iterative manner which leads to dense primitive exploration.

The method proposed in this work eventually produces a set of affordance certainty functions based on the set of primitives detected in the perceived environment. Our goal is to use this information as a basis for task planning (e.g. [16], [17]), whole-body action planning with contacts (e.g. [18], [19]) or whole-body contact-based control (e.g. [20]).

The remainder of this paper is structured as follows: Section II defines fundamental unimanual grasp affordances that form the basis of the hierarchical structure for affordance extraction (Layer **L0**). Section III discusses the extension of the set of fundamental grasp affordances towards higher-level affordances of bimanual whole-body manipulation (Layers **L1**-**L4**). Section IV discusses several examples for extracted affordances based on captured RGB-D data. Finally, Section V concludes the paper and discusses future work.

II. Fundamental Affordances

The proposed hierarchical affordance formalism relies on a set of fundamental grasp affordances. These affordances are detected based on a simplified environmental representation in terms of environmental primitives as introduced in our previous work in [10] and [11]. An exemplary scene with detected environmental primitives is depicted in Fig. 1. In the following, we will further discuss the formalization of environmental primitives and affordances used throughout this paper.

A. Environmental Primitives

Let $\Pi = \{p_1, \dots, p_k\}$ denote a set of environmental. Each primitive p_i provides information concerning its shape, orientation and extent. This information will later be fed into different layers of the affordance hierarchy.

1) Shape Functions: We define a set of shape functions that determine the degree to which p_i belongs to an associated shape class. Possible shape functions, matching the current capabilities of our perceptual pipeline, are:

$$\begin{aligned} &\text{planar}(p_i) \in [0,1] \\ &\text{circular}(p_i) \in [0,1] \\ &\text{spherical}(p_i) \in [0,1] \\ &\text{cylindrical}(p_i) \in [0,1] \end{aligned} \tag{1}$$

The extension of the system to further shape classes is possible and straightforward.

2) Distance Function: The distance function $d(p_i, \boldsymbol{x}, \boldsymbol{v})$ describes the extent of the primitive at the point $\boldsymbol{x} \in \mathbb{R}^3$ in direction $\boldsymbol{v} \in \mathbb{R}^3$, $\|\boldsymbol{v}\| = 1$ (see Fig. 2):

$$d(p_i, \boldsymbol{x}, \boldsymbol{v}) = \max \left\{ \lambda \in \mathbb{R}^+ : \boldsymbol{x} \pm \frac{\lambda}{2} \boldsymbol{v} \in p_i \right\} \tag{2}$$

Both, $\boldsymbol{x}$ and $\boldsymbol{v}$ relate to the primitive's local coordinate frame. For simplification, we will write $d_x(p_i, \boldsymbol{x}) := d(p_i, \boldsymbol{x}, \mathbf{1}_x)$ and similarly for the other axes $\mathbf{1}_y$ and $\mathbf{1}_z$.

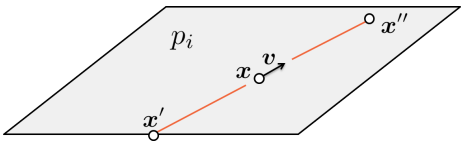

Fig. 2: $d(p_i, \boldsymbol{x}, \boldsymbol{v})$ describes the length of the longest possible symmetric line, i.e. $\|\boldsymbol{x} - \boldsymbol{x}'\| = \|\boldsymbol{x} - \boldsymbol{x}''\|$ in the direction of $\boldsymbol{v}$ that lies entirely inside p_i, i.e. $\boldsymbol{x}', \boldsymbol{x}'' \in p_i$. In this case: $d(p_i, \boldsymbol{x}, \boldsymbol{v}) = \|\boldsymbol{x}' - \boldsymbol{x}''\|$.

3) Orientation Function: The orientation function $\text{up}(p_i)$ describes the orientation of p_i with respect to the global up-vector $\boldsymbol{u}$[1]. It computes the angle between the primitive's normal vector $\boldsymbol{n}(p_i)$ and $\boldsymbol{u}$:

$$\text{up}(p_i) = \arccos\left(\frac{\boldsymbol{n}(p_i) \cdot \boldsymbol{u}}{\|\boldsymbol{n}(p_i)\| \cdot \|\boldsymbol{u}\|} \right) \in [0, \pi] \tag{3}$$

B. Affordance Certainty Functions

An affordance as understood in our approach is a function Θ defined over the Cartesian product of a space $\mathcal{S}$ of end-effector poses and the set Π of environmental primitives:

$$\Theta_a : \Pi \times \mathcal{S} \rightarrow [0,1] \tag{4}$$

$\mathcal{S}$ defines the underlying representation of end-effector poses relevant to the affordance. It will be chosen as $\mathcal{S} = SE(3)$ for unimanual affordances and as $\mathcal{S} = SE(3) \times SE(3)$ for bimanual affordances[2]. The certainty function $\Theta_a(p, \boldsymbol{e})$ for an affordance a describes how certain the perceptual system is about the existence of a for a given end-effector pose $\boldsymbol{e} \in \mathcal{S}$. Mathematical operations can be applied to combine certainty functions for different affordances in order to construct joint certainties for higher-level affordances. Subsequently, Θ_a can be used in order to propose end-effector poses based on the existence certainty of the corresponding affordance a.

C. Fundamental Grasp Affordances

The most fundamental affordance for whole-body loco-manipulation is *grasping*, while grasping in this context is understood as bringing an end-effector in contact with an environmental primitive. Grasp taxonomies, e.g. [21], usually differ between *precision grasps* and *power grasps*. Power

[1] In our implementation: $\boldsymbol{u} = \mathbf{1}_z$

[2] $SE(3)$ denotes the special euclidean group. We refer to $\boldsymbol{e} \in SE(3)$ as a homogeneous matrix $\boldsymbol{e} \in \mathbb{R}^{4\times4}$. The rotational and translational components of $\boldsymbol{e}$ will be denoted as $t(\boldsymbol{e}) \in \mathbb{R}^3$ and $R(\boldsymbol{e}) \in \mathbb{R}^{3\times3}$.

Fig. 3: The two grasp types considered throughout this work: a *prismatic grasp* (left) and a *platform grasp* (right).

grasps are specifically important for establishing reliable contacts with environmental objects.

