Validation of Whole-Body Loco-Manipulation Affordances for Pushability and Liftability

Peter Kaiser, Markus Grotz, Eren E. Aksoy, Martin Do, Nikolaus Vahrenkamp and Tamim Asfour

***Abstract*— Autonomous robots that are intended to work in disaster scenarios like collapsed or contaminated buildings need to be able to efficiently identify action possibilities in unknown environments. This includes the detection of environmental elements that allow interaction, such as doors or debris, as well as the utilization of fixed environmental structures for stable whole-body loco-manipulation. Affordances that refer to whole-body actions are especially valuable for humanoid robots as the necessity of stabilization is an integral part of their control strategies.**

Based on our previous work we propose to apply the concept of affordances to actions of stable whole-body loco-manipulation, in particular to pushing and lifting of large objects. We extend our perceptual pipeline in order to build large-scale representations of the robot's environment in terms of environmental primitives like planes, cylinders and spheres. A rule-based system is employed to derive whole-body affordance hypotheses from these primitives, which are then subject to validation by the robot. An experimental evaluation demonstrates our progress in detection, validation and utilization of whole-body affordances.

I. INTRODUCTION

Autonomous or semi-autonomous robotic systems designed for disaster response in hazardous environments are often requested to have a humanoid structure because they have to perform tasks that have originally been carried out by humans. These tasks include the usage of tools highly specialized for the human body as well as the ability to establish stabilizing contacts to environmental elements like handrails, handles and stairs. In this work we apply the concept of affordances to actions of whole-body loco-manipulation, i.e. actions that involve the whole body for stabilization, locomotion or manipulation.

In our work whole-body affordance hypotheses are assigned to environmental primitives based on properties like shape, orientation and extent. Starting from registered RGB-D images, we employ a perceptual pipeline that segments the scene into environmental primitives and labels affordance hypotheses to these. Before the actual execution of an action based on a whole-body affordance, we employ a validation step in which the robot tries to touch the underlying environmental primitive in order to examine the existence of the affordance (see Fig. 1).

The research leading to these results has received funding from the European Union Seventh Framework Programme under grant agreement no 611832 (WALK-MAN) and grant agreement no 270273 (Xperience).

The authors are with the Institute for Anthropomatics and Robotics, High Performance Humanoid Technologies Lab (H^2T), at the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany. peter.kaiser@kit.edu

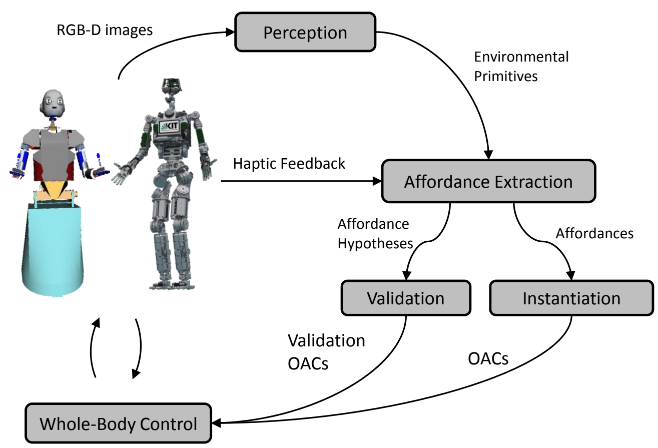


Fig. 1: Validation process of affordance hypotheses: The perceptual pipeline identifies environmental primitives which are assigned affordance hypotheses. These hypotheses pass a validation step in which the robot establishes contact to the primitive. Depending on the measured haptic feedback and the resulting level of confidence, the affordance hypotheses become actual affordances and can be instantiated as *Object Action Complexes* (OACs). These OACs can then be executed by the robot.

A. Related Work

The concept of affordances was originally introduced by Gibson [1] in the field of cognitive psychology for describing the perception of action possibilities. It states that an object suggest actions to the agent due to the object's shape and the agent's capabilities. A chair, for example, affords *sitting*, a cup *drinking* and a staircase *climbing*. Various works focus on learning grasp affordances, e.g. by initial visual perception and subsequent interaction with an object [2] [3] or by focusing on either haptic [4] or visual [5] perception. Further applications of affordance-based strategies can be found in locomotion and navigation, e.g. [6] [7], human-robot interaction, e.g. [8] [9] and symbolic planning, e.g. [10] [11]. An overview over attempts to formalize affordances with robotic applications in mind is presented in [7]. Several works have addressed the extraction of whole-body affordances, either based on predefined environmental models, e.g. [12], or by focusing on specific whole-body actions like stair-climbing, e.g. [13], or pushing and lifting of objects, e.g. [14]. In contrast to these works we create an intermediate simplified representation of the environment in terms of environmental primitives that allows the recognition of many different types of affordances.

Affordances are strongly related to *Object Action Complexes* (OACs) [15], a framework for the representation of sensorimotor experience and behaviors based on the coupling of objects and actions. Affordances as understood in this work can be regarded as preconditions for the instantiation of OACs: If there exists an OAC that couples an object o and an action a, then o must have suggested a in terms of an affordance. In this context, we think that the research on whole-body motion with contacts [16] [17] [18] can benefit from whole-body affordances by breaking down complex problems into separate parts that can be handled by a high level reasoning process, similar to [19].

In our previous work we proposed a framework for extracting whole-body affordance hypotheses based on environmental primitives derived from RGB-D camera images [20]. Kinematic reachability maps [21] [22] have been used in order to discard hypotheses clearly out of stable reach as well as to determine possible end-effector positions for affordance utilization. A first experimental evaluation of the concept was presented in [23], where perceived *supportability* and *leanability* affordance hypotheses were either accepted or rejected based on the resistance force of the underlying environmental primitive.

