2018 IEEE International Conference on Robotics and Automation (ICRA)
May 21-25, 2018, Brisbane, Australia

Affordance-Based Multi-Contact Whole-Body Pose Sequence Planning for Humanoid Robots in Unknown Environments

Peter Kaiser, Christian Mandery, Andreas Boltres and Tamim Asfour

Abstract— **Despite impressive advances of humanoid robotics, the autonomous planning of whole-body loco-manipulation actions in unknown environments is still an open problem. In our previous work, we addressed two fundamental aspects related to this problem: 1) the autonomous detection of end-effector contact opportunities in unknown environments and 2) the goal-directed planning of multi-contact pose sequences, which can serve as the starting point for motion planning and control approaches of reduced complexity. Both problems suffer from the extensive amounts of possible solutions, particularly due to the complexity of humanoid robots and the multitude of available contact opportunities.**

In this paper, we propose a method for the planning of whole-body multi-contact tasks based on our previous work on vision-based detection of loco-manipulation affordances and whole-body multi-contact pose sequence planning. We demonstrate a combined approach for planning multi-contact pose sequences with a focus on the utilization of available end-effectors for stabilizing contacts with the environment during loco-manipulation tasks. The method is evaluated in simulation in multiple exemplary scenarios based on actual sensor data and the humanoid robot ARMAR-4.

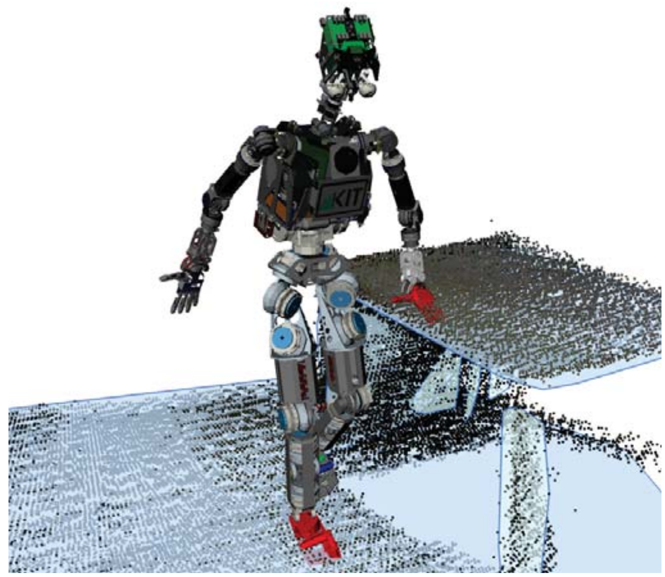

Fig. 1: Exemplary multi-contact pose of the humanoid robot ARMAR-4 with additional hand support.

I. Introduction

Humanoid robots are designed to operate in unstructured, human-centered environments, which are not entirely known in advance. Within these environments, the robots are intended to demonstrate human-like *whole-body loco-manipulation* capabilities, i. e. the execution of combined locomotion and manipulation actions that incorporate the whole body. Such actions are particularly important for realizing stable locomotion over challenging terrain, e. g. rubble or stairs, where multi-contact motions that utilize the robot hands are desired to enhance stability (see Fig. 1). Due to the high dimensionality of the configuration space of a humanoid robot, the imprecise and semantically unlabeled environmental perception, and the infeasible amount of possible environmental contacts, the problem of planning whole-body multi-contact loco-manipulation actions is challenging and considered unsolved. While existing approaches employ sophisticated, high-dimensional motion planners for planning complex whole-body actions, these approaches commonly assume entirely known environments and require excessively long planning times, making their online application hard.

To solve the problem of multi-contact motions for humanoid robots, we identify three distinct subproblems, which need to be addressed: 1) the detection of possibilities for support contact in the unknown environment, 2) the planning of pose sequences based on the available contact opportunities and 3) the planning and control of the actual motion trajectory based on the generated intermediate poses. In our previous work, we proposed methods and solutions to the first two problems:

1) In [1], we presented a perceptive-cognitive system for the autonomous detection of loco-manipulation affordances in unknown environments, which has been successfully applied in the context of shared autonomous control of humanoid robots [2]. In this control mode, the cognitively challenging roles of action and task planning based on detected affordances are leveraged to a human pilot. See Section III-A for a more detailed introduction.
2) Following the *contacts-before-motion* paradigm [3], we proposed in [4] an approach for the generation of whole-body multi-contact pose sequences, which is inspired by techniques from natural language processing. In this approach, an n-gram model learned from observed human motions is used to model transition probabilities between characteristic whole-body poses. See Section III-B for a more detailed introduction.