Works like [22] and [23] show that humans indeed heavily rely on power grasps when performing tasks of daily living. According to these works the predominant power grasp types are the prismatic grasp and the circular grasp. However, in our application that is primarily focused on stabilization, we believe that the prismatic grasp together with the platform grasp (see Fig. 3) are suitable for implementing basic behaviors of whole-body loco-manipulation. Extension to further power grasp types as well as to precision grasp types for dexterous manipulation is possible and planned in our future work.

The existence of grasp affordances in the environment depends on the dimensions of environmental objects with respect to the perceiving agent's embodiment. Fig. 4 depicts the end-effector parameters considered for the extraction of grasp affordances. Table I lists possible values for these parameters for the embodiments of an average human and the humanoid robots ARMAR-III [24] and ARMAR-4 [25].

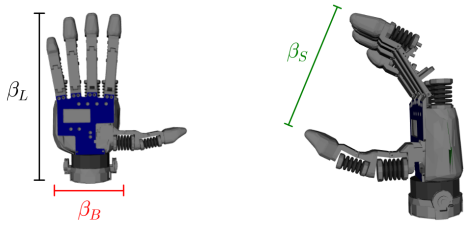

Fig. 4: The body-scaled parameters β_L, β_B and β_S as foundation for perceiving grasp affordances. The parameters refer to *Hand Length*, *Hand Breadth* and *Hand Span*, respectively, as defined in [26].

TABLE I: Comparison of body-scaled parameters for an average human and the humanoid robots ARMAR-III and ARMAR-4 in cm.

Parameter	Sym.	Human[3]	ARMAR-III	ARMAR-4
Hand Length	β_L	19.71	17.0	16.0
Hand Breadth	β_B	8.97	10.0	6.5
Hand Span	β_S	12.42	13.0	10.0
Shoulder Length	β_{Sh}	$0.258H$	40.0	40.0

The main building block of affordance certainty functions are threshold-based decision functions applied to properties of environmental primitives. In this work we employ a sigmoid function as a continuous version of such a decision function:

$$\text{sigm}_{\lambda,\beta}(x) = \frac{1}{1+e^{-\lambda(x-\beta)}} \in (0,1) \tag{5}$$

In Fig. 5 a visualization of $\text{sigm}_{\lambda,\beta}(x)$ and two of its variations is shown.

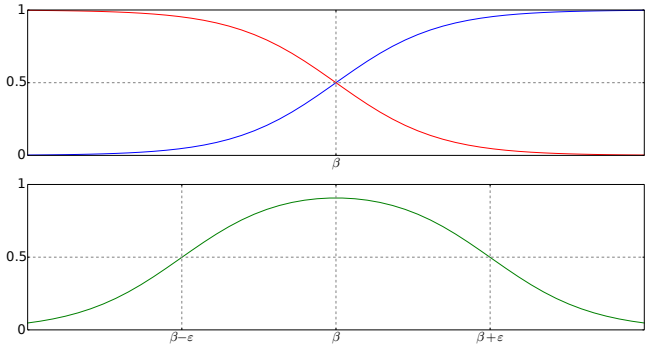

Fig. 5: The sigmoid functions $\text{sigm}_{\lambda,\beta}(x)$ (**blue**) and $\text{sigm}_{-\lambda,\beta}(x)$ (**red**). The bottom plot displays a sigmoid-based interval function $\text{sigm}_{\lambda,\beta-\varepsilon}(x) \cdot \text{sigm}_{-\lambda,\beta+\varepsilon}(x)$ (**green**).

There are two predominant types of inputs to the sigmoid function: translations $t \in \mathbb{R}$ in *mm* and rotations $r \in [0, 2\pi)$ in *radians*. The parameter λ of the sigmoid function can be fixed with respect to the input type, resulting in two instances of the decision function Γ^t and Γ^r[4]:

$$\begin{aligned} \Gamma^t_\beta(x) &= \text{sigm}_{\lambda_t,\beta}(x) \\ \Gamma^r_\beta(x) &= \text{sigm}_{\lambda_r,\beta}(x) \end{aligned} \tag{6}$$

For the matter of simplicity we also define the complements $\bar{\Gamma}^t_\beta$ and $\bar{\Gamma}^r_\beta$:

$$\begin{aligned} \bar{\Gamma}^t_\beta(x) &= 1 - \Gamma^t_\beta(x) = \text{sigm}_{-\lambda_t,\beta}(x) \\ \bar{\Gamma}^r_\beta(x) &= 1 - \Gamma^r_\beta(x) = \text{sigm}_{-\lambda_r,\beta}(x) \end{aligned} \tag{7}$$

Finally, we define an interval function based on the above sigmoids, which will be a common pattern in the affordances developed later:

$$\begin{aligned} \Delta^t_{\beta,\varepsilon}(x) &= \Gamma^t_{\beta-\varepsilon}(x) \cdot \bar{\Gamma}^t_{\beta+\varepsilon}(x) \\ \Delta^r_{\beta,\varepsilon}(x) &= \Gamma^r_{\beta-\varepsilon}(x) \cdot \bar{\Gamma}^r_{\beta+\varepsilon}(x) \end{aligned} \tag{8}$$

Now we can express the fundamental power grasp affordances in terms of certainty functions Θ composed of the defined sigmoid-based decision functions Γ, $\bar{\Gamma}$ and Δ. Note that $\bar{\Gamma}$ and Δ are essentially shortcuts for products of differently parameterized Γ-functions.

[3]The hand measures for the human embodiment in Table I refer to the hand of an average male adult [26]. The shoulder length is given relative to the body height H as described in [27].

[4]In our implementation: $\lambda_t = 1, \lambda_r = 20$

1) Platform Grasp: For modeling platform grasp affordances (**G-Pl**), we consider the body-scaled parameters for the hand length β_L and hand breadth β_B (see Fig. 4 and Table I):

$$\Theta_{\text{G-Pl}}(p, \boldsymbol{e}) = \Gamma^t_{\beta_B}(d_x(p, \boldsymbol{e})) \cdot \Gamma^t_{\beta_L}(d_y(p, \boldsymbol{e})) \tag{9}$$

The above equation states that a platform grasp is applicable to a primitive p at an end-effector pose $\boldsymbol{e}$ if p is large enough to fit a bounding box of the end-effector's dimensions around $\boldsymbol{e}$. The axes x and y refer to the local end-effector coordinate systems as shown in Fig. 6.