B. Contribution

In this paper we introduce improvements to various parts of the perceptual pipeline originally proposed in [20]:

- Point cloud registration methods allow us to extract affordances from large-scale environments constructed in the robot's memory instead of just the currently perceived scene (see Sec. II-A).
- In addition to planar primitives, cylinders and spheres can now be properly extracted and have been integrated into the affordance extraction methods (see Sec. II-C).
- The affordance extraction methods are extended in order to provide information on *pushability* and *liftability* of objects (see Sec. II-D).
- Inverted reachability maps have been integrated in order to find solutions to the problem of robot placement for affordance utilization (see Sec. II-F).
- We propose a strategy for computing grasp points for detected environmental primitives (see Sec. II-E).

The complete pipeline as employed in this work is depicted in Fig. 2. An experiment with the humanoid robot ARMAR-III [24] demonstrates the successful identification of pushable and liftable objects and the validation of the underlying affordances based on the validation process depicted in Fig. 1 (see Sec. III).

II. AFFORDANCE EXTRACTION PIPELINE

The affordance extraction pipeline proposed in this work is outlined in Fig. 2: Each captured RGB-D image is registered with previous frames and assembled into a larger point cloud. This point cloud passes several processing steps in which associated structures are segmented and further refined into environmental primitives. Affordances are assigned to these

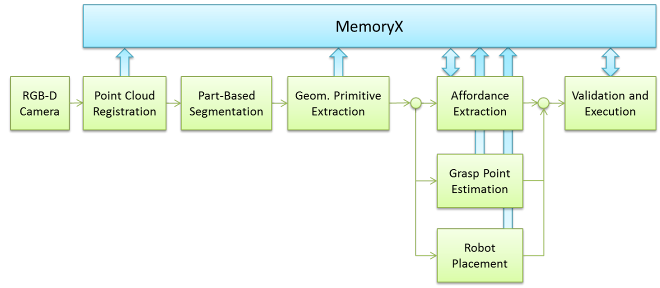
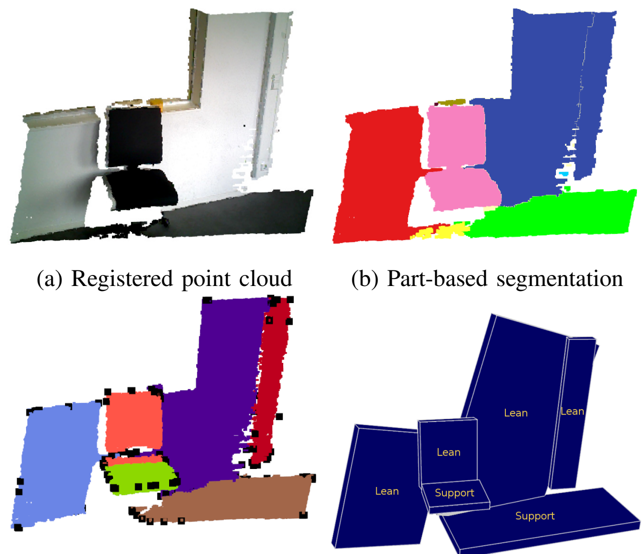


Fig. 2: The affordance extraction pipeline: RGB-D images are registered, segmented and classified into geometric primitives. Based on the resulting primitives, affordance hypotheses are derived and validated. Other possible outcomes of the pipeline are hypotheses for grasp points and robot locations.



(a) Registered point cloud (b) Part-based segmentation (c) Geometric primitives and grasp points (d) Affordance hypotheses

Fig. 3: The intermediate steps of the proposed perceptual pipeline. **(a)** RGB-D images are registered and merged into a combined point cloud. **(b)** The point cloud is segmented based on the convexity of surface patches. Each color represents one segment. **(c)** Geometric primitives are fitted into the segmented point cloud. **(d)** Affordances are extracted from the primitives.

primitives based on shape, orientation and extent, and are then subject to further validation by the robot. Two optional steps can be carried out in order to guide the validation process: The computation of grasp point hypotheses and suitable robot locations. Fig. 3 displays the intermediate results of the individual pipeline steps which will be explained in detail in the following sections.

A. Point Cloud Registration

Registration is the process of aligning sequentially captured point clouds such that they can be merged into a single RGB-D representation. In our experiments we use the state of the art open source SLAM library RTAB-Map [25] in combination with an Asus Xtion Pro depth sensor. The real-time capable RTAB-Map registration method bases on the tracking of 2D local image features among consecutive frames. RTAB-Map also features loop closure detection and graph pose optimization. Fig. 3a and Fig. 4 show examples for registered point clouds that serve as initial data for the affordance extraction pipeline.

B. Part-Based Segmentation

Once the captured point clouds are registered, we segment the scene into plausible and distinct regions by employing the segmentation method from [26]. In contrast to conventional segmentation methods that involve model fitting or learning techniques, this approach grows locally connected convex surface regions bounded by concavities. Convexly connected neighbor surface patches are then merged together resulting in a final scene segmentation. We refer the interested reader to [26] for further details on the segmentation method. Fig. 3b depicts the final part-based segmentation result of the registered point cloud shown in Fig. 3a. In the following we denote the segmentation of a registered point cloud $\mathcal{P}$ as a set of segments $s_i \subset \mathcal{P}$:

$$\mathcal{S} = \{s_1, \dots, s_n\} \tag{1}$$

C. Geometric Primitive Extraction

The next step in the perceptual pipeline from RGB-D images to affordances is the extraction of environmental primitives from the segmentation result $\mathcal{S}$. Each segment s_i is matched against geometric models, e.g. *plane*, *cylinder* or *sphere*, by using RANSAC [27]. More sophisticated methods for primitive extraction exist, e.g. [28] and [29], but they come at a much higher computational cost.

