Affordance-Based Grasping and Manipulation in Real World Applications

Christoph Pohl*, Kevin Hitzler*, Raphael Grimm, Antonio Zea, Uwe D. Hanebeck and Tamim Asfour

***Abstract*— In real world applications, robotic solutions remain impractical due to the challenges that arise in unknown and unstructured environments. To perform complex manipulation tasks in complex and cluttered situations, robots need to be able to identify the interaction possibilities with the scene, i. e. the affordances of the objects encountered. In unstructured environments with noisy perception, insufficient scene understanding and limited prior knowledge, this is a challenging task. In this work, we present an approach for grasping unknown objects in cluttered scenes with a humanoid robot in the context of a nuclear decommissioning task. Our approach combines the convenience and reliability of autonomous robot control with the precision and adaptability of teleoperation in a semi-autonomous selection of grasp affordances. Additionally, this allows exploiting the expert knowledge of an experienced human worker. To evaluate our approach, we conducted 75 real world experiments with more than 660 grasp executions on the humanoid robot ARMAR-6. The results demonstrate that high-level decisions made by the human operator, supported by autonomous robot control, contribute significantly to successful task execution.**

I. Introduction

The decommissioning of nuclear power plants is one of the most challenging problems that affect many countries around the world. A report of the International Atomic Energy Agency from 2019 shows that by the end of 2018, there were 451 operational, 55 incomplete, 81 planned and 172 permanently shutdown reactors worldwide [1]. In total, 144 of these shutdown reactors have been in a decommissioning process while only 5 of them have reached the final decommissioning phase. Until now, most of the involved processing steps cannot be automated as they require complex object manipulation by humans with expert knowledge, as e. g. the decontamination of plant components. Due to the fact that these humans need to deal with radioactive material, they are often exposed to hazardous and exhausting working conditions, i. e. they suffer from a high cognitive load and are forced to work under strict safety restrictions such as wearing whole-body protective suits [2]. Furthermore, most of these tasks need to take place directly inside the partially decommissioned nuclear power plants. This means that people have to work in a strongly restricted working environment with limited space that cannot easily be modified for automation. The precarious working conditions, the requirement to work in a human-designed environment with limited space as well as the need for complex manipulation tasks makes the decommissioning process a suitable application for humanoid robots. As humanoid robots are designed to work in human-centered environments, they are able to move around, adapt to changing working conditions and fulfill tasks with physical abilities comparable to humans.

Fig. 1. The humanoid robot ARMAR-6 grasping and placing unknown objects in a cluttered scene in context of a nuclear decommissioning task.

However, to perform complex manipulation tasks, robots need to be able to identify interaction possibilities within the environment, i. e. they need to detect object affordances [3]. In unstructured environments with noisy perception, unknown objects and limited prior knowledge, this poses a major challenge. The considered scenario in this work deals with dismantled plant components of a nuclear power plant that have to be picked up from a box and placed at a predefined location before they get decontaminated by humans. As these objects have arbitrary shapes, colors, textures and occur in any given configuration, they can be considered as completely unknown. Therefore, robust approaches are required that enable grasp affordance extraction and execution by a humanoid robot in cluttered scenes.

In this work, we present a robust approach for grasping unknown objects in cluttered scenes with a humanoid robot in the context of a nuclear decommissioning task. Our approach includes an autonomous method for extracting and selecting grasp affordances and a semi-autonomous method with an intuitive human-robot interface that can be used by the human operator to select extracted affordances in virtual reality (VR). While the former does not require any decision-making by humans, the latter allows combining the advantages of an autonomous robot with the expert knowledge of

The research leading to these results has received funding from the German Federal Ministry of Education and Research (BMBF) under the competence center ROBDEKON (13N14678).

The authors are with the Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Karlsruhe, Germany. {pohl, hitzler, raphael.grimm, antonio.zea, uwe.hanebeck, asfour}@kit.edu

*The first two authors contributed equally to this work

an experienced human worker. As a result, the operator is always able to actively intervene in the scene. In context of nuclear plant decommissioning, this can be very helpful as a worker can, for instance, decide *where* and *which* object to grasp for decontamination. At the same time, the human operator can focus on the important high-level decisions and is no longer exposed to the high cognitive load caused by the complexity of robot control. To evaluate our approach in a real world scenario, we conducted 75 experiments on the humanoid robot ARMAR-6 (see Figure 1) and compared three different *operational modes*: a *manual* mode in which grasp affordances are defined by a human operator with a 2D mouse-screen interface, a *semi-autonomous* mode where the operator selects the affordances in VR and an *autonomous* mode in which the robot makes decisions based on a predefined heuristic. Our results show that a semi-autonomous execution reduces the number of grasp failures drastically, compared to a fully autonomous and manual operational mode.

In the following sections, we provide an overview of existing real world applications and describe how the general concept of affordances can be used and integrated into the scenario of nuclear decommissioning tasks. After looking into the implementation details, we discuss the robustness of our approach by evaluating the experiment results.