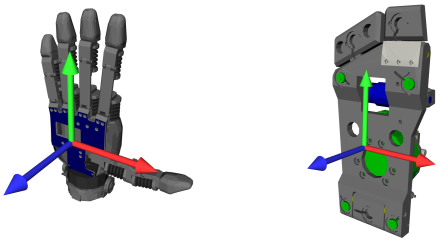

Fig. 6: The assumed end-effector coordinate systems exemplary for a hand and a foot of ARMAR-4. The axis z points into grasp direction (blue) and y points into the direction of the longest end-effector extent (green).

2) Prismatic Grasp: For modeling prismatic grasp affordances (**G-Pr**), we consider the body-scaled parameters for the hand span β_S and hand breadth β_B (see Fig. 4 and Table I):

$$\Theta_{\text{G-Pr}}(p, \boldsymbol{e}) = \Gamma^t_{\beta_B}(d_x(p, \boldsymbol{e})) \cdot \bar{\Gamma}^t_{\beta_S}(d_y(p, \boldsymbol{e})) \tag{10}$$

The above equation states that a prismatic grasp is applicable to a primitive p at an end-effector pose $\boldsymbol{e}$ if p is large enough in x-direction to fit the hand breadth and small enough in y-direction to fit into the open hand.

Note that the formalization of the fundamental grasp affordances discussed in this section does not consider reachability, stability or similar indices. Such information can be included in selection strategies based on the set of extracted affordances.

III. Affordance Hierarchy

In this section we will define higher-level affordances based on the previously discussed fundamental grasp affordances $\Theta_{\text{G-Pl}}$ and $\Theta_{\text{G-Pr}}$ (**L0**-Affordances). Higher-level affordances result from combining lower-level affordances with additional environmental properties or from combining multiple lower-level affordances. In the following we will discuss higher-level unimanual and bimanual affordances.

A. Unimanual Affordances (*L1*)

The fundamental unimanual grasp affordances $\Theta_{\text{G-Pl}}$ and $\Theta_{\text{G-Pr}}$ can be combined with additional environmental properties to form higher-level unimanual affordances. For example, the unimanual affordances *lean* (**Ln**) and *support*

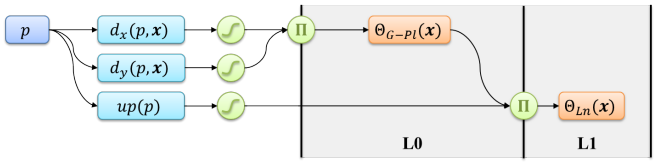

Fig. 7: Properties of the environmental primitive p pass sigmoid decision functions. The results are multiplied in order to form certainties for fundamental affordances, in this case for a platform grasp affordance. Affordance certainty functions can participate in forming certainties for more abstract affordances. The colors refer to environmental primitives (purple), their properties (blue), mathematical operations **(green)** and affordance certainties (orange).

(**Sp**) result from the certainty function for a platform grasp affordance in combination with the degree of horizontality of the underlying primitive p:

$$\Theta_{\text{Ln}}(p, \boldsymbol{e}) = \Theta_{\text{G-Pl}}(p, \boldsymbol{e}) \cdot \Delta^r_{\pi,\varepsilon}(\text{up}(p)) \tag{11}$$

$$\Theta_{\text{Sp}}(p, \boldsymbol{e}) = \Theta_{\text{G-Pl}}(p, \boldsymbol{e}) \cdot \Delta^r_{0,\varepsilon}(\text{up}(p)) \tag{12}$$

Fig. 7 shows a visualization of the computational process behind the certainty function Θ_{Ln}. The unimanual affordances discussed in this section appear in the second layer (**L1**) of the affordance hierarchy. Table II defines further higher-level unimanual affordances.

B. Bimanual Affordances (*L2-L4*)

The extraction of bimanual affordances based on their unimanual counterparts is one of the major contributions of the hierarchical affordance framework. Note that for bimanual affordances the underlying space $\mathcal{S}$ of end-effector poses changes from $SE(3)$ to $SE(3) \times SE(3)$ to account for the availability of two end-effectors. The new end-effector pose space $\mathcal{S}$ allows further environmental properties to be considered in the affordance extraction process, properties that put the two end-effector poses into relation:

1) Distance: The distance $d(\boldsymbol{e}_1, \boldsymbol{e}_2)$ between the two end-effector poses $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$:

$$d(\boldsymbol{e}_1, \boldsymbol{e}_2) = \|t(\boldsymbol{e}_1) - t(\boldsymbol{e}_2)\| \in \mathbb{R}^+ \tag{13}$$

2) Angle: The angle $\alpha(\boldsymbol{e}_1, \boldsymbol{e}_2)$ between the two end-effector poses $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$, assuming the grasp direction $R(\boldsymbol{e}_*) \cdot \mathbf{1}_y$ is the same:

$$\alpha(\boldsymbol{e}_1, \boldsymbol{e}_2) = \arccos\Big(\big(R(\boldsymbol{e}_1) \cdot \mathbf{1}_x\big) \cdot \big(R(\boldsymbol{e}_2) \cdot \mathbf{1}_x\big)\Big) \tag{14}$$

3) Relative Orientation: The relative orientation $\text{up}(\boldsymbol{e}_1, \boldsymbol{e}_2)$ of the two end-effector poses $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ with respect to the global up-vector $\boldsymbol{u}$ (see Eq. 3).

$$\text{up}(\boldsymbol{e}_1, \boldsymbol{e}_2) = \arccos\left(\frac{(t(\boldsymbol{e}_1) - t(\boldsymbol{e}_2)) \cdot \boldsymbol{u}}{\|(t(\boldsymbol{e}_1) - t(\boldsymbol{e}_2))\|}\right) \tag{15}$$

The elementary bimanual grasp types **Bi-G-Pl** and **Bi-G-Pr** can be defined by evaluating the individual unimanual

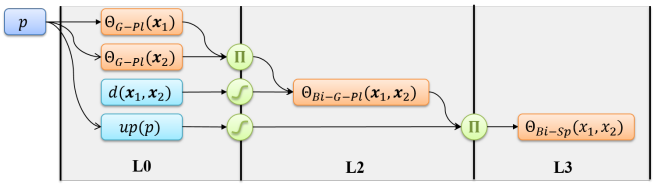

Fig. 8: Bimanual affordance certainty functions $\Theta_{\text{Bi-}*}$ are combinations of the underlying unimanual affordance certainty functions evaluated at the respective end-effector poses $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$. Additional properties can be taken into consideration that put $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ into relation, in this case $d(\boldsymbol{e}_1, \boldsymbol{e}_2)$ for their distance.