In this paper, we aim at combining the above methods into a joint approach for the initially defined problems

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

of environmental contact affordance perception and whole-body pose planning. Our approach allows to generate multi-contact pose sequences for whole-body loco-manipulation tasks based on affordances detected in unknown environments. We demonstrate the validity of our approach and its ability to generate feasible pose sequences in four exemplary scenarios.

The remainder of this work is structured as follows: Section II discusses related approaches in the two areas outlined above. Section III provides details on our previously proposed approaches to autonomous affordance detection and whole-body multi-contact pose planning, before introducing the combined approach in Section III-C. Section IV evaluates the combined approach based on real sensor data using the simulated humanoid robot ARMAR-4, before Section V concludes the paper and discusses future work.

II. Related Work

The problem of whole-body multi-contact motion planning for loco-manipulation actions in unknown environments is central in the field of humanoid robotics and has been approached with a variety of different motivations. Many researchers focused on particular aspects of the problem, e. g. in the areas of motion planning, contact and footstep planning or visual perception. In the following, related work will be presented concerning the two essential aspects of our approach: the detection of whole-body loco-manipulation affordances and the generation of whole-body motions for humanoid robots given a model of the environment.

A. Whole-Body Affordances

The psychological theory of *affordances* [5] attempts to explain the process of perception of action and interaction possibilities of humans and animals with the world. It suggests that affordances, i. e. action possibilities, arise based on properties of perceived objects and the perceiving agent's capabilities. Extensive surveys on the application of affordances in robotics can be found in [6], [7]. The idea of *whole-body affordances* has been introduced in [8], referring to affordances that relate to whole-body loco-manipulation actions. State-of-the-art approaches to the detection of action possibilities with humanoid robots commonly define affordances as templates, containing 3D representations of objects and conceptual descriptions of their utilization. Due to the embedded object descriptions, 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 [9]–[11]. Template-based approaches work well in environments, where principle structures are known with only small tolerated variations. A conceptually related approach for the detection of loco-manipulation affordances was used in [12] for locomotion planning in unknown environments.

B. Generation of Whole-Body Motions

Many authors have demonstrated the generation of whole-body motions by using optimization techniques, such as quadratic programming [13]–[17], or by using randomized sampling-based algorithms, such as variants of rapidly-exploring random trees [18], [19]. These approaches however cannot be used to efficiently solve the problem of multi-contact locomotion, where long motion sequences with frequent contact changes paired with a large number of DoFs result in a high-dimensional search space. A reduction in problem complexity can be achieved for such motion tasks by sequencing given motion primitives to form the resulting whole-body motion [20], [21]. A popular approach to simplifying the search space for multi-contact motions formulates the problem in a *contacts-before-motion approach* [3], [22], which separates the problem into two steps that can be solved independently. In the first step, a generated motion is represented as a sequence of *stance poses* or *configuration poses* that specify the order and locations of contacts to be used during the motion [3], [12], [23], [24]. The generation of multi-contact pose sequences presented in this paper also aims to solve this first step, with its input being represented by the perceived whole-body affordances of the environment. In a second step not tackled by this paper, these stance poses can then be linked to a continuous trajectory that e. g. satisfies given stability constraints [25], [26].

III. Approach

Before introducing the proposed approach in Section III-C as the main contribution of this paper, we will briefly introduce our previously proposed works in the areas of loco-manipulation affordance detection (Section III-A) and whole-body multi-contact pose sequence generation (Section III-B).

A. Whole-Body Loco-Manipulation Affordances

In [1], we proposed a formalization of the affordance concept based on *affordance belief functions* Θ_a, defined over the space of end-effector poses:

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

Affordance belief functions assess the existence of affordances by mapping Dempster-Shafer belief expressions $d \in \mathcal{D}$ to end-effector poses $\boldsymbol{x} \in SE(3)$. In order to effectively reason about complex loco-manipulation affordances, we formalized the hierarchical composition of affordance belief functions, which enables the consistent integration of affordance-related evidence on multiple layers. Evidence gained on lower-level affordances, e. g. *platform graspability*, is considered in the joint belief in higher-level affordances, e. g. *supportability*. Fig. 2 shows the composition rule for the affordance belief function Θ_{Support}, representing *supportability* affordances. The composition is based on the lower-level affordance belief function $\Theta_{\text{G-Platform}}$ for *platform graspability* and the property belief functions $\Theta_{\text{Horizontal}}$ and Θ_{Fixed} which estimate horizontal orientation and assumed mobility of a primitive p based on geometric properties.