II. Related Work

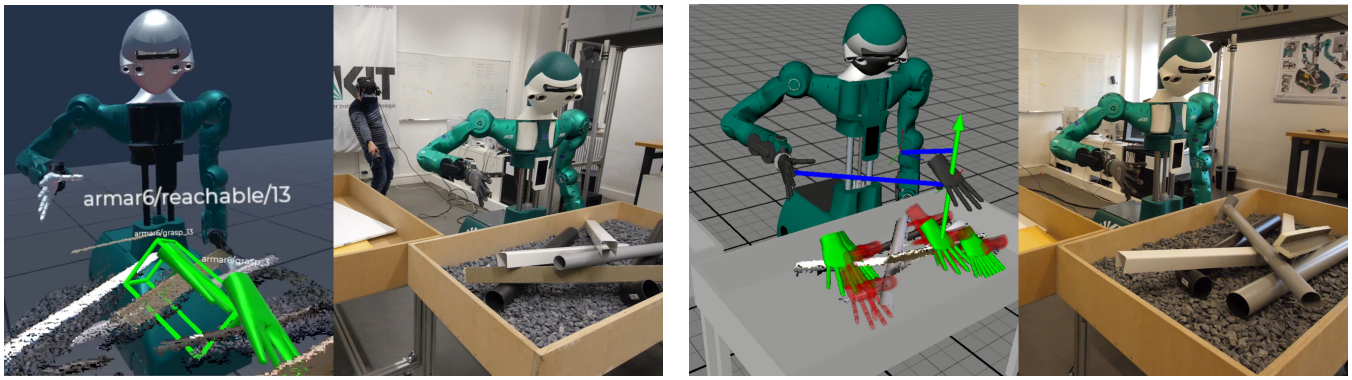
For most real world applications in industrial or disaster response scenarios, robotic solutions remain hard to realize due to the many challenges that arise in unstructured and unknown environments. Previous studies of humanoid robots in real world applications have shown that they are well suited for human-centered environments, such as disaster response scenarios [4], [5], [6], industrial manufacturing [7], [8] and collaborative maintenance and repair tasks [9], as they are versatile, mobile and can use tools designed for humans to fulfill complex loco-manipulation tasks. To accomplish this, they need to understand which actions can be performed in unknown scenarios that lead to successful task execution, i. e. they need to recognize task-related affordances [3], [10]. However, as perceptual uncertainty in unknown environments increases the complexity of the system drastically and might lead to failures, more robust methods are required to use humanoid robots in real world scenarios.

The humanoid robot ARMAR-6 [9], for example, was particularly designed for collaborative maintenance and repair tasks in industrial warehouses. The high degree of autonomy allows the robot to recognize the need for help, learn new tasks from human observation and support workers by performing complex tasks such as grasping, mobile manipulation and bimanual collaboration. Although many of these tasks can be executed in a robust way, failures can occur in different situations that need to be recovered. In unknown scenarios with no prior knowledge about the environment, this becomes very challenging. This can also be seen in the work of [8], where the position-controlled humanoid robot HRP-4 and the torque-controlled TORO are used for autonomous assembly tasks in aircraft manufacturing. Initial concerns for the stability of bipedal robots were outweighed by their improved versatility and smaller size compared to multi-legged solutions. The experiment results show that humanoid robots can even operate in complex scenarios, however, there are still limitations regarding robust perception and failure recovery strategies which, e. g. in case of falling, can even lead to unrecoverable hardware failures.

To overcome the challenges of fully autonomous robot control, researchers look into semi-autonomous or teleoperated solutions for real world applications. In the well-known DARPA Robotics Challenge (DRC) [4], for example, the participants had to tackle eight tasks to investigate the efficiency of supervised robotics in disaster response scenarios. The authors of [11] report that human-robot interaction had a considerable influence on the performance of the teams. The two top-ranked teams at DRC [12], [13] employed bipedal humanoid robots to handle difficult tasks in unstructured environments. However, due to even higher complexity, applications of humanoids to actual disaster sites were still out of reach [14].

In context of nuclear decommissioning and decontamination, the work of [15] shows that the current robotic solutions used in the industry have little or no autonomy, most of them being remotely controlled via master-slave systems or manual teleoperation. These systems often require a high bandwidth of communication between the human operator and the robot. However, this cannot always be provided in disaster and nuclear decommissioning sites. Furthermore, a pilot study in the work of [16] shows that an autonomous solution can increase the reliability of the task execution in decommissioning scenarios, while decreasing the cognitive load of the operator. Similar conclusions are drawn in [17] which confirms that a semi-autonomous solution should be preferred over manual teleoperation in order to exploit the expert knowledge of an operator in the form of high-level decision making. To reduce the high cognitive load on human operators, [6] implement a full-body suit for manual teleoperation as well as different methods for autonomous manipulation of familiar objects. The experiments show that some tasks could be accomplished autonomously while others still required the utilization of a 6D mouse interface for precise manipulation. The work of [18] presents a teleoperational framework that includes a low-level real-time control for manipulation and high-level teleoperation for locomotion, using VR and motion tracking. The proposed system of [19] only depends on a VR-headset in combination with a motion controller to track the hand-pose of the operator for remote control of a robotic gripper. However, both approaches rely on fast data transmission and low-level teleoperation for manipulation which puts high cognitive load on the human operator. Thus, they are less suitable for disaster response or decommissioning tasks.