One of the main drawbacks of low-level feature-based segmentation methods is the possible under-segmentation of the scene, i.e. multiple distinct object segments happen to be merged, for instance due to noise in the depth cues. In such cases, a simple model fitting approach as proposed above is prone to error. To tackle the under-segmentation problem we customize the state of the art model fitting approach provided in [30].

Let the scene $\mathcal{S}$ be segmented into segments s_i. For each segment and its associated point cloud $\mathcal{P}_{s_i}$, our approach computes a set of disjoint geometric primitives

$$\Psi = \{\psi_1, \dots, \psi_m\}, \tag{2}$$

each of which defining either a *plane*, a *cylinder* or a *sphere*. The primitives ψ_i are represented by inlier point clouds

$$\mathcal{P}_{\psi_i} \subset \mathcal{P}_{s_i} \tag{3}$$

together with a corresponding set of outliers, i.e. segment points that have not been assigned to any of the primitives ψ_i:

$$\mathcal{O}_{s_i} = \mathcal{P}_{s_i} \setminus \bigcup_{j=1}^{m} \mathcal{P}_{\psi_j} \tag{4}$$

To partition a segment s_i into distinct primitives, we iteratively apply RANSAC to the segment point cloud $\mathcal{P}_{s_i}$. In each iteration, we compute fitting scores δ_{plane}, δ_{cylinder} and δ_{sphere} based on the maximum number of inliers for the three possible models. The model with the highest fitting score is instantiated as a new primitive ψ_{best}. Before adding ψ_{best} to the set of discovered primitives, the underlying point cloud $\mathcal{P}_{\psi_{\text{best}}}$ is further partitioned in a clustering process based on Euclidean distances between points. This step avoids distant clusters of points to be merged into one single primitive.

We repeat the same procedure over the remaining outliers $\mathcal{O}$ to generate further primitives, until the number of outliers, $|\mathcal{O}|$, is less than a threshold τ_{min}. The complete iterative primitive extraction approach is outlined in Algorithm 1.

Algorithm 1 primitiveExtraction($\mathcal{S}$, τ_{min}, τ_{max})
Input Segmentation $\mathcal{S}$
Input Minimum and maximum point cloud sizes τ_{min}, τ_{max}
Output A set of environmental primitives Ψ

```
Ψ ← ∅
for each s ∈ S do
    O ← s
    while |O| ∈ (τ_min, τ_max) do
        ψ_plane ← RANSAC_plane(O)
        ψ_cylinder ← RANSAC_cylinder(O)
        ψ_sphere ← RANSAC_sphere(O)
        ψ_best ← arg max_{ψ∈{ψ_plane, ψ_cylinder, ψ_sphere}} |P_ψ|
        if ψ_best = ∅ then
            break
        Ψ_new ← euclideanClustering(P_{ψ_best})
        Ψ ← Ψ ∪ Ψ_new
        O ← O \ P_{ψ_best}
return Ψ
```

Fig. 3c depicts the primitives extracted from the scene segmentation shown in Fig. 3b. Note that scene parts that are segmented into single segments due to under-segmentation, e.g. the chair, are now successfully partitioned into distinct primitives.

D. Affordance Extraction

In [20] we proposed to assign hypotheses for whole-body affordances like *support*, *lean*, *grasp* or *hold* to environmental primitives based on properties like shape, orientation and extent. Large vertical planes for instance are assumed to indicate *lean*-affordances. The types of affordances previously considered by us have mainly been chosen due to their importance for whole-body stabilization. However, further possible whole-body affordances exist and are of special interest when manipulating large, and possibly heavy, objects, for instance for removing debris from a blocked pathway. Examples for whole-body affordances indicating manipulability of objects are *pushability* and *liftability*, which are experimentally evaluated in this work. We extended the affordance hypotheses rule set from [20] to include extraction rules for *pushability* and *liftability* (see Table I).

While an exhaustive evaluation of the available types of whole-body affordances still remains to be done, *pushability*