While a core concept of the proposed formalization is the integration of affordance-related evidence from multiple sources, the principle source of affordance-related information is visual perception. The experiments in [1] and [2]

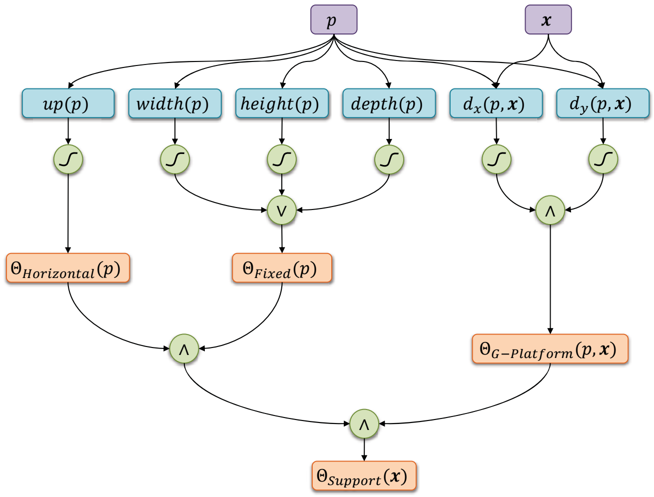

Fig. 2: The composition of an affordance belief function Θ_{Support} for *support* affordances.

demonstrate that visual affordance detection based on the above formalism is a viable and feasible approach to the detection of action possibilities in unknown environments for whole-body loco-manipulation on real humanoid robots.

B. Multi-Contact Whole-Body Pose Planning

The generation of multi-contact pose sequences is based on a statistical approach first proposed in [4], which uses an n-gram model to describe transition probabilities between whole-body poses. This use of an n-gram model represents a linguistic point of view towards the problem of multi-contact motion generation, wherein *words* represent whole-body poses and *sentences* represent multi-contact motions. The n-gram model describes the conditional probability

$$P\big(k_i \mid (k_{i-n+1}, \ldots, k_{i-1})\big) \tag{2}$$

of observing a certain whole-body pose k_i, given the $n-1$ preceding poses. The probability of a pose sequence of arbitrary length can then be determined by multiplication of all conditional probabilities in that pose sequence:

$$P\big((k_1, \ldots, k_N)\big) = \prod_{i=1}^{N} P\big(k_i \mid (k_{i-n+1}, \ldots, k_{i-1})\big). \tag{3}$$

The four individual steps for learning this statistical model for pose transitions from human observation are outlined in the following, while further details on the process can be found in [4]:

1) Motion Acquisition: Whole-body motion demonstrations of human subjects are captured together with the motions of environmental elements utilized in the demonstration, for which we use a marker-based passive-optical motion capture system[1]. From the captured marker trajectories, the 6D root pose and joint angles are reconstructed for the 104-DoF *Master Motor Map* (MMM) human reference model [27], which is based on existing biomechanics literature. Since fingers and some other joints of the MMM are excluded in this work, 40 of the available joints are used for this reconstruction and subsequently for the definition of whole-body configuration poses.

2) Support Pose Segmentation: Reconstructed MMM motions are segmented with regard to the used *support poses* of the human subject [28]. We consider both the velocities of human body segments and their distances to supporting environmental objects to detect supporting contacts between the human and the environment. For a given set of support contacts at any point in time, the corresponding support pose is determined and hence the motion is represented as a sequence of these support poses.

3) Configuration Pose Classification: Since support poses extracted from the motion segmentation are defined only through the set of used support contacts, these poses are further subdivided into *configuration poses*[2], which define concrete values for the 40 considered joints of the MMM human reference model. As described in [4], 8 to 16 possible configuration poses are extracted from human motion data for each of the support poses, yielding a total number of 111 possible configuration poses. Using the pose retargeting capabilities of the MMM framework [27], these configuration poses can subsequently be transferred to the kinematic embodiment of an arbitrary humanoid robot model. By further processing the segmented motion data from the previous step, each occurrence of a support pose in the segmented motion data is assigned to its closest matching configuration pose, minimizing spatial deviations for the four end-effectors of the human body. As a result, we represent captured human motions as sequences of configuration poses.

4) Model Learning: In the final step, the n-gram model[3] for pose transitions is learned from a sufficiently large set of representative training motions. Training the n-gram model is based on textual representations of the observed motions as configuration pose sequences, facilitated using the *SRILM Toolkit* [29]. In addition, we are learning a spatial translation model which describes the displacement of the whole-body center of mass along the straight line towards the goal position of a locomotion.

Given the statistical representation of transitions between whole-body poses in human motion, in [4] we proposed an algorithm for planning multi-contact motions that is based on maximizing the n-gram model likelihood of the generated pose sequence while sufficing all constraints imposed by a given locomotion task. The algorithm is shown as pseudocode in Algorithm 1. Since the number of possible pose sequences grows exponentially with the length, we cannot consider all possible pose sequences, but use an informed breadth-first search to explore the space of these possible pose sequences. Additionally, pruning techniques (explained in more detail in [4]) are employed to limit the growth of active paths during the breadth-first search (lines 7–16).