To overcome these challenges, we propose a robust approach that makes use of both an autonomous method for affordance extraction and selection and a semi-autonomous interface for intuitive affordance selection in VR. In our



(a) (b)

Fig. 2. Affordance selection by a human operator in the *semi-autonomous* mode (a) and by the robot in the *autonomous* mode (b). The grasp affordances are extracted based on visual information from the scene and provided to the operator and the robot, respectively. Hands that are displayed in green represent grasp candidates that are reachable by the robot, while candidates that are not reachable from the robot's position are displayed in red.

previous work [20], a pilot interface for the teleoperated selection of whole-body affordances was presented that increases the autonomy of the system while decreasing the communication bandwidth as well as the operator's cognitive load. In this work, we build on the concept of affordance-based teleoperation and show how this can improve the task execution performance in complex grasp scenarios in the context of nuclear decommissioning tasks with unknown objects in cluttered scenes. As a result, the human operator is able to operate from a safe distance and can focus on important high-level decisions such as the decontamination sequence of objects, without having to worry about robot control.

III. System Setup

A. Decommissioning Scenario

During the decommissioning of a nuclear power plant, surfaces and objects are checked for residual radiation before they can be decontaminated and approved for disposal. In an example workflow, human workers divide the relevant objects into pieces of a maximum size of $80 \times 80\,\text{cm}$ and put them into a standardized box. These boxes will then be transported to the decontamination facilities where the objects are picked up from the box and subsequently cleaned, e. g. using high pressure water jetting. After a clearance measurement, the objects can be further processed in the regular waste disposal system.

To evaluate the industrial feasibility of robotic solutions in a similar decommissioning procedure, this work considers the subtask of retrieving the unknown, dismantled objects from the delivered box and transporting them for further processing. The evaluated scenario consists of grasping unknown objects from a cluttered box and placing them into a second, empty box. The task is executed on the humanoid robot ARMAR-6 in a static indoor environment and implemented using the robot development environment *ArmarX* [21]. An example setup can be seen in Figure 1. In order to draw meaningful conclusions about how much teleoperation is required in such a real world application, the experiments are performed in three different operational modes with varying degrees of autonomy: *manual*, *semi-autonomous* and *autonomous*.

There is no prior knowledge used for grasping the unknown objects, except the position and dimensions of the boxes. All affordances are derived from the images and point clouds of an RGB-D camera in the head of the robot. The remote interaction of a human operator with the scene in the *semi-autonomous* mode is realized by a teleoperation interface in VR, where the user can choose between multiple grasp affordances. This intuitive remote control of the robot allows for the intervention in case of failures and reduces the mental strain on the human operator.

B. Virtual Reality Interface

The objective of VR-assisted teleoperation is to allow a user to control a robot and interact with the robot's environment immersively, while physically being in a different place - possibly several kilometers away. A key challenge is letting the user know about the robot's surroundings, especially in scenarios with sparse data and unreliable communication. Usually, the user can only observe the robot's workspace from sensor data provided by the robot, e. g. video streams, laser scans, or point clouds. Given that most of these sensors are located on the robot's head, a straightforward solution would be to stream the camera images directly onto the head-mounted display of the operator. However, this approach limits the user's point of view to the robot's head, and even a binocular view would not allow the user to fully exploit the captured depth information.

Instead, we chose to project the RGB-D data into the three-dimensional world as a point cloud, together with a 3D model of the robot with correct pose and joint positions. This permits the user to move freely and choose the virtual point of view that best suits them. The projected sensor data takes the robot's pose into account, i. e. objects in the point cloud remain fixed in place even if the robot is moving. Then, by decoupling the robot's head from the operator's virtual point of view, we obtain three key advantages: First, users can

perceive depth information optimally, in particular curvatures and corners, by viewing the point cloud from any angle they choose. Second, the robot can move at any moment and at any speed without dragging or affecting the user's point of view, avoiding confusion. Third, the virtual world is rendered locally and decoupled from the sensor data stream. This means that if the user moves their head, they will see the results immediately, even if communication with the robot suffers from high latency and the point cloud updates slowly. This last point is critical for minimizing motion sickness, one of the key challenges in VR environments.

IV. Approach

In our approach, we distinguish between three cases for affordance extraction and selection. Each of them has a different level of autonomy and thus, needs to be regarded separately. To clearly differentiate between these cases, we refer to them as *operational modes* that can be described as follows:

1) *Manual mode*: Based on the perceived point cloud of the robot, a human operator manually sets an affordance in the scene. The corresponding action for the defined affordance is directly executed on the robot. Consequently, there is no autonomy included regarding affordance extraction or selection.
2) *Semi-Autonomous mode*: In the semi-autonomous mode, the affordances are extracted autonomously by the robot and visualized in VR (Figure 2a). A human operator then selects one of the given affordances and thus, decides upon the next action, e. g. where and which object to grasp next that is executed on the robot.
3) *Autonomous mode*: In this mode, the affordance extraction and selection is done fully autonomously by the robot without human interaction (Figure 2b). During action execution, a human operator is still able to stop the execution if necessary (e. g. in case of failure).