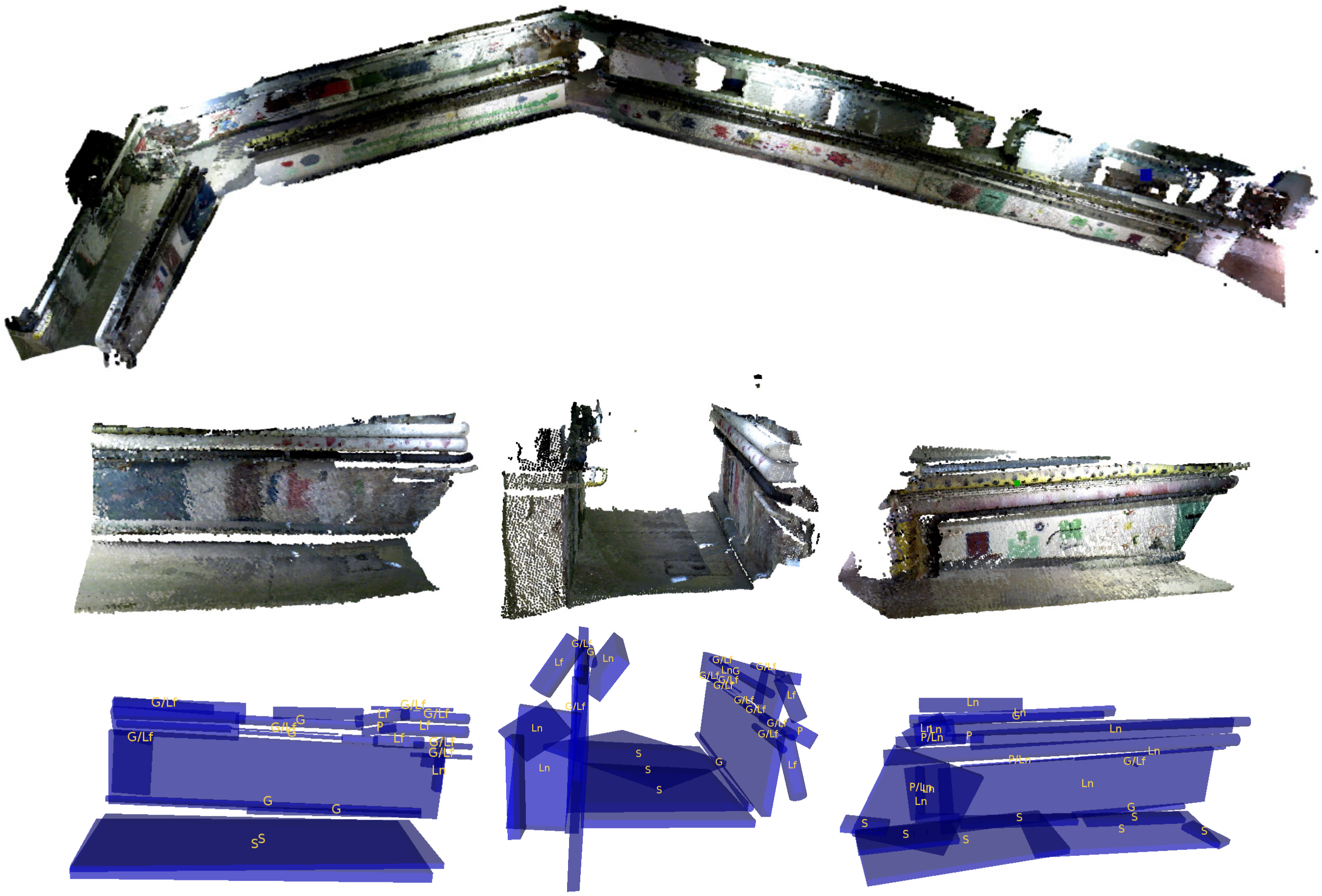


Fig. 4: The perceptual pipeline applied to a large scale registered point cloud from a tunnel scenario (**top**). The system successfully identifies planar and cylindrical primitives, i.e. floor, walls and pipes, and assigns affordance hypotheses to these (**bottom**). The affordance tags refer to Table I.

and *liftability* are certainly essential. In a way similar to [23], we integrate and evaluate the processes of affordance perception, validation and utilization on a real-world robotic system considering the two new affordance types.

The constants λ_i from Table I are currently application specific. However, we think that there is a set of affordance extraction parameters that will work reasonably well for our scenarios due to the following reasons:

- Research shows that agents infer affordances based on a body-scaled metric, i.e. with respect to the proportions of their bodies [31].
- We primarily focus our studies on disaster scenarios that contain at least partly intact man-made structures. These structures usually have standardized dimensions known beforehand.

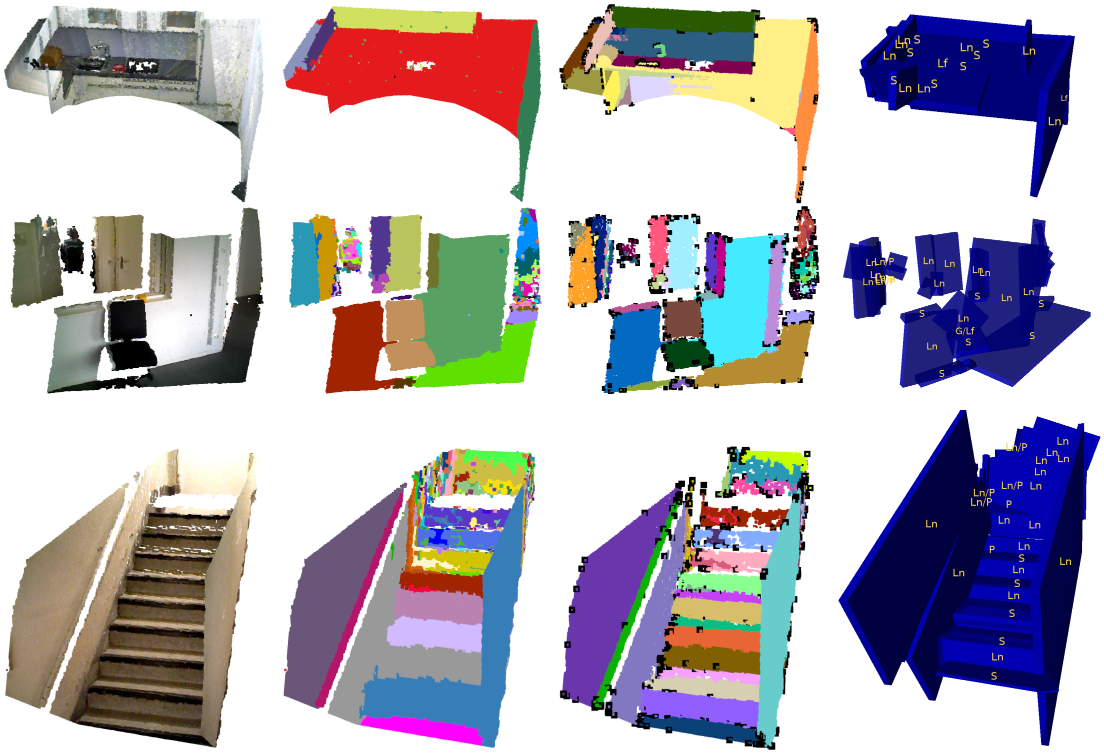
Fig. 4 and Fig. 6 visualize the environmental primitives extracted from two exemplary point clouds: A large scan of a tunnel and a highly cluttered industrial scene. The proposed framework successfully identifies the existing primitives in both scenes. The resulting primitives are assigned meaningful whole-body affordances based on the rules from Table I.