[1] VICON MX system with ten T10 cameras running at 100 Hz.

[2] Dubbed *shape poses* in [4].

[3] In this work: $n = 5$, Witten-Bell smoothing. See [4] for details.

Algorithm 1 Pose Sequence Planning (modified from [4])

1: *activePaths* $\leftarrow$ heap()
2: Insert Path(*startPose*) into *activePaths*
3: $i \leftarrow 0$
4: **loop**
5: $\quad i \leftarrow i + 1$
6: $\quad$ *bestPath* $\leftarrow$ ExtractMax(*activePaths*)
7: $\quad$ **if** ($i \mod$ *prunePeriod*) $= 0$ **then**
8: $\quad\quad$ *pruneDist* $\leftarrow$ *bestPath.distance* $-$ *pruneThresh*
9: $\quad\quad$ *newPaths* $\leftarrow$ heap()
10: $\quad\quad$ **for all** *path* $\in$ *activePaths* **do**
11: $\quad\quad\quad$ **if** *path.distance* $\geq$ *pruneDist* **then**
12: $\quad\quad\quad\quad$ Insert *path* into *newPaths*
13: $\quad\quad\quad$ **end if**
14: $\quad\quad$ **end for**
15: $\quad\quad$ *activePaths* $\leftarrow$ *newPaths*
16: $\quad$ **end if**
17: $\quad$ **if** *activePath.distance* $\geq$ *distance* **and** *bestPath.endPose* $=$ *endPose* **then**
18: $\quad\quad$ **return** *bestPath* $\quad\triangleright$ Solution found
19: $\quad$ **else** $\quad\triangleright$ *bestPath* is not a solution
20: $\quad\quad$ *expandedPaths* $\leftarrow$ ExpandPath($bestPath$)
21: $\quad\quad$ **for all** *path* $\in$ *expandedPaths* **do**
22: $\quad\quad\quad$ Score *path* using n-gram model
23: $\quad\quad\quad$ Insert *path* into *activePaths*
24: $\quad\quad$ **end for**
25: $\quad$ **end if**
26: **end loop**

Algorithm 2 Expansion of Pose Sequence

1: **function** ExpandPath(*path*)
2: $\quad$ *newPaths* $\leftarrow$ list()
3: $\quad\triangleright$ Consider all 111 possible configuration poses
4: $\quad$ **for all** *cp* $\in$ *allCPs* **do**
5: $\quad\quad$ *newPath* $\leftarrow$ *path*
6: $\quad\quad$ Append configuration pose *cp* to $newPath$
7: $\quad\quad$ **if** !MaxContactDurExceeded(*newPath*) **and** ValidatePose(*newPath.lastPose*) **then**
8: $\quad\quad\quad$ Append *newPath* to *newPaths*
9: $\quad\quad$ **end if**
10: $\quad$ **end for**
11: $\quad$ **return** *newPaths*
12: **end function**

In each iteration of Algorithm 1, the most promising pose sequence is extracted from the heap *activePaths* (line 6) and expanded with the addition of all valid successor poses (line 20). This path expansion, as outlined in Algorithm 2, considers both the validity of the appended pose (ValidatePose) and a limitation on how long a given contact may be kept (MaxContactDurExceeded), which is detailed in [4].

In [4], valid successor poses in ValidatePose (Algorithm 2, line 7) are determined from a manual definition of the locomotion task, where available hand contacts are provided as intervals of the straight distance from the start to the goal position. In contrast, with the combined approach introduced in this work, we propose to determine possible successor poses directly by considering affordances extracted from visual perception in the ValidatePose function.

C. The Combined Approach

The central idea of the combined approach is that the set of contact possibilities, required for pose sequence planning, can be generated from visually detected *supportability* affordances. The critical part of Algorithm 1 in this regard is the determination of suitable and valid nodes for expansion (see ExpandPath and ValidatePose in Algorithm 2), which in our previous work has been implemented via predefined intervals of support availability. Instead of relying on such a manual definition of possible contacts, the proposed approach queries affordance belief functions generated from visual affordance detection for available support possibilities. More formally, the perceptive-cognitive process for affordance detection produces a set of geometric primitives

$$\Psi = \{p_1, \ldots, p_K\}, \tag{4}$$

and a set of associated affordance belief functions Θ_a, which map end-effector poses to Dempster-Shafer belief assignments. Out of the set of possible affordances proposed in [1], we will focus on *supportability* affordances Θ_{Support} in this work. In order to enable the effective processing of affordance belief functions, geometric primitives $p \in \Psi$ are sampled into end-effector pose samplings $\mathcal{S}_p$, which are in contact with the primitive boundary ∂p:

$$\mathcal{S}_p = \{s_1, \ldots, s_N\}, \;\; s_i \in \partial p \subset SE(3). \tag{5}$$