affordance certainty functions. In addition, we impose constraints on the grasp distance which is constrained by the hand length β_L to the lower end and by the shoulder distance β_{Sh} to the upper end, e.g.:

$$\begin{aligned}\Theta_{\text{Bi-G-Pl}}(p, \boldsymbol{e}_1, \boldsymbol{e}_2) = \Theta_{\text{G-Pl}}(p, \boldsymbol{e}_1) \cdot \Theta_{\text{G-Pl}}(p, \boldsymbol{e}_2) \\ \cdot \Gamma^t_{\beta_{\text{L}}}(d(\boldsymbol{e}_1, \boldsymbol{e}_2)) \cdot \bar{\Gamma}^t_{\beta_{\text{Sh}}}(d(\boldsymbol{e}_1, \boldsymbol{e}_2))\end{aligned} \tag{16}$$

Based on these fundamental bimanual grasping affordances (layer **L2** in Table II), we differentiate affordances for four further bimanual grasp types that are directly usable for executing high-level bimanual skills. These bimanual grasp types are differed by considering the underlying unimanual grasp type (as in Eq. 16) as well as the mutual orientation α of the end-effector grasp poses. We define an aligned type ($\alpha = 0$) and an opposing type ($\alpha = \pi$) of each of the bimanual grasp types defined above, e.g.:

$$\begin{aligned}\Theta_{\text{Bi-G-Pl-Al}}(p, \boldsymbol{e}_1, \boldsymbol{e}_2) = \Theta_{\text{Bi-G-Pl}}(p, \boldsymbol{e}_1, \boldsymbol{e}_2) \\ \cdot \Delta^r_{0,\varepsilon}(\text{up}(\boldsymbol{e}_1, \boldsymbol{e}_2)) \\ \cdot \Delta^r_{0,\varepsilon}(\alpha(\boldsymbol{e}_1, \boldsymbol{e}_2))\end{aligned} \tag{17}$$

The remaining three bimanual grasp types can be found in Table II (layer **L3**). As an example for a higher-level bimanual affordance, we define a *bimanual support* affordance (**Bi-Sp**) in a similar way as **Sp**, just replacing the underlying grasp affordance (see Fig. 8):

$$\Theta_{\text{Bi-Sp}}(p, \boldsymbol{e}) = \Theta_{\text{Bi-G-Pl}}(p, \boldsymbol{e}_1, \boldsymbol{e}_2) \cdot \Delta^r_{0,\varepsilon}(\text{up}(p)) \tag{18}$$

Table II and Fig. 9 show the full affordance hierarchy as defined in this work. This hierarchy is not to be understood as complete, it's more an incrementally growing hierarchy for the affordance exploration process in our experimental setups.

IV. Experimental Evaluation

For making the process of affordance extraction computationally tractable, we pursue a sampling-based approach. We reduce the space $\mathcal{S}$ to the set of poses that lie on the boundary of at least one of the available primitives. For example, in the case of $\mathcal{S} = SE(3)$:

$$\mathcal{S}_{\text{red}} = \left\{\boldsymbol{e} \in SE(3) : \exists p \in \Pi, \boldsymbol{e} \in \partial p\right\} \tag{19}$$

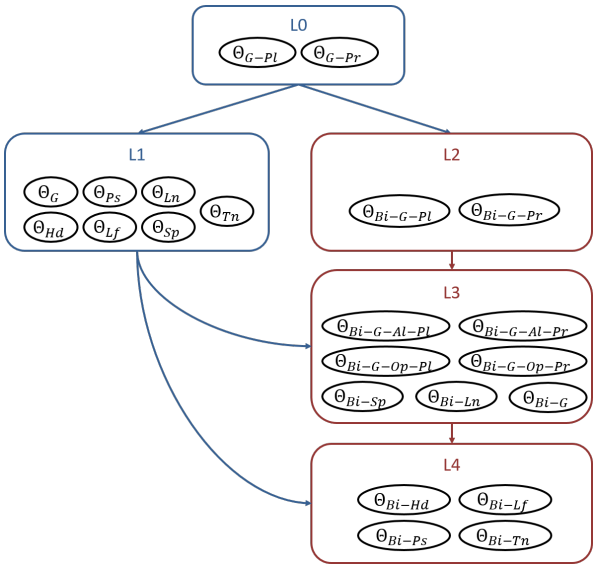

Fig. 9: The hierarchical structure induced by the affordance definitions from Table II. Layers that contain unimanual affordances are drawn in blue while layers with bimanual affordances are drawn in red.

while ∂p denotes the boundary of the primitive p. This reduction works in a similar way if $\mathcal{S} = SE(3) \times SE(3)$. We now define all certainty functions Θ_* to evaluate to zero for inputs $\boldsymbol{e} \notin \mathcal{S}_{\text{red}}$. The rationale behind this is that all affordances defined in Table II are based on fundamental grasp affordances, i.e. they all require the end-effectors to be in contact with one of the available primitives. By definition of the fundamental grasp affordances, suitable contact with a primitive p only occurs if $\boldsymbol{e} \in \partial p$. This reduced space $\mathcal{S}_{\text{red}}$ can efficiently be sampled using spatial and orientational discretization step sizes Δx and $\Delta\varphi$.[5]

In the following we will evaluate our approach to affordance extraction based on three exemplary scenes from the area of whole-body loco-manipulation. One of these examples, the turning of an industrial valve, has been implemented and executed on a real robotic platform.[6]

A. Example I: Staircase

Fig. 10 visualizes the affordance certainty functions $\Theta_{\text{G-Pl}}$ and $\Theta_{\text{G-Pr}}$ computed for a perceived staircase. The example demonstrates the capability of the perceptual pipeline to generate a reasonable segmentation of the scene into environmental primitives as well as the successful extraction of certainty functions for various affordances. It also shows that the formalization of the proposed affordance hierarchy can, to some degree, handle perceptual inaccuracies, e.g. it properly assigns high certainties for prismatic grasps along the whole handrail although the perceptual process

[5]Throughout this work, we used discretization step sizes of $\Delta x = 3\,\text{cm}$ and $\Delta\varphi = \frac{\pi}{2}$ rad

[6]All three experiments have been performed using the embodiment of ARMAR-III (see Table I).