A complete overview of our affordance extraction and execution pipeline, including the different operational modes, is given in Figure 3. In the following sections, we provide a detailed description of each individual step in the pipeline.

A. Preprocessing and Segmentation

In the first step (I.), the point cloud of the RGB-D camera is registered and transformed into the robot's coordinate frame. As the location of the box is known in advance, the point cloud can be filtered such that all points lying outside the box or belonging to the box itself are cropped. After preprocessing, the point cloud contains only the visible parts of the objects.

Since we need to cope with arbitrary configurations of unknown objects in a box, the segmentation must be versatile and robust against clutter. The best results regarding geometrical consistency were obtained with euclidean clustering, followed by a region growing algorithm. In many cases, this results in an oversegmentation and thus, parts of the objects that are separated through occlusion are assigned to different labels. Additionally, there is no guarantee that the segments are temporally consistent over multiple camera frames. A segmented object part can have two completely unrelated labels in two successive point clouds. These factors need to be considered in the following steps.

B. Affordance Extraction

The affordance extraction step (II. and III.) differs depending on the used *operational mode*. In the *manual mode*, affordances are defined directly by the operator that draws a line in the received camera image in a 2D mouse-screen interface to identify the region of interest in which a grasp should be performed. More specifically, this line is used as a mask to select the points corresponding to an object in the scene. The first two principal components of this selected area are then used to determine the orientation of the grasp affordance. To grasp an object in a robust way, the pose of the hand is oriented orthogonally to the first and along the second principal component as shown in Figure 3.

In the *semi-autonomous* and *autonomous* mode, the affordance extraction is based on the labeled point cloud and thus, depends on the preprocessing and segmentation from the previous step. For each identified segment, the *Object-Oriented Bounding Boxes* (OOBB) and a principal component analysis (PCA) is computed as shown in Figure 3, III. Following this, multiple grasp candidates are generated in fixed intervals and along the largest axis of the OOBB. If the two largest axes of the OOBB have the same length, a random axis among them is chosen for the generation of candidates. From the PCA, grasp candidates are extracted in the same way as manual grasp affordances, i. e. orthogonal to the first principal component. To account for sensor noise and temporal inconsistencies in the segmentation, the generated affordances are filtered over time. Grasp candidates of the last 10 timesteps are stored and clustered in the local spatio-temporal neighborhood. Clusters with less than 5 members are discarded as outliers produced by noise. For the remaining clusters, the grasp candidate representing the spatial median pose is selected as cluster representative. Afterward, the actions corresponding to the resulting grasp affordances are checked for self-collision and reachability with respect to the current robot pose and the affordance is labeled accordingly.

C. Affordance Selection

The goal of the affordance selection is to choose the next affordance, which is linked to an action to be executed by the robot, i. e. *where* and *which* object to grasp next. Similar to the previous step, the affordance selection depends on the chosen operational mode. In the *manual mode*, no further selection is required as only a single affordance is generated in the pipeline (II.). In the *semi-autonomous* and *autonomous* mode, however, the affordance is either selected by the human operator (IV.) or by the robot (V.), respectively.

In the *semi-autonomous* mode (IV.), this is realized in an intuitive way through a VR interface as shown in Figure 2. Available affordances are visualized in the form of colored virtual "hands" at the corresponding position in