Affordance belief functions express belief in the existence of an affordance with respect to an end-effector pose $\boldsymbol{x} \in SE(3)$. In the aspired application of this work however, the orientational aspect of $\boldsymbol{x}$ is neglected at this stage. The existence of an affordance is further assessed by evaluating the *expected probability* $E(\Theta_a(\boldsymbol{x}))$ [30] of the Dempster-Shafer valued affordance belief functions. These considerations allow the representation of generated affordance belief functions Θ_a as a point cloud $\mathcal{P}_{p,a,\lambda}$, which can be organized in an octree data structure for efficient spatial access:

$$\mathcal{P}_{p,a,\lambda} = \big\{\mathfrak{t}(\boldsymbol{x}) : \boldsymbol{x} \in \partial p, E(\Theta_a(\boldsymbol{x})) > \lambda\big\}. \tag{6}$$

The notion $\mathfrak{t}(\boldsymbol{x}) \in \mathbb{R}^3$ refers to the translational component of $\boldsymbol{x} \in SE(3)$. Validation of poses is performed within Algorithm 2 during path expansion and an expanded path is only considered valid if the added configuration pose can be validated, i. e. if this pose is considered kinematically feasible given the detected affordances of the scene. The affordance-based pose validation method is implemented in Algorithm 3. For a pose to be valid, all end-effectors which are used in supporting contacts (in this work, hands and feet) must be validated in a three-step process:

1) First (lines 3–8), the octree data structure is searched for suitable *supportability* affordances in the vicinity

Algorithm 3 Affordance-Based Validation of Poses

```
Require:
    affOct – Octree of affordance pointcloud
    maxCandDist – Maximum distance for support aff.
 1: function VALIDATEPOSE(pose)
 2:     for all end-effector ee used as a contact in pose do
 3:         ▷ Check availability of support affordances
 4:         eePos ← ee.pos                    ▷ Forward kinematics
 5:         affCand ← OCTREE-RADIUS-SEARCH(
                affOct, eePos, maxCandDist)
 6:         if affCand = ∅ then
 7:             return false          ▷ No support afford. found
 8:         end if
 9:
10:         ▷ Rank available support affordances based on
                expected probability and distance
11:         jointAnglesFound ← false
12:         scoredCand ← list()
13:         for all c ∈ affCand do
14:             eeDist ← ||c.pos − eePos||_2
15:             score ← E(Θ_Supp.(c.pos)) + (1 − eeDist/maxCandDist)
16:             Add (c.pos, score) to scoredCand
17:         end for
18:         Sort scoredCand by score
19:
20:         ▷ IK-based refinement of configuration pose
21:         for all c ∈ scoredCand do
22:             Use IK to find joint angles to reach c.pos
23:             if solution found then
24:                 Save joint angles for pose
25:                 jointAnglesFound ← true
26:                 Abort loop
27:             end if
28:         end for
29:         if !jointAnglesFound then
30:             return false            ▷ No joint angles found
31:         end if
32:     end for
33:     return true
34: end function
```

of the end-effector position. If no affordance is found within a distance of *maxCandDist*, the candidate pose is discarded.

2) Next (lines 11–18), available *supportability* affordances are ranked based on a combined metric that includes the distance of the affordance to the end-effector (lower is better) and the expected probability of the affordance (higher is better).
3) Finally (lines 21–31), the joint angles that are provided by the configuration pose are altered through inverse kinematics (IK), such that the end-effector reaches the position of the *supportability* affordance to be used. Affordances are tested with decreasing score and once the IK finds a solution to reach the given affordance, the computed joint angles are stored and the next end-effector is considered. Note that the locomotion displacement between different configuration poses is not altered in this process. In the case that none of the available *supportability* affordances is reachable, we deem the configuration pose invalid.