Fig. 5 depicts the different steps of the perceptual pipeline for multiple registered point clouds. The part-based segmentation results are illustrated in the second column. Note that the problem of under-segmentation appears in almost all of the samples independent of the scene context. The proposed iterative primitive extraction method yields a finer granularity here. The third column contains the derived geometric primitives together with grasp point hypotheses (black dots). Refer to section II-E for an explanation of the grasp point extraction process. Environmental primitives and extracted affordances are shown in the last column of Fig. 5.

Further examples of whole-body affordances extracted from environmental primitives, particularly in terms of pushability and liftability, are presented in Sec. III in the context of affordance validation.

E. Grasp Points

We investigated two possible extensions to the perceptual pipeline in order to enrich the information provided with environmental primitives: The computation of possible grasp points and robot locations. This additional information can be beneficial for the affordance validation procedure as well as for the later execution.



(a) Registered Point Cloud (b) Segmented Point Cloud (c) Primitives/Grasp Points (d) Affordances

Fig. 5: Visualization of the intermediate pipeline results of the affordance extraction process. The example scenarios are a kitchen sideboard (**top**), a chair in front of a window (**middle**) and a staircase (**bottom**). The affordance tags refer to Table I.

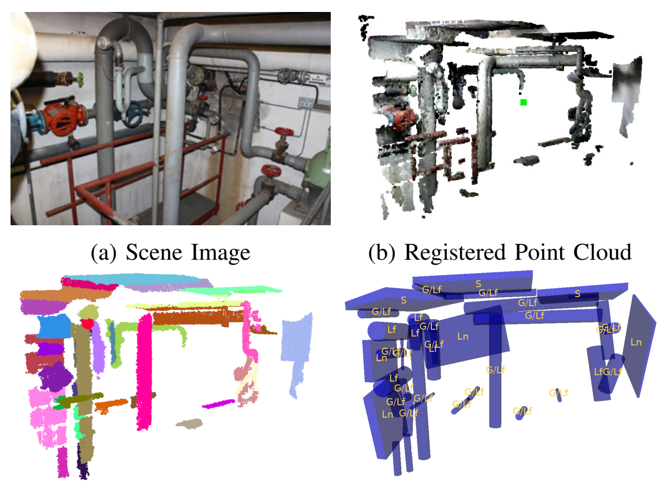
TABLE I: Exemplary rule set for affordance extraction and possible validation strategies.

Affordance	Shape	Parameters	Conditions[1,2]	Valid.
Support (S)	Planar	Normal $\boldsymbol{n}$ Area a	$\boldsymbol{n} \uparrow \boldsymbol{z}_{world}$ $a \geq \lambda_1$	(1a)
Lean (Ln)	Planar	Normal $\boldsymbol{n}$ Area a	$\boldsymbol{n} \perp \boldsymbol{z}_{world}$ $a \geq \lambda_2$	(1a)
Grasp (G)	Planar	Normal $\boldsymbol{n}$ Area a	$a \in [\lambda_3, \lambda_4]$	(3)
	Cylindrical	Radius r Direction $\boldsymbol{d}$	$r \in [\lambda_5, \lambda_6]$ $\|\boldsymbol{d}\| \leq \lambda_7$	
	Spherical	Radius r	$r \in [\lambda_8, \lambda_9]$	
Hold (H)	Cylindrical	Radius r Direction $\boldsymbol{d}$	$r \in [\lambda_{10}, \lambda_{11}]$ $\|\boldsymbol{d}\| \geq \lambda_{12}$	(2a)
Push (P)	Planar	Normal $\boldsymbol{n}$ Area a	$\boldsymbol{n} \perp \boldsymbol{z}_{world}$ $a \leq \lambda_{13}$	(1b)
Lift (Lf)	Planar	Normal $\boldsymbol{n}$ Area a	$a \leq \lambda_{15}$	(2b)
	Cylindrical	Radius r Direction $\boldsymbol{d}$	$r \leq \lambda_{15}$ $\|\boldsymbol{d}\| \leq \lambda_{16}$	
	Spherical	Radius r	$r \leq \lambda_{17}$	

1 The values λ_i are implementation-specific constants.

2 The operator $\uparrow$ tells if two vectors point into the same direction:

$$\boldsymbol{v} \uparrow \boldsymbol{w} \leftrightarrow \frac{\boldsymbol{v} \cdot \boldsymbol{w}}{\|\boldsymbol{v}\| \cdot \|\boldsymbol{w}\|} \approx 1$$



(a) Scene Image (b) Registered Point Cloud

(c) Geometric Primitives (d) Affordances

Fig. 6: The perceptual pipeline applied to an exemplary industrial scene containing a high amount of clutter (**top left**). The system successfully identifies planar and cylindrical primitives and assigns affordance hypotheses to these (**bottom right**). Note that for the sake of clarity, we show only a slice of the entire registered point cloud. The affordance tags refer to Table I.

For the computation of grasp points, we project the inlier point cloud $\mathcal{P}_\psi$ of each primitive ψ to the primitive's boundary $\partial\psi$. We then create a convex hull around these projected inliers:

$$\mathcal{G} \leftarrow \text{convexHull}\Big(\big\{\text{project}(x, \partial\psi) : x \in \mathcal{P}_\psi\big\}\Big) \qquad (5)$$

The convex hull approximates the smallest set of points that encloses the boundary of the primitive. Those resulting sparse hull points represent potential grasp points for the respective primitive. Black dots indicate computed grasp points in Fig. 3c and Fig. 5.