TABLE II: Affordance Hierarchy

	Label	Certainty Function	Expression $\Theta_*(p, \boldsymbol{e})$ or $\Theta_*(p, \boldsymbol{e}_1, \boldsymbol{e}_2)$	Description
L0	G-Pl	$\Theta_{\text{G-Pl}}(\boldsymbol{e})$	$\Gamma^t_{\beta_R}(d_x(p,\boldsymbol{e})) \cdot \Gamma^t_{\beta_{I'}}(d_y(p,\boldsymbol{e}))$	*Platform grasp*
	G-Pr	$\Theta_{\text{G-Pr}}(\boldsymbol{e})$	$\Gamma^t_{\beta_R}(d_x(p,\boldsymbol{e})) \cdot \bar{\Gamma}^t_{\beta_S}(d_y(p,\boldsymbol{e}))$	*Prismatic grasp*
L1	G	$\Theta_{\text{G}}(\boldsymbol{e})$	$\max\{\Theta_{\text{G-Pl}}(p,\boldsymbol{e}), \Theta_{\text{G-Pr}}(p,\boldsymbol{e})\}$	*Grasp (platform or prismatic)*
	Sp	$\Theta_{\text{Sp}}(\boldsymbol{e})$	$\Theta_{\text{G-Pl}}(p,\boldsymbol{e}) \cdot \Delta^r_{0,\varepsilon}(\text{up}(p))$	*Support*
	Ln	$\Theta_{\text{Ln}}(\boldsymbol{e})$	$\Theta_{\text{G-Pl}}(p,\boldsymbol{e}) \cdot \Delta^r_{\pi,\varepsilon}(\text{up}(p))$	*Lean*
	Hd	$\Theta_{\text{Hd}}(\boldsymbol{e})$	$\Theta_{\text{G-Pr}}(p,\boldsymbol{e}) \cdot \Gamma^t_{\lambda_1}(d_x(p,\boldsymbol{e}))$	*Hold*
	Lf	$\Theta_{\text{Lf}}(\boldsymbol{e})$	$\Theta_{\text{G-Pr}}(p,\boldsymbol{e}) \cdot \bar{\Gamma}^t_{\lambda_2}(d_x(p,\boldsymbol{e})) \cdot \bar{\Gamma}^t_{\lambda_3}(d_z(p,\boldsymbol{e}))$	*Lift*
	Ps	$\Theta_{\text{Ps}}(\boldsymbol{e})$	$\Theta_{\text{G-Pl}}(p,\boldsymbol{e}) \cdot \bar{\Gamma}^t_{\lambda_4}(d_x(p,\boldsymbol{e})) \cdot \bar{\Gamma}^t_{\lambda_5}(d_y(p,\boldsymbol{e}))$	*Push*
	Tn	$\Theta_{\text{Tn}}(\boldsymbol{e})$	$\Theta_{\text{G-Pr}}(p,\boldsymbol{e}) \cdot \text{circular}(p)$	*Turn*
L2	Bi-G-Pl	$\Theta_{\text{Bi-G-Pl}}(\boldsymbol{e}_1,\boldsymbol{e}_2)$	$\Theta_{\text{G-Pl}}(p,\boldsymbol{e}_1) \cdot \Theta_{\text{G-Pl}}(p,\boldsymbol{e}_2) \cdot \Gamma^t_{\beta_{\text{L}}}(d(\boldsymbol{e}_1,\boldsymbol{e}_2)) \cdot \bar{\Gamma}^t_{\beta_{\text{Sh}}}(d(\boldsymbol{e}_1,\boldsymbol{e}_2))$	*Bimanual platform grasp*
	Bi-G-Pr	$\Theta_{\text{Bi-G-Pr}}(\boldsymbol{e}_1,\boldsymbol{e}_2)$	$\Theta_{\text{G-Pr}}(p,\boldsymbol{e}_1) \cdot \Theta_{\text{G-Pr}}(p,\boldsymbol{e}_2) \cdot \Gamma^t_{\beta_{\text{L}}}(d(\boldsymbol{e}_1,\boldsymbol{e}_2)) \cdot \bar{\Gamma}^t_{\beta_{\text{Sh}}}(d(\boldsymbol{e}_1,\boldsymbol{e}_2))$	*Bimanual prismatic grasp*
L3	Bi-G	$\Theta_{\text{Bi-G}}(\boldsymbol{e}_1,\boldsymbol{e}_2)$	$\max\{\Theta_{\text{Bi-G-Pl}}(\boldsymbol{e}_1,\boldsymbol{e}_2), \Theta_{\text{Bi-G-Pr}}(\boldsymbol{e}_1,\boldsymbol{e}_2)\}$	*Bimanual grasp (platform or prism.)*
	Bi-G-Al-Pl	$\Theta_{\text{Bi-G-Al-Pl}}(\boldsymbol{e}_1,\boldsymbol{e}_2)$	$\Theta_{\text{Bi-G-Pl}}(p,\boldsymbol{e}_1,\boldsymbol{e}_2) \cdot \Delta^r_{0,\varepsilon}(\text{up}(\boldsymbol{e}_1,\boldsymbol{e}_2)) \cdot \Delta^r_{0,\varepsilon}(\alpha(\boldsymbol{e}_1,\boldsymbol{e}_2))$	*Bimanual aligned platform grasp*
	Bi-G-Al-Pr	$\Theta_{\text{Bi-G-Al-Pr}}(\boldsymbol{e}_1,\boldsymbol{e}_2)$	$\Theta_{\text{Bi-G-Pr}}(p,\boldsymbol{e}_1,\boldsymbol{e}_2) \cdot \Delta^r_{0,\varepsilon}(\text{up}(\boldsymbol{e}_1,\boldsymbol{e}_2)) \cdot \Delta^r_{0,\varepsilon}(\alpha(\boldsymbol{e}_1,\boldsymbol{e}_2))$	*Bimanual aligned prismatic grasp*
	Bi-G-Op-Pl	$\Theta_{\text{Bi-G-Op-Pl}}(\boldsymbol{e}_1,\boldsymbol{e}_2)$	$\Theta_{\text{Bi-G-Pl}}(p,\boldsymbol{e}_1,\boldsymbol{e}_2) \cdot \Delta^r_{0,\varepsilon}(\text{up}(\boldsymbol{e}_1,\boldsymbol{e}_2)) \cdot \Delta^r_{\pi,\varepsilon}(\alpha(\boldsymbol{e}_1,\boldsymbol{e}_2))$	*Bimanual opposed platform grasp*
	Bi-G-Op-Pr	$\Theta_{\text{Bi-G-Op-Pr}}(\boldsymbol{e}_1,\boldsymbol{e}_2)$	$\Theta_{\text{Bi-G-Pr}}(p,\boldsymbol{e}_1,\boldsymbol{e}_2) \cdot \Delta^r_{0,\varepsilon}(\text{up}(\boldsymbol{e}_1,\boldsymbol{e}_2)) \cdot \Delta^r_{\pi,\varepsilon}(\alpha(\boldsymbol{e}_1,\boldsymbol{e}_2))$	*Bimanual opposed prismatic grasp*
	Bi-Sp	$\Theta_{\text{Bi-Sp}}(\boldsymbol{e}_1,\boldsymbol{e}_2)$	$\Theta_{\text{Bi-G-Pl}}(p,\boldsymbol{e}_1,\boldsymbol{e}_2) \cdot \Delta^r_{0,\varepsilon}(\text{up}(p))$	*Bimanual support*
	Bi-Ln	$\Theta_{\text{Bi-Ln}}(\boldsymbol{e}_1,\boldsymbol{e}_2)$	$\Theta_{\text{Bi-G-Pl}}(p,\boldsymbol{e}_1,\boldsymbol{e}_2) \cdot \Delta^r_{\pi,\varepsilon}(\text{up}(p))$	*Bimanual lean*
L4	Bi-Hd	$\Theta_{\text{Bi-Hd}}(\boldsymbol{e}_1,\boldsymbol{e}_2)$	$\Theta_{\text{Bi-G-Al-Pr}}(p,\boldsymbol{e}_1,\boldsymbol{e}_2) \cdot \Theta_{\text{Hd}}(\boldsymbol{e}_1) \cdot \Theta_{\text{Hd}}(\boldsymbol{e}_2)$	*Bimanual hold*
	Bi-Lf	$\Theta_{\text{Bi-Lf}}(\boldsymbol{e}_1,\boldsymbol{e}_2)$	$\Theta_{\text{Bi-G-Al-Pr}}(p,\boldsymbol{e}_1,\boldsymbol{e}_2) \cdot \Theta_{\text{Lf}}(\boldsymbol{e}_1) \cdot \Theta_{\text{Lf}}(\boldsymbol{e}_2)$	*Bimanual lift*
	Bi-Ps	$\Theta_{\text{Bi-Ps}}(\boldsymbol{e}_1,\boldsymbol{e}_2)$	$\Theta_{\text{Bi-G-Al-Pl}}(p,\boldsymbol{e}_1,\boldsymbol{e}_2) \cdot \Theta_{\text{Ps}}(\boldsymbol{e}_1) \cdot \Theta_{\text{Ps}}(\boldsymbol{e}_2)$	*Bimanual push*
	Bi-Tn	$\Theta_{\text{Bi-Tn}}(\boldsymbol{e}_1,\boldsymbol{e}_2)$	$\Theta_{\text{Bi-G-Op-Pr}}(p,\boldsymbol{e}_1,\boldsymbol{e}_2) \cdot \text{circular}(p)$	*Bimanual turn*