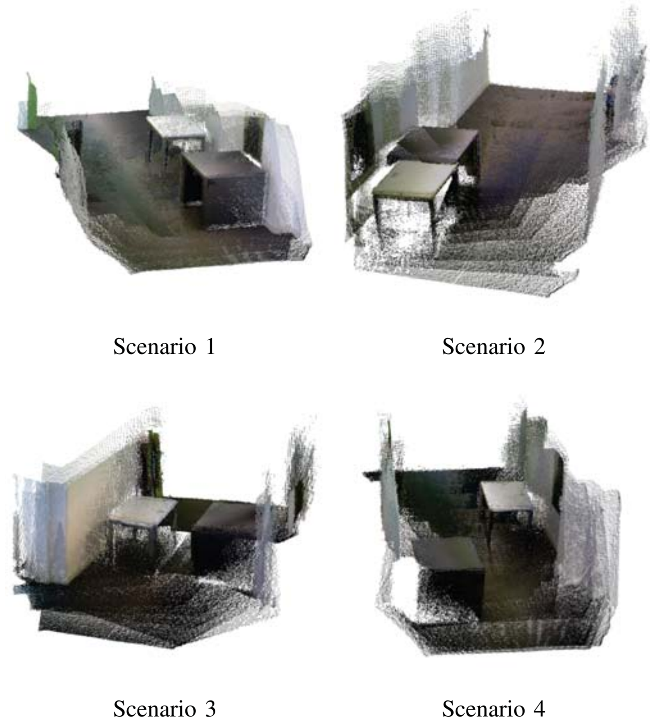

Fig. 3: Four evaluation scenarios for walking with support contact opportunities composed from different arrangements of tables in a hallway. The scenes are represented as registered point clouds.

IV. Evaluation

The proposed combination of loco-manipulation affordance detection and whole-body pose sequence planning is evaluated based on a set of four exemplary hallway scenarios with different arrangements of tables (see Fig. 3). The tables provide opportunities for supporting hand contacts along a defined locomotion path. The scenes, which are assumed to be entirely unknown to the robot, are represented as registered point clouds captured using an *ASUS Xtion Pro* sensor. In the evaluation scenarios, we employ the proposed approach to plan locomotion pose sequences with supporting end-effector contacts for the humanoid robot ARMAR-4. Fig. 4 shows the geometric primitives extracted from the scenes, as described in Section III-A. All considered examples define the target locomotion trajectory as a straight path along the hallway. The trajectories are chosen such that sequences of contact and non-contact phases with the same end-effector (Scenario 1a and Scenario 2), unreachable contact opportunities (Scenario 1b), as well as sequences of simultaneous or alternating contact phases with both end-effectors (Scenario 3 and Scenario 4) are captured in the

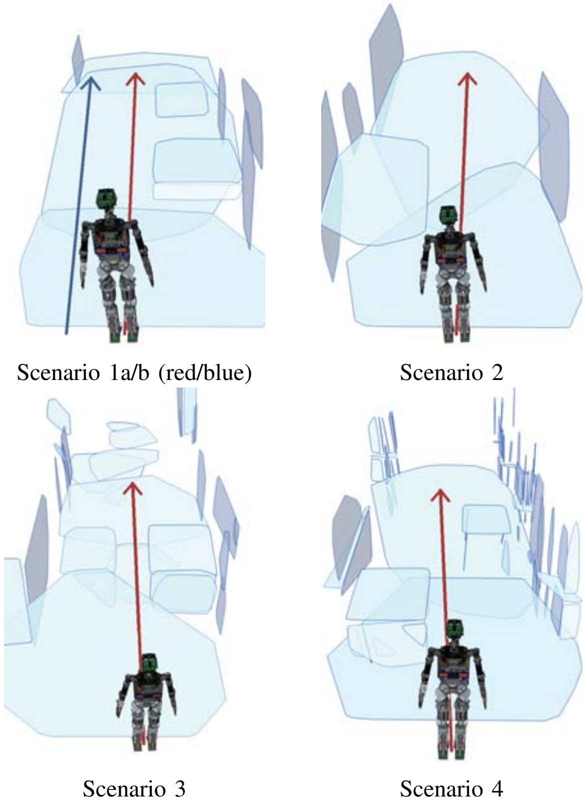

Fig. 4: Visualization of geometric primitives obtained from the evaluation scenarios shown in Fig. 3, including aspired straight locomotion trajectories.

evaluation. In all evaluation scenarios, the locomotion is defined to start and stop in the configuration pose *LFRF_1*, which represents a neutral double-foot support. The defined locomotion distance varies between $2\,\mathrm{m}$, $4\,\mathrm{m}$ and $6\,\mathrm{m}$.

For model training, the dataset from [4] is used, consisting of 137 human motion recordings from the *KIT Whole-Body Human Motion Database*[4] [31] that have been processed as described in Section III-B. These recordings represent walking motions, in which different *supportability* affordances from handrails and tables have been used during locomotion tasks. The employed dataset is symmetric with regard to left/right hand supports.

The proposed affordance-based pose sequence planner is able to successfully find pose sequences for ARMAR-4 with appropriate utilization of environmental support opportunities in all evaluated scenarios. The solution pose sequences generated for the evaluation scenarios 1a, 2, 3 and 4 are visualized in Fig. 5, where selected intermediate robot poses are depicted with end-effector contact indicated by red highlighting. The detected *supportability* affordance belief functions are visualized as green areas in the respective first pictures. The examples demonstrate that the affordance-based pose sequence planner is able to produce meaningful multi-contact poses for crucial points in a desired whole-body locomotion trajectory, which can be used for subdividing the subsequent trajectory planning into multiple computationally feasible subproblems, as explained in Section II-B.