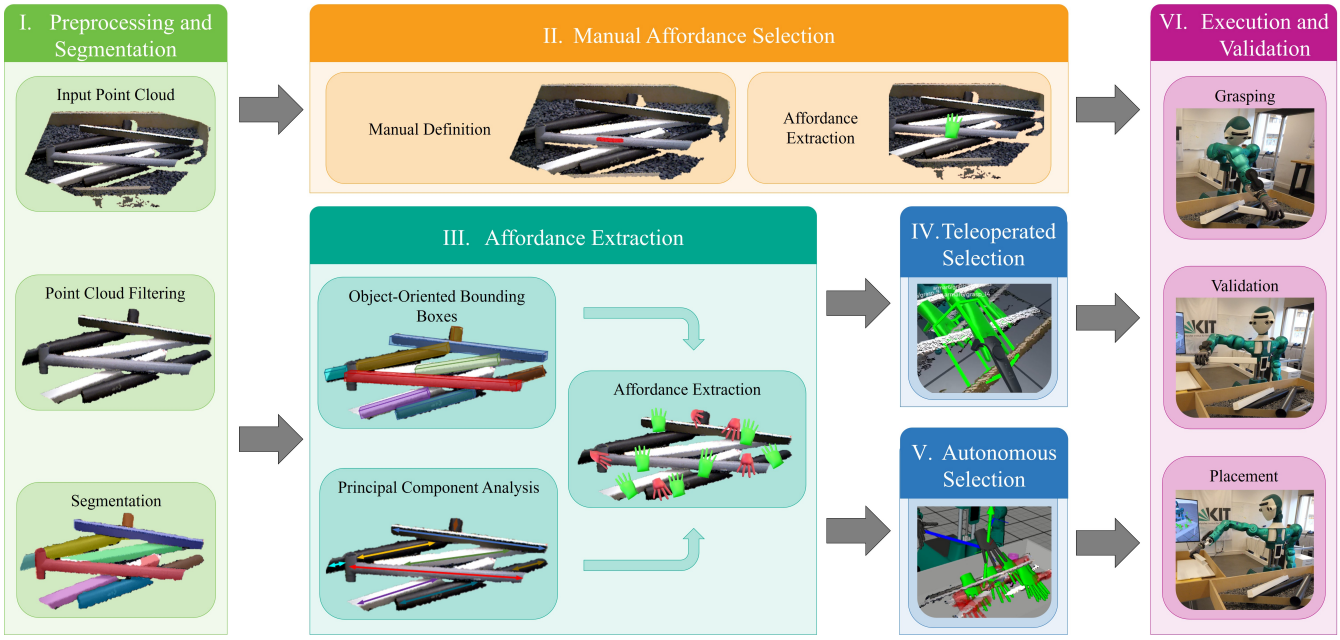


Fig. 3. The affordance extraction and execution pipeline: In a first step, the input point cloud is preprocessed and points belonging to the box itself are removed before a segmentation is performed (I.). In a second step, the grasp affordances are either set manually by a human operator (II.) or extracted fully autonomously (III.), given the segmented point cloud. The resulting affordances are then selected semi-autonomously by the operator in virtual reality (IV.) or autonomously by the robot (V.), depending on the operational mode. Finally, the corresponding action is executed by the robot (VI.).

the scene, where reachable grasp affordances are displayed with green color and affordances that are not executable from the current robot position are colored red. The list of available affordances is updated regularly, based on the frame rate and processing time in the affordance extraction component. During the affordance selection, the operator is able to remove all existing affordances by pressing a button on a hand-held controller and wait for new affordances. The human-robot interface then becomes very intuitive and easy to use: First, the human operator waits for extracted affordances to appear in VR. In the meantime, the operator can move around in the virtual scene in a third-person view to perceive the scene. To make a decision, a virtual hand can be selected by pointing to the target and pressing a trigger on the controller. The robot will then execute the action associated with the selected affordance. After finishing the task, the robot clears and provides a new list of affordances as the scene changes due to physical interaction.

In case of *autonomous* affordance selection (V.), a simple heuristic is used that includes two conditions: First, the highest grasp affordance in the scene is selected since we assume that the highest objects are easier to grasp than those occluded or blocked by other objects. The second condition compares the position of the newly selected affordance to the positions of the last three grasp executions if there are any previous grasp attempts that failed. If the chosen affordance is too close to a previously failed grasp, it is neglected in the selection process. This prevents the robot from trying to grasp the same object again, after failing in previous grasp executions. A comparison of the *semi-autonomous* and *autonomous* affordance selection is depicted in Figure 2.

D. Execution and Validation

Once a grasp affordance is selected, the execution and validation step (VI.) is identical for all operational modes. First, the robot moves to a pre-pose, a position and orientation directly above the corresponding target object, using velocity control. The pre-pose is calculated by adding a fixed distance value along the z-axis to the grasp pose. After that, the robot moves towards the target until contact with the object is established, which is detected by the 6D force-torque sensor in the wrist. As soon as a certain force-torque threshold is exceeded, the robot follows a predefined grasp trajectory to close its hand. After the object is fully grasped, the robot moves its hand to the final location using Via-Point Movement Primitives (VMP) [22] to place the object in the second box.

During execution, the operator is always able to abort the current action of the robot, e.g. if a grasp execution fails due to slippage or unstable grasping. In this case, the hand automatically opens again and the robot arm returns to the initial position, where the process starts from the beginning. Additionally, the robot stores the object information such as position and orientation in its memory so that it can be considered in the following *autonomous* affordance selection (see subsection IV-C).

V. Experiments

To evaluate our approach and the level of autonomy required in a real world scenario for successful task completion, we conducted 75 experiments on the humanoid robot ARMAR-6 with more than 660 grasp executions.

Affordance selection	Grasp executions	Successful grasps	Successful grasps [%]	Failed grasps	Failed grasps [%]	Grasp executions	Successful grasps	Failed grasps	Duration [min]
	Total					Mean ± Standard deviation			
Manual	227	150	66%	77	34%	9.1 ± 2.9	6.0 ± 0.0	3.1 ± 2.9	4:59 ± 1:21
Semi-Autonomous	199	143	72%	56	28%	8.0 ± 2.5	5.7 ± 0.5	2.2 ± 2.6	5:28 ± 1:32
Autonomous	236	135	57%	101	43%	9.4 ± 2.9	5.4 ± 0.8	4 ± 3.1	5:22 ± 1:15

Fig. 4. Results of the grasp selection experiments on ARMAR-6, with six objects placed in the box in each of the 25 experiments per operational mode. This results in a total of 150 objects possible to grasp in the *manual, semi-autonomous* and *autonomous* mode. On the left side, the statistics are accumulated over all 25 experiments for each affordance selection method, i.e. the *operational mode*. The right side shows the mean and standard deviation of the success and failure rates for a single series of grasping the six unknown objects in the box. The color of the entries corresponds to the relative quality of the results for one evaluation criterion. A darker shade of blue indicates better results.