$\varepsilon = \frac{\pi}{8}$, $\lambda_1, \ldots, \lambda_5$ implementation-specific constants. For example, λ_2 and λ_3 characterize the maximum dimensions liftable objects.

segmented the handrail into several primitives of different types.

B. Example II: Ladder

Fig. 11 shows the affordance certainty functions $\Theta_{\text{G-Pr}}$, Θ_{Sp} and $\Theta_{\text{Bi-G-Op-Pr}}$ computed for a perceived ladder. Bimanual certainty functions are visualized by connecting the respective end-effector poses with a line colored according to the certainty value. For clarity we applied a certainty threshold of 0.7 in the visualization, reducing the amount of bimanual configurations displayed.

The resulting certainty functions $\Theta_{\text{Bi-G-Op-Pr}}$ and Θ_{Sp} provide initial information for planning a bimanual trajectory for climbing the ladder. In this case a footstep would be regarded as a platform grasp. Note that further information, e.g. concerning reachability and stability, is required for actual trajectory planning.

C. Example III: Valve

In our last example, we confronted the system with a DRC-inspired scenario: Turning an industrial valve (see Fig. 12). In order to demonstrate the feasibility of our perceptual pipeline and the usefulness of the resulting hints for action execution, we tested this example on the humanoid robot ARMAR-III.

The perceptual pipeline successfully identifies two predominant primitives in the scene: The valve and the wall (see Fig. 12a). Both of these primitives are planar, the valve however receives a high circularity score ($\text{circular}(p) \approx 1$). In this particular example we focus on the *bimanual turn* affordance (**Bi-Tn**, see Fig. 12b). The system successfully identifies end-effector poses for bimanual turning based on bimanual prismatic grasping with opposing end-effectors (**Bi-G-Op-Pr**). In the next step, we choose the most certain affordance among the available affordances (see Fig. 12c). Note that one of the criteria involved in $\Theta_{\text{Bi-Tn}}$ a horizontal relative orientation of the two end-effector poses. This is why the end-effector poses shown in Fig. 12c are chosen over the alternatives depicted in Fig. 12b. The end-effector poses that correspond to the chosen affordance used to execute a predefined skill for valve turning (see Fig. 12d). A recording of this experiment is presented as video attachment enclosed with this publication. The video is not accelerated.