Table I lists performance measurements for the evaluation scenarios for trajectory lengths of $2\,\mathrm{m}$, $4\,\mathrm{m}$ and $6\,\mathrm{m}$, respectively. In order of appearance, the columns of Table I represent the runtime of the affordance detection process (excluding the generation of geometric primitives), the resulting sizes of Θ_{Support}, the runtime of the affordance filtering process from Eq. 6, the runtime of the octree generation, the defined locomotion distance, the number of iterations and validation steps performed by the planner, the average runtime of the expansion step outlined in Algorithm 2 and the runtime of the planning process. Note that the first four values only depend on the scenario, not on the locomotion distance. The figures show that affordance-based pose sequence planning is computationally feasible with a maximum total runtime of about $16\,\mathrm{s}$ in scenario 2. The planning duration roughly corresponds to the intuitive problem complexity and the choice of the pruning parameterization (*pruneThresh* and *prunePeriod* in Algorithm 1). Since the employed pruning strategy unconditionally purges paths whose modeled translation falls behind the current path by a static threshold, large areas of continuous contact opportunities (such as the two adjacent tables left to the robot in scenario 2), which increase the number of possible successor poses, lead to larger heap sizes and thus to longer planning times.

V. Conclusion and Future Work

In this work, we proposed a novel approach towards affordance-based planning of whole-body multi-contact locomotion actions for humanoid robots in unknown environments. We tackle this problem by extending and combining two previously proposed methods for the detection of affordances in unknown environments and for the planning of whole-body pose sequences based on n-gram pose transition models learned from human motion data. Our approach produces a sequence of whole-body poses which consider environmental contact opportunities as indicated by detected affordances. The approach has been implemented and evaluated using four exemplary scenarios for multi-contact locomotion based on real sensor data.

In our future work, we will continue to pursue the proposed approach and aim at completing it by implementing methods for whole-body motion planning and whole-body control based on the affordance-informed generation of whole-body pose sequences. We further intend to apply the concepts, which are evaluated in a kinematic simulation at this point, to dynamically simulated robots and to the real humanoid robot ARMAR-4, which requires means for the dynamic stabilization of generated motions. The consideration of dynamic motions further raises the challenging question, how dynamic properties of human motion demonstrations can be appropriately transferred to the robot.

[4]The motions can be found at `https://motion-database.humanoids.kit.edu/details/motions/<ID>/` with ID $\in \{395, 396, 677, 678, 679, 681, 705, 724\}$.

Fig. 5: Solution paths for the humanoid robot ARMAR-4 in the evaluation scenarios 1a (*first row*), 2 (*second row*), 3 (*third row*) and 4 (*third row*). See Fig. 3 and Fig. 4 for descriptions of the scenario setups. The affordance belief function Θ_{Support} is visualized in the leftmost pictures, in which support contact opportunities are highlighted green. Note that the presented solutions are sequences of configuration poses with end-effector contact information, not continuous motion trajectories.

TABLE I: Runtimes of the different steps of the proposed approach in four scenarios with a total amount of 15 configurations.

Scenario	Aff. Extr.	Aff. Sampling Sz.	Aff. Point Cloud	Octree	Distance	#Iterations	#Validations	Expansion	Planning
Scenario 1a	250 ms	707,760	173 ms	9.17 ms	2 m	1,044	5,024	4.19 ms	4.41 s
					4 m	2,573	12,773	4.21 ms	10.96 s
					6 m	4,690	22,174	2.65 ms	12.76 s
Scenario 1b	253 ms	707,760	205 ms	9.33 ms	2 m	1,065	5,363	1.15 ms	1.25 s
					4 m	1,001	5,205	1.20 ms	1.23 s
					6 m	2,811	14,203	1.42 ms	4.12 s
Scenario 2	296 ms	906,572	276 ms	9.52 ms	2 m	2,001	9,573	5.76 ms	11.62 s
					4 m	3,708	16,569	3.40 ms	12.77 s
					6 m	4,712	21,465	3.22 ms	15.43 s
Scenario 3	358 ms	660,856	221 ms	5.35 ms	2 m	1,007	4,894	3.37 ms	3.42 s
					4 m	1,009	4,802	2.15 ms	2.23 s
					6 m	2,008	9,558	2.01 ms	4.16 s
Scenario 4	533 ms	934,268	295 ms	8.26 ms	2 m	459	2,211	1.67 ms	0.78 s
					4 m	1,892	9,505	2.30 ms	4.43 s
					6 m	2,400	11,647	1.82 ms	4.51 s

REFERENCES

[1] 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 (IROS)*, 2016.