A. Experimental Setup

In every experiment, the robot has to pick up six unknown objects from a cluttered box and place them into another. For each operational mode (*manual*, *semi-autonomous* and *autonomous*) we conducted 25 experiments, i. e. 150 objects that need to be grasped. To be consistent over all operational modes, we used the same six (unknown) objects throughout the experiments, consisting of construction materials with different shapes, colors and sizes such as pipes, connectors and cable canals, which were randomly arranged by a human to represent a cluttered scene in a box. Due to the fact that the cluttered scene is human-made and randomly arranged, it is not easily reproducible and differs between every experiment. In total, this results in 450 objects to be grasped by the robot over all experiments.

The experiments are evaluated based on the number of grasp attempts required to empty the box. To limit the number of executions, an experiment was canceled after a maximum of 10 failures or an execution time of 10 minutes. In addition, an experiment was canceled if the objects were no longer reachable for the robot, i.e. the robot could not reach the remaining objects. The number of successful and failed grasps was counted by a neutral person (i. e. a referee) and independent of the human operator. Throughout the 25 experiments in each operational mode, the person acting as a human operator stays the same. A grasp attempt was considered as successful if, and only if, an object was grasped, manipulated and successfully placed in the second box and as a failure otherwise.

B. Experimental Results

The experiment results are given in Figure 4. In case of *manual* affordance selection, all 150 objects could be grasped successfully after 227 grasp executions, which results in a total success rate of 66% and a failure rate of 34% over 25 experiments. This means that, on average, 9.1 ± 2.9 grasp attempts were required to empty a box with six objects. As the manual affordance selection was able to grasp all six objects in each experiment, the mean of successful grasps is 6.0 whereas the mean failed grasps are given by 3.1 ± 2.9. The rate of successful grasps shows that, even in difficult scenarios, our system is very robust regarding grasp execution, since all objects could successfully be grasped in the experiments. The high number of grasp executions and failures, however, implies that many interactions with the scene were required before an object could be grasped successfully. This could be due to the fact that in the *manual* affordance selection, the robot directly executes the grasp affordance selected by the human without checking for collision and reachability which can result in a failure.

In the *semi-autonomous* mode, 143 of the 150 objects were grasped successfully after 199 grasp executions. Consequently, 72% of all selected actions were successful. On average, 8.0 ± 2.5 grasp executions were required in every single experiment to empty the box with 6 objects, while 5.7 ± 0.5 grasp executions are successful and 2.2 ± 2.6 result in failures. The results of the *semi-autonomous* affordance selection show that the high-level decisions and expert knowledge of the operator have a significant impact on the successful outcome of the experiments. Besides the high number of successful grasps, the teleoperation resulted in very few failed grasp attempts. The ability of the operator to perceive the scene from different perspectives as well as the intuitive and automated interaction with the scene presents a good trade-off between autonomy and manual teleoperation. Nevertheless, in some cases, the objects ended up in unreachable positions, where either no valid grasp affordances could be extracted or the execution of the available action would have resulted in a potential collision with the box.

The results of the *autonomous* affordance selection show that 135 of 150 objects could be grasped successfully in a total of 236 executions. On average, 5.4 ± 0.8 grasp executions were successful in each experiment and 4 ± 3.1 resulted in failure. In other words, 9.4 ± 2.9 grasp executions are required to grasp all six objects in the box. The experiments demonstrate that the heuristic used to autonomously select the next affordance is able to cope with simple object arrangements in the scene. In the case of very complex and

cluttered scenes, however, selecting the highest grasp in the scene is not always the best choice, as these objects might be blocked by underlying objects or not easily graspable by the robot. Consequently, an attempt to grasp that object could fail and make the scene even more complex such that objects are no longer reachable for the robot.

The mean execution times of $4{:}59 \pm 1{:}21$ (*manual*), $5{:}28 \pm 1{:}32$ (*semi-autonomous*) and $5{:}22 \pm 1{:}15$ (*autonomous*) show that the *manual* affordance selection takes less time than the *semi-autonomous* and *autonomous* selection. This is due to the fact that in the latter two cases, the affordances are extracted fully autonomously by the robot. In these operational modes, there is an additional waiting time to ensure that enough affordances are extracted and can be filtered over time while in the *manual* case, the affordances were directly defined by the human operator with a 2D mouse-screen interface and do not require additional waiting time. Based on our observations during the experiments, we can further conclude that the high execution time in the *semi-autonomous* mode is related to the high flexibility in virtual reality which allows the user to freely move around and choose different virtual points of view, i.e. the human operator first evaluated the scene in greater detail in the virtual 3D environment before making a decision.

C. Comparison and Summary

The experiments show that the *semi-autonomous* affordance selection covers 143 of 150 objects, i. e. 95% of all cases. At the same time, it has the highest grasp success rate compared to the other operational modes and consequently, the lowest number of failures over all experiments. The *autonomous* affordance selection, on the contrary, requires more executions for the same number of objects as it fails to clear the box in complex scenes. However, it is still able to grasp 90% of the objects over all experiments. The fact that all objects of the experiments could be grasped in case of *manual* affordance selection indicates that the operator is able to create very precise grasps and even deal with complex scenes. Nevertheless, as manual teleoperation cannot easily cope with challenges regarding robot control such as checking for reachability or avoiding collision, it leads to more failures which is also shown by the high number of grasp executions. Furthermore, it puts the highest cognitive load on the human operator.