V. Conclusion

We presented an hierarchical approach to the extraction of whole-body affordances, representing these affordances in terms of continuous certainty functions. Higher-level affordances result from combining certainties for lower-level affordances with additional environmental properties. The system produces valuable hints for action execution, e.g. possible end-effector poses, that can be used as an input for various planning components. The effectiveness of these hints was evaluated in multiple experiments based

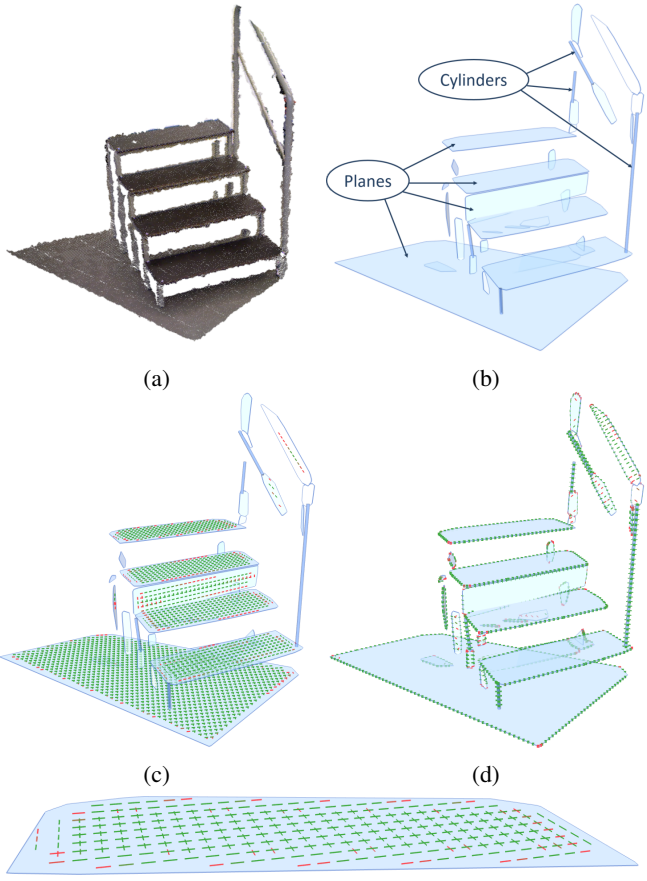

Fig. 10: Visualizations of different affordance certainty functions applied to a perceived staircase (a). Figure (b) displays the primitives extracted from the point cloud. Figures (c) and (d) visualize the affordance certainty functions $\Theta_{\text{G-Pl}}$ and $\Theta_{\text{G-Pr}}$, respectively. Figure (e) displays the visualization of a certainty function Θ in greater detail. Each sampled end-effector pose appears as a line colored from green (high certainty) to red (low certainty). Very low certainties ($\Theta \approx 0$) are omitted. The end-effector pose space for the visualized unimanual affordances is $\mathcal{S} = SE(3)$

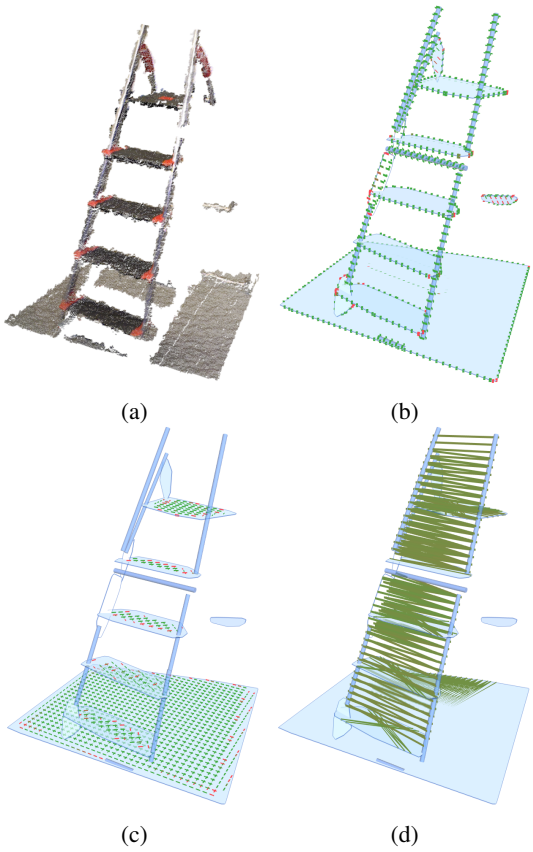

Fig. 11: Visualizations of different affordance certainty functions applied to a perceived ladder (a). The affordances shown are (b) $\Theta_{\text{G-Pr}}$ and (c) Θ_{Sp} and (d) $\Theta_{\text{Bi-G-Op-Pr}}$. The end-effector pose space for the visualized unimanual affordances is $\mathcal{S} = SE(3)$ and for the bimanual affordance $\mathcal{S} = SE(3) \times SE(3)$.

on captured point cloud data. One of these examples was implemented on a real robot platform demonstrating the feasibility of the complete approach.

A. Future Work

Currently, our affordance hierarchy is based on two grasp types that we consider predominant in actions of whole-body loco-manipulation: the platform grasp and the prismatic grasp. However, established grasp taxonomies define further types of power grasps that will be integrated into our framework in order to gain higher flexibility in affordance definition as well as in the produced hints for action execution. The system is capable of including additional sources of affordance certainties. These certainties can come from a human operator, autonomous exploration and validation or higher-level reasoning. Our future work will address the question of introducing such additional certainty functions into the affordance extraction process. Another extension to the existing method is the (partly) automatic acquisition of the affordance certainty functions, which are currently manually defined (see Table II).

References

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

[2] 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, p. 447, 2007.

[3] N. Krüger, C. Geib, J. Piater, R. Petrick, M. Steedman, F. Wörgötter, A. Ude, T. Asfour, D. Kraft, D. Omrčen, A. Agostini, and R. Dillmann, "Object-Action Complexes: Grounded Abstractions of Sensorimotor Processes," *Robotics and Autonomous Systems*, vol. 59, no. 10, pp. 740–757, 2011.

[4] J. Borràs and T. Asfour, "A whole-body pose taxonomy for loco-manipulation tasks," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015.

[5] A. Romay, S. Kohlbrecher, D. C. Conner, and O. von Stryk, "Achieving Versatile Manipulation Tasks with Unknown Objects by Supervised Humanoid Robots based on Object Templates," in *IEEE-RAS International Conference on Humanoid Robots*, pp. 249–255, 2015.