[2] 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 Int. Conf. on Humanoid Robots*, 2016.

[3] K. Bouyarmane and A. Kheddar, "Humanoid robot locomotion and manipulation step planning," *Advanced Robotics*, vol. 26, no. 10, pp. 1099–1126, 2012.

[4] C. Mandery, J. Borràs, M. Jöchner, and T. Asfour, "Using language models to generate whole-body multi-contact motions," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 5411–5418, 2016.

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

[6] 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.

[7] 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 Transactions on Cognitive and Developmental Systems*, 2016.

[8] 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.

[9] 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.

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

[11] 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.

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

[13] L. Saab, O. E. Ramos, F. Keith, N. Mansard, P. Soueres, and J.-Y. Fourquet, "Dynamic whole-body motion generation under rigid contacts and other unilateral constraints," *IEEE Transactions on Robotics*, vol. 29, no. 2, pp. 346–362, 2013.

[14] J. Schulman, Y. Duan, J. Ho, A. Lee, I. Awwal, H. Bradlow, J. Pan, S. Patil, K. Goldberg, and P. Abbeel, "Motion planning with sequential convex optimization and convex collision checking," *The International Journal of Robotics Research*, vol. 33, no. 9, pp. 1251–1270, 2014.

[15] Y. Tassa, T. Erez, and E. Todorov, "Synthesis and stabilization of complex behaviors through online trajectory optimization," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 4906–4913, 2012.

[16] T. Erez and E. Todorov, "Trajectory optimization for domains with contacts using inverse dynamics," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4914–4919, 2012.

[17] M. Posa, C. Cantu, and R. Tedrake, "A direct method for trajectory optimization of rigid bodies through contact," *The International Journal of Robotics Research*, vol. 33, no. 1, pp. 69–81, 2014.

[18] S. Dalibard, A. Nakhaei, F. Lamiraux, and J.-P. Laumond, "Whole-body task planning for a humanoid robot: A way to integrate collision avoidance," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, pp. 355–360, 2009.

[19] J. Kuffner, K. Nishiwaki, S. Kagami, M. Inaba, and H. Inoue, "Motion planning for humanoid robots under obstacle and dynamic balance constraints," in *IEEE International Conference on Robotics and Automation (ICRA)*, vol. 1, pp. 692–698, 2001.

[20] A. Settimi, D. Caporale, P. Kryczka, M. Ferrati, and L. Pallottino, "Motion primitive based random planning for loco-manipulation tasks," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, pp. 1059–1066, 2016.

[21] Y. Zhang, J. Luo, K. Hauser, R. Ellenberg, P. Oh, H. A. Park, M. Paldhe, and C. G. Lee, "Motion planning of ladder climbing for humanoid robots," in *IEEE International Conference on Technologies for Practical Robot Applications (TePRA)*, pp. 1–6, 2013.

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

[23] A. Escande, A. Kheddar, and S. Miossec, "Planning contact points for humanoid robots," *Robotics and Autonomous Systems*, vol. 61, no. 5, pp. 428–442, 2013.

[24] J. Vaillant, A. Kheddar, H. Audren, F. Keith, S. Brossette, A. Escande, K. Bouyarmane, K. Kaneko, M. Morisawa, P. Gergondet, E. Yoshida, S. Kajita, and F. Kanehiro, "Multi-contact vertical ladder climbing with an hrp-2 humanoid," *Autonomous Robots*, vol. 40, no. 3, pp. 561–580, 2016.

[25] K. Bouyarmane and A. Kheddar, "Using a multi-objective controller to synthesize simulated humanoid robot motion with changing contact configurations," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4414–4419, 2011.

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

[27] C. Mandery, O. Terlemez, M. Do, N. Vahrenkamp, and T. Asfour, "Unifying representations and large-scale whole-body motion databases for studying human motion," *IEEE Transactions on Robotics*, vol. 32, no. 4, pp. 796–809, 2016.

[28] C. Mandery, J. Borràs, M. Jöchner, and T. Asfour, "Analyzing whole-body pose transitions in multi-contact motions," in *IEEE/RAS International Conference on Humanoid Robots*, pp. 1020–1027, 2015.

[29] A. Stolcke, "Srilm – an extensible language modeling toolkit," in *International Conference on Spoken Language Processing*, vol. 2, pp. 901–904, 2002.

[30] A. Jøsang, "A logic for uncertain probabilities," *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, pp. 279–311, 2001.

[31] C. Mandery, O. Terlemez, M. Do, N. Vahrenkamp, and T. Asfour, "The kit whole-body human motion database," in *International Conference on Advanced Robotics (ICAR)*, pp. 329–336, 2015.