The comparison demonstrates that exploiting the expert knowledge and the high-level decisions of an experienced human operator has a significant impact on successful task execution while keeping the cognitive load to a minimum. Additionally, it shows that more than 90% and 95% of all cases in complex scenes can already be covered by the *autonomous* and *semi-autonomous* mode, respectively. As a consequence, the operator only has to make use of the *manual* affordance selection in the remaining 5% of the cases. Due to the fact that the operational mode can be switched during run-time, our approach is capable to combine these methods and thus, can provide a robust solution that is applicable in real world scenarios.

VI. Conclusion

In this paper, we presented a robust approach for grasping unknown objects in cluttered scenes with a humanoid robot in the context of a nuclear decommissioning task. Our approach combines the advantages of autonomous robot control as well as the exploitation of expert knowledge of an experienced human operator by the semi-autonomous selection of grasp affordances. We evaluated our approach in 75 real world experiments with more than 660 grasp executions on the humanoid robot ARMAR-6 in which we compared three different levels of autonomy for selecting and extracting grasp affordances.

The experiments show that the *semi-autonomous* selection of grasp affordances through an operator can drastically reduce the number of failures to 28%, compared to 43% failures in the *autonomous*, and 34% failures in the *manual* mode. Therefore, it can contribute significantly to successful task execution while keeping the cognitive load of the operator to a minimum. Furthermore, the comparison of the three operational modes demonstrates that, even in complex scenarios, more than 90% and 95% of all cases can already be covered by the *autonomous* and *semi-autonomous* mode. Thus, *manual* selection is only required in 5% of the remaining cases. However, due to the fact that not all cases can be covered *semi-autonomously* and the human operator should always be able to actively intervene in the scene, we propose to use a combination of these methods to provide a robust solution for real world scenarios.

In the future, we plan to extend our approach by integrating mobility and adding more affordances such as pushing, lifting, and bimanual manipulation. This would lead to further improvements of the approach in terms of versatility, i. e. the ability to consider additional manipulation actions. Furthermore, we want to improve the autonomous manipulation, e.g. by learning from human decisions for affordance selection and learning from experience in order to detect and predict failures during task execution.

Considering that a human operator manually sets or selects an affordance in the *manual* and *semi-autonomous* mode, respectively, it is not yet clear how much human learning can influence the success of the robot's task execution. Therefore, we plan to investigate the success rate in similar experiments with varying human operators in a broader study. In this context, the cognitive load on the human operator could further be evaluated in a quantifiable way.

References

[1] *Nuclear Power Reactors in the World*. No. 2 in Reference Data Series, Vienna: International Atomic Energy Agency, 2019.

[2] J. Petereit, J. Beyerer, T. Asfour, S. Gentes, B. Hein, U. D. Hanebeck, F. Kirchner, R. Dillmann, H. H. Gotting, M. Weiser, M. Gustmann, and T. Egloffstein, "ROBDEKON: Robotic Systems for Decontamination in Hazardous Environments," in *IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pp. 249–255, 2019.

[3] J. J. Gibson, "The theory of affordances," in *The Ecological Approach to Visual Perception*, ch. 8, pp. 119–137, Houghton Mifflin, 1979.

[4] M. Spenko, S. Buerger, and K. Iagnemma, eds., *The DARPA Robotics Challenge Finals: Humanoid Robots To The Rescue*, vol. 121 of *Springer Tracts in Advanced Robotics*. Springer International Publishing, 2018.

[5] T. Yoshiike, M. Kuroda, R. Ujino, Y. Kanemoto, H. Kaneko, H. Higuchi, S. Komura, S. Iwasaki, M. Asatani, and T. Koshiishi, "The Experimental Humanoid Robot E2-DR: A Design for Inspection and Disaster Response in Industrial Environments," *IEEE Robotics & Automation Magazine*, vol. 26, no. 4, pp. 46–58, 2019.

[6] T. Klamt, M. Kamedula, H. Karaoguz, N. Kashiri, A. Laurenzi, C. Lenz, D. Leonardis, E. Mingo Hoffman, L. Muratore, D. Pavlichenko, F. Porcini, D. Rodriguez, Z. Ren, F. Schilling, M. Schwarz, M. Solazzi, M. Felsberg, A. Frisoli, M. Gustmann, P. Jensfelt, K. Nordberg, J. Rossmann, L. Baccelliere, U. Suss, N. G. Tsagarakis, S. Behnke, X. Chen, D. Chiaradia, T. Cichon, M. Gabardi, P. Guria, and K. Holmquist, "Flexible Disaster Response of Tomorrow: Final Presentation and Evaluation of the CENTAURO System," *IEEE Robotics & Automation Magazine*, vol. 26, pp. 59–72, Dec 2019.