The extracted grasp points can prove to not exist due to various reasons, e.g. errors in the perceptual process or kinematic limitations of the robot. Reachability maps [21], [22] provide means for discarding kinematically unreachable grasp points based on the current robot pose.

F. Robot Location

Reachability maps can assist in assorting grasp points by kinematic feasibility based on the current robot location. However, in our scenario the robot creates a registered representation of its environment with attached information on primitives and affordances based on multiple views. It is particularly possible for the robot to locomote within its environment in order to optimize access to specific environmental primitives. Vahrenkamp et al. proposed inverted reachability maps that suggest robot locations based on desired end effector poses [32]. Fig. 7 displays the results of an inverted reachability query for a chosen grasp point in an exemplary scene.

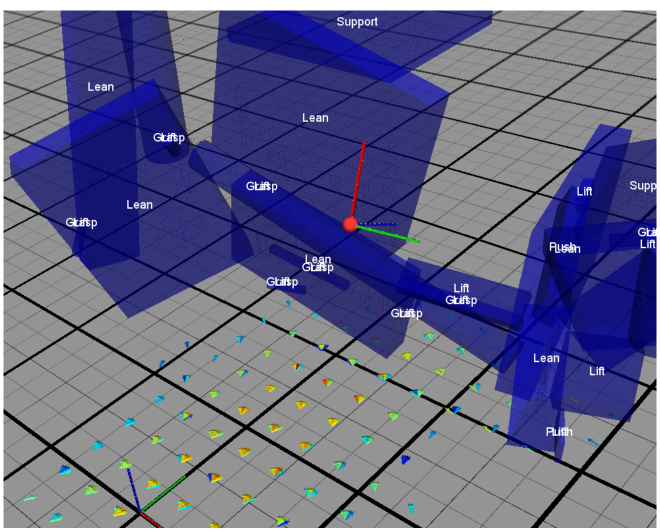


Fig. 7: Example queries to an inverted reachability map based on a chosen grasp point (**red ball**). Possible robot locations are represented by colored fans with hotter colors indicating higher reachability ratings.

III. VALIDATION OF AFFORDANCE HYPOTHESES

The strategies for affordance extraction outlined above are purely based on visual perception. Extracted affordances are in fact *affordance hypotheses* and therefore subject to further investigation and validation by the robot (see Fig. 1). While precomputed reachability maps can help to discard non utilizeable affordances (as proposed in [20]), there is no reliable mechanism for verifying the existence of whole-body affordances without establishing contact to the underlying primitives. Related approaches observe object bahavior during action execution in order to evaluate learned affordances, e.g. [33].

Referring to the last column of Table I, different force-based validation strategies exist based on the type of the affordance to investigate:

(1) Touch the primitive and exert forces along the primitive's normal $\mathbf{n}$. Compare the resistance force against a minimum ϑ_1 **(1a)** or a maximum ϑ_2 **(1b)**.
(2) Grasp the primitive and exert forces along the expected direction of utilization. Compare the resistance force against a minimum ϑ_3 **(2a)** or a maximum ϑ_4 **(2b)**.
(3) Push the primitive and perceive the caused effect.

Considering further sensor modalities apart from contact forces is of interest and can lead to more sophisticated and accurate validation strategies. Validating the *pushability* or *graspability* of very light objects for instance might not result in reliable resistance force feedback. Possible solutions for cases like these include tactile feedback or the comparison of RGB-D images before and after the push, similar to [34].

A. Experiment

In this section we describe an experiment carried out on the humanoid robot ARMAR-III, demonstrating the perception and validation of affordance hypotheses for *pushability* and *liftability*. In the experiment ARMAR-III is facing a cluttered arrangement of different obstacles that block its way: A chair, a box and a pipe (see Fig. 8, top left corner). The robot has no a-priori knowledge on the types or locations of the employed obstacles, the only information it gets results from the perceptual pipeline discussed above. The robot executes the following strategy:

1) Move to a given start position in front of the obstacles.
2) Capture and register multiple depth images in order to obtain a wide-angle view of the scene.
3) Run the perceptual pipeline outlined in section II.
4) Pick the most disturbing primitive that can be moved away, i.e. that carries a pushability or liftability affordance hypothesis.
5) Validate the affordance according to the validation rules from Table I (explained in section III).
6) If the validation was successful, utilize the affordance in order to remove the obstacle.
7) Repeat until no further obstacles are found.

Fig. 8 displays snapshots of different stages of the experiment: perception (first column), validation (second column) and execution (third column). The perception stage displays the initial obstacle arrangement and its representation after the perceptual pipeline in terms of primitives and affordance hypotheses. The validation stage includes the establishment of contact with the selected primitive and the affordance validation based on the obstacle's resistant force. In the execution phase, the robot has validated the affordance in question and starts pushing or lifting the obstacle, respectively.

The robot successfully identifies all three obstacles and starts by validating the liftability of the pipe (Fig. 8, **first row**) The validated liftability is then exploited for moving the obstacle away. In the next steps the robot identifies the chair and the box as pushable obstacles and validates these affordances accordingly (Fig. 8, **second row**, **third row**). The last row of Fig. 8 displays a repetition of the previous scene with a fixed box. The robot again assigns a pushability hypothesis to the box, but fails to validate this hypothesis. Hence, the corresponding push cannot be executed.

IV. CONCLUSION AND FUTURE WORK

Based on our previous results we proposed a perceptual pipeline that allows the extraction of whole-body affordances based on detected environmental primitives. The pipeline starts from RGB-D camera images and imposes methods for point cloud registration, part-based segmentation and geometric primitive regression. The results of the perceptual pipeline can be further processed for discovering grasp points on the detected primitives and robot poses beneficial for the utilization of affordances.