[6] 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,

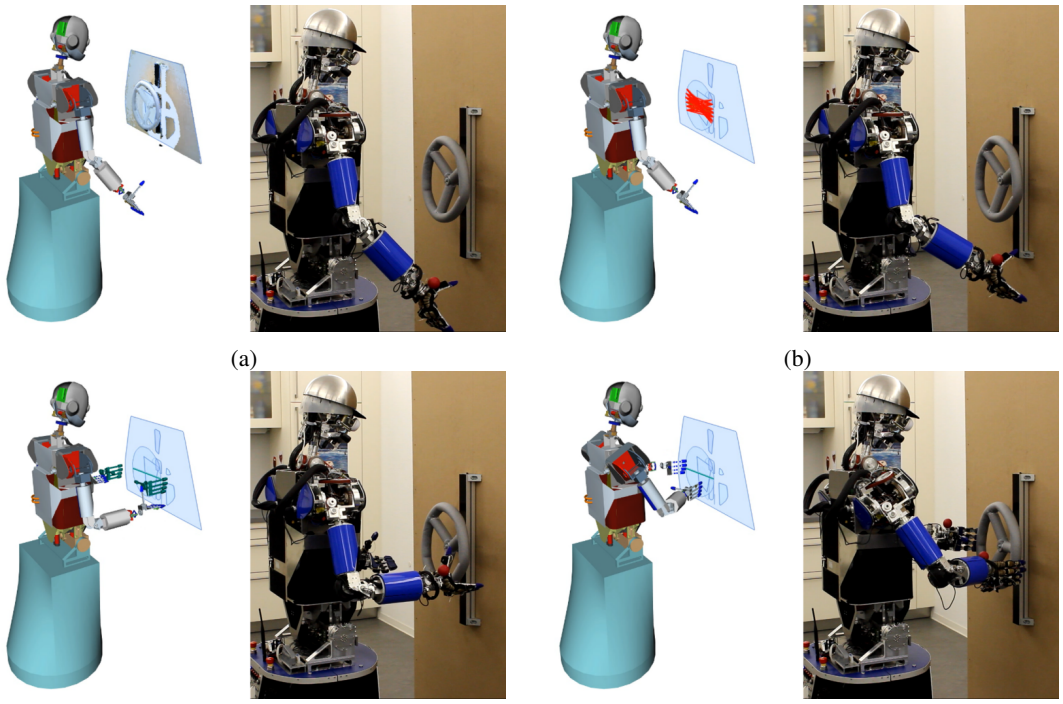

Fig. 12: The different stages of the valve-turning experiment for affordance extraction and affordance-driven action execution: (a) extraction of environmental primitives, (b) extraction of affordances for bimanual turning, (c) end-effector pose proposition based on extracted affordance certainties, (d) bimanual execution of a valve turning skill.

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.

[7] 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, 2015.

[8] R. R. Murphy, "Meta-analysis of Autonomy at the DARPA Robotics Challenge Trials," *Journal of Field Robotics*, vol. 32, no. 2, pp. 189–191, 2015.

[9] P. Kaiser, D. Gonzalez-Aguirre, F. Schültje, J. Borràs, N. Vahrenkamp, and T. Asfour, "Extracting Whole-Body Affordances from Multimodal Exploration," in *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, pp. 1036–1043, 2014.

[10] 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 (IJHR)*, 2015.

[11] 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 (Humanoids)*, 2015.

[12] S. C. Stein, M. Schoeler, J. Papon, and F. Wörgötter, "Object partitioning using local convexity," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 304–311, 2014.

[13] A. Myers, C. L. Teo, C. Fermüller, and Y. Aloimonos, "Affordance detection of tool parts from geometric features," in *ICRA*, 2015.

[14] A. Berner, J. Li, D. Holz, J. Stuckler, S. Behnke, and R. Klein, "Combining contour and shape primitives for object detection and pose estimation of prefabricated parts," in *IEEE International Conference on Image Processing (ICIP)*, pp. 3326–3330, 2013.

[15] D. I. Kim and G. S. Sukhatme, "Semantic Labeling of 3d Point Clouds with Object Affordance for Robot Manipulation," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5578–5584, 2014.

[16] S. Dalibard, A. Nakhaei, F. Lamiraux, and J.-P. Laumond, "Manipulation of Documented Objects by a Walking Humanoid Robot," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, pp. 518–523, 2010.

[17] M. Levihn and M. Stilman, "Using Environment Objects as Tools: Unconventional Door Opening," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2502–2508, 2014.

[18] K. Hauser, T. Bretl, and J.-C. Latombe, "Non-gaited humanoid locomotion planning," in *IEEE/RAS International Conference on Humanoid Robots*, pp. 7–12, 2005.

[19] S. Lengagne, J. Vaillant, E. Yoshida, and A. Kheddar, "Generation of whole-body optimal dynamic multi-contact motions," *The International Journal of Robotics Research*, vol. 32, no. 9-10, pp. 1104–1119, 2013.

[20] L. Sentis, J. Park, and O. Khatib, "Compliant Control of Multicontact and Center-of-Mass Behaviors in Humanoid Robots," *IEEE Transactions on Robotics*, vol. 26, no. 3, pp. 483–501, 2010.

[21] M. R. Cutkosky, "On grasp choice, grasp models, and the design of hands for manufacturing tasks," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, pp. 269–279, 1989.

[22] I. M. Bullock, J. Z. Zheng, S. De La Rosa, C. Guertler, and A. M. Dollar, "Grasp Frequency and Usage in Daily Household and Machine Shop Tasks," *IEEE Transactions on Haptics*, vol. 6, no. 3, pp. 296–308, 2013.

[23] M. Vergara, J. Sancho-Bru, V. Gracia-Ibáñez, and A. Pérez-González, "An introductory study of common grasps used by adults during performance of activities of daily living," *Journal of Hand Therapy*, vol. 27, pp. 225–234, 2014.

[24] T. Asfour, K. Regenstein, P. Azad, J. Schröder, A. Bierbaum, N. Vahrenkamp, and R. Dillmann, "ARMAR-III: An integrated humanoid platform for sensory-motor control," in *IEEE/RAS International Conference on Humanoid Robots*, pp. 169–175, 2006.

[25] T. Asfour, J. Schill, H. Peters, C. Klas, J. Bucker, C. Sander, S. Schulz, A. Kargov, T. Werner, and V. Bartenbach, "Armar-4: A 63 dof torque controlled humanoid robot," in *IEEE/RAS International Conference on Humanoid Robots*, pp. 390–396, 2013.

[26] J. W. Garrett, "The adult human hand: some anthropometric and biomechanical considerations," *The Journal of the Human Factors and Ergonomics Society*, vol. 13, no. 2, pp. 117–131, 1971.

[27] D. A. Winter, *Biomechanics and motor control of human movement*. New York: John Wiley & Sons, 1990.