[7] I. Kumagai, F. Kanehiro, M. Morisawa, T. Sakaguchi, S. Nakaoka, K. Kaneko, H. Kaminaga, S. Kajita, M. Benallegue, and R. Cisneros, "Toward Industrialization of Humanoid Robots: Autonomous Plasterboard Installation to Improve Safety and Efficiency," *IEEE Robotics & Automation Magazine*, vol. 26, pp. 20–29, Dec 2019.

[8] A. Kheddar, M. A. Roa, P.-b. Wieber, F. Chaumette, F. Spindler, G. Oriolo, L. Lanari, A. Escande, K. Chappellet, F. Kanehiro, P. Rabate, S. Caron, P. Gergondet, A. Comport, A. Tanguy, C. Ott, B. Henze, G. Mesesan, and J. Englsberger, "Humanoid Robots in Aircraft Manufacturing: The Airbus Use Cases," *IEEE Robotics & Automation Magazine*, vol. 26, pp. 30–45, Dec 2019.

[9] T. Asfour, M. Wächter, L. Kaul, S. Rader, P. Weiner, S. Ottenhaus, R. Grimm, Y. Zhou, M. Grotz, and F. Paus, "ARMAR-6: A High-Performance Humanoid for Human-Robot Collaboration in Real World Scenarios," *IEEE Robotics & Automation Magazine*, vol. 26, no. 4, pp. 108–121, 2019.

[10] P. Kaiser, E. E. Aksoy, M. Grotz, and T. Asfour, "Towards a Hierarchy of Loco-Manipulation Affordances," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2839–2846, IEEE, Oct 2016.

[11] A. Norton, W. Ober, L. Baraniecki, D. Shane, A. Skinner, and H. Yanco, "Perspectives on human-robot team performance from an evaluation of the DARPA robotics challenge," in *Springer Tracts in Advanced Robotics*, vol. 121, pp. 631–666, 2018.

[12] J. Lim, H. Bae, J. Oh, I. Lee, I. Shim, H. Jung, H. M. Joe, O. Sim, T. Jung, S. Shin, K. Joo, M. Kim, K. Lee, Y. Bok, D.-G. Choi, B. Cho, S. Kim, J. Heo, I. Kim, J. Lee, I. S. Kwon, and J.-H. Oh, "Robot System of DRC-HUBO+ and Control Strategy of Team KAIST in DARPA Robotics Challenge Finals," in *Springer Tracts in Advanced Robotics*, vol. 121, pp. 27–69, 2018.

[13] M. Johnson, B. Shrewsbury, S. Bertrand, T. Wu, D. Duran, M. Floyd, P. Abeles, D. Stephen, N. Mertins, A. Lesman, J. Carff, W. Rifenburgh, P. Kaveti, W. Straatman, J. Smith, M. Griffioen, B. Layton, T. de Boer, T. Koolen, P. Neuhaus, and J. Pratt, "Team IHMC's Lessons Learned from the DARPA Robotics Challenge Trials," *Journal of Field Robotics*, vol. 32, pp. 192–208, mar 2015.

[14] C. G. Atkeson, P. Franklin, M. Gennert, J. P. Graff, P. He, A. Jaeger, J. Kim, K. Knoedler, L. Li, C. Liu, X. Long, B. P. W. Babu, T. Padir, F. Polido, G. G. Tighe, X. Xinjilefu, N. Banerjee, D. Berenson, C. P. Bove, X. Cui, M. DeDonato, R. Du, and S. Feng, "No falls, no resets: Reliable humanoid behavior in the DARPA robotics challenge," in *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, pp. 623–630, IEEE, Nov 2015.

[15] D. W. Seward and M. J. Bakari, "The Use of Robotics and Automation in Nuclear Decommissioning," in *22nd International Symposium on Automation and Robotics in Construction (ISARC)*, Sep 2005.

[16] N. Marturi, A. Rastegarpanah, C. Takahashi, M. Adjigble, R. Stolkin, S. Zurek, M. Kopicki, M. Talha, J. A. Kuo, and Y. Bekiroglu, "Towards advanced robotic manipulation for nuclear decommissioning: A pilot study on tele-operation and autonomy," *International Conference on Robotics and Automation for Humanitarian Applications - Conference Proceedings*, pp. 1–8, 2017.

[17] A. Settimi, C. Pavan, V. Varricchio, M. Ferrati, E. Mingo Hoffman, A. Rocchi, K. Melo, N. G. Tsagarakis, and A. Bicchi, "A Modular Approach for Remote Operation of Humanoid Robots in Search and Rescue Scenarios," in *Modelling and Simulation for Autonomous Systems*, pp. 192–205, 2014.

[18] L. Penco, N. Scianca, V. Modugno, L. Lanari, G. Oriolo, and S. Ivaldi, "A Multi-Mode Teleoperation Framework for Humanoid Loco-Manipulation," *IEEE Robotics & Automation Magazine*, vol. 26, pp. 73–82, dec 2019.

[19] I. Jang, J. Carrasco, A. Weightman, and B. Lennox, "Intuitive bare-hand teleoperation of a robotic manipulator using virtual reality and leap motion," in *Towards Autonomous Robotic Systems* (K. Althoefer, J. Konstantinova, and K. Zhang, eds.), (Cham), pp. 283–294, Springer International Publishing