The pipeline was evaluated with several exemplary scenes, producing reasonable affordance hypotheses. It needs to be mentioned that due to its simplicity the affordance extraction method is prone to misconception. However, we consider the results to be affordance hypotheses that need to be validated by the robot before being utilized. We have implemented the perceptual pipeline on the humanoid robot ARMAR-III, demonstrating the feasibility of extraction and validation of affordances.

In our future work we will further examine the space of whole-body affordances in terms of completeness, extraction rules and validation strategies. Interesting attempts have been made to create ontologies of affordances [35] or manipulation actions [36], which could be useful for creating a complete set of whole-body affordances. We are planning to add further sensor modalities in order to reliably ground the discovered affordances even in difficult cases. Automatic selection of validation strategies based on a measure of reliability assigned to each affordance will be a central aspect.

REFERENCES

[1] J. J. Gibson, *The ecological approach to visual perception*. Psychology Press, 1979.

[2] M. Popović, D. Kraft, L. Bodenhagen, E. Başeski, N. Pugeault, D. Kragic, T. Asfour, and N. Krüger, "A strategy for grasping unknown objects based on co-planarity and colour information," *Robotics and Autonomous Systems*, vol. 58, no. 5, pp. 551–565, 2010.

[3] 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, 2011.

[4] A. Bierbaum, M. Rambow, T. Asfour, and R. Dillmann, "Grasp affordances from multi-fingered tactile exploration using dynamic potential fields," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, pp. 168–174, 2009.

[5] A. ten Pas and R. Platt, "Localizing grasp affordances in 3-D points clouds using taubin quadric fitting," in *International Symposium on Experimental Robotics (ISER)*, 2014.

[6] S. L. Gay, O. L. Georgeon, and C. Wolf, "Autonomous object modeling based on affordances for spatial organization of behavior," pp. 94–99, 2014.

[7] 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," vol. Adaptive Behavior, no. 15, p. 447, 2007.

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

[9] H. S. Koppula and A. Saxena, "Anticipating Human Activities using Object Affordances for Reactive Robotic Response," in *Robotics: Science and Systems (RSS)*, 2013.

[10] E. Ugur, E. Oztop, and E. Sahin, "Goal emulation and planning in perceptual space using learned affordances," *Robotics and Autonomous Systems*, vol. 59, no. 7, pp. 580–595, 2011.

[11] E. Ugur and J. Piater, "Bottom-Up Learning of Object Categories, Action Effects and Logical Rules: From Continuous Manipulative Exploration to Symbolic Planning," in *International Conference on Robotics and Automation*, pp. 119–139, 5 2015.

[12] M. Fallon, S. Kuindersma, S. Karumanchi, M. Antone, T. Schneider, H. Dai, C. Prez 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.

[13] S. Oßwald, A. Görög, A. Hornung, and M. Bennewitz, "Autonomous Climbing of Spiral Staircases with Humanoids," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4844–4849, 2011.

[14] 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.

[15] N. Krüger, C. Geib, J. Piater, R. Petrick, M. Steedman, F. Wörgötter, A. Ude, T. Asfour, D. Kraft, D. Omren, A. Agostini, and R. Dillmann, "Object-action complexes: Grounded abstractions of sensorimotor processes," *Robotics and Autonomous Systems*, vol. 59, pp. 740–757, 2011.

[16] 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, pp. 1104–1119, 2013.

[17] L. Sentis, "Compliant control of whole-body multi-contact behaviors in humanoid robots," in *Motion Planning for Humanoid Robots*, pp. 29–66, Springer, 2010.

[18] 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.

[19] D. Leidner, A. Dietrich, F. Schmidt, C. Borst, and A. Albu-Schäffer, "Object-centered hybrid reasoning for whole-body mobile manipulation," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1828–1835, 2014.

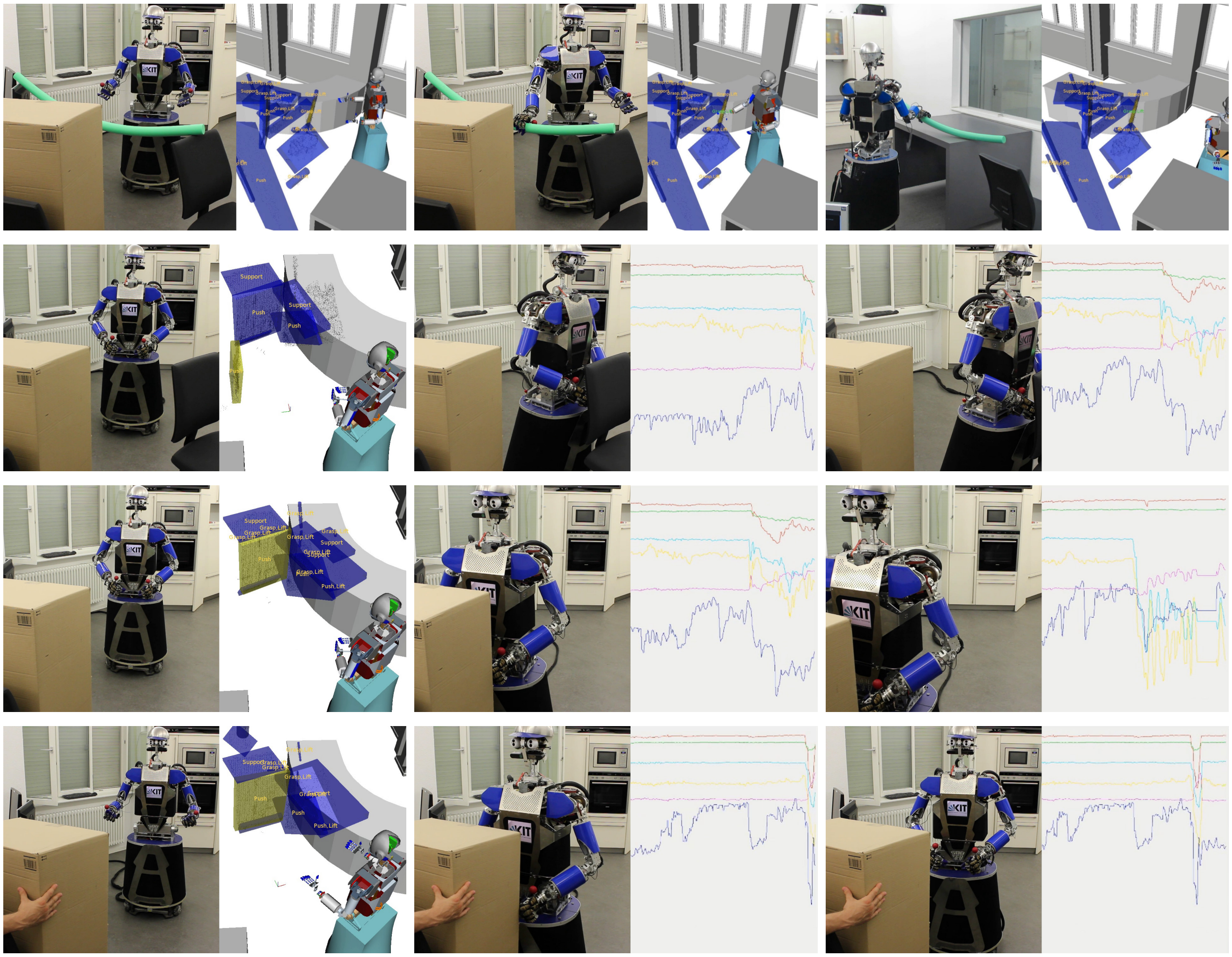
[20] P. Kaiser, D. Gonzalez-Aguirre, F. Schültje, J. Borràs, N. Vahrenkamp, and T. Asfour, "Extracting Whole-Body Affordances from Multimodal Exploration," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, pp. 1036–1043, 2014.

[21] F. Zacharias, C. Borst, and G. Hirzinger, "Capturing robot workspace structure: representing robot capabilities," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3229–3236, 2007.

[22] N. Vahrenkamp, T. Asfour, and R. Dillmann, "Efficient inverse kinematics computation based on reachability analysis," *International Journal of Humanoid Robotics*, vol. 9, no. 4, 2012.

[23] 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.

[24] T. Asfour, K. Regenstein, P. Azad, J. Schröder, N. Vahrenkamp, and R. Dillmann, "ARMAR-III: An Integrated Humanoid Platform for Sensory-Motor Control," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, pp. 169–175, 2006.



(a) Perception (b) Validation (c) Execution

Fig. 8: The three stages of perception, validation and execution of whole-body affordances in four different scenarios: A pipe that can be grasped and lifted (**first row**), a chair that can be pushed (**second row**), a box that can be pushed (**third row**) and a box that is fixed and cannot be pushed (**fourth row**). The plots visualize the force amplitudes (*y-axis*) measured in the robot's left wrist over time (*x-axis*), while the blue curve represents the force in pushing direction.

[25] M. Labbe and F. Michaud, "Online global loop closure detection for large-scale multi-session graph-based slam," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2661–2666, IEEE, 2014.

[26] S. Stein, F. Wörgötter, M. Schoeler, J. Papon, and T. Kulvicius, "Convexity based object partitioning for robot applications," in *International Conference on Robotics and Automation (ICRA)*, pp. 3213–3220, 2014.

[27] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[28] R. Schnabel, R. Wahl, and R. Klein, "Efficient RANSAC for Point-Cloud Shape Detection," *Computer graphics forum*, vol. 26, no. 2, pp. 214–226, 2007.

[29] Y. Li, X. Wu, Y. Chrysathou, A. Sharf, D. Cohen-Or, and N. J. Mitra, "GlobFit: Consistently Fitting Primitives by Discovering Global Relations," *ACM Transactions on Graphics*, vol. 30, no. 4, 2011.

[30] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *International Conference on Robotics and Automation (ICRA)*, pp. 1–4, 2011.

[31] W. H. Warren, "Perceiving affordances: Visual guidance of stair climbing," *Journal of Experimental Psychology*, vol. 10, no. 5, pp. 683–703, 1984.

[32] N. Vahrenkamp, T. Asfour, and R. Dillmann, "Robot placement based on reachability inversion," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1970–1975, 2013.

[33] E. Ugur and J. Piater, "Emergent Structuring of Interdependent Affordance Learning Tasks," in *International Conference on Development and Learning and on Epigenetic Robotics*, pp. 481–486, 2014.

[34] D. Schiebener, A. Ude, and T. Asfour, "Physical interaction for segmentation of unknown textured and non-textured rigid objects," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4959–4966, 2014.

[35] K. M. Varadarajan and M. Vincze, "AfNet: The affordance network," in *Proceedings of the 11th Asian Conference on Computer Vision*, pp. 512–523, 2012.

[36] F. Wörgötter, E. E. Aksoy, N. Krüger, J. Piater, A. Ude, and M. Tamosiunaite, "A Simple Ontology of Manipulation Actions based on Hand-Object Relations," *IEEE Transactions on Autonomous Mental Development*, vol. 5, no. 2, pp. 117–134, 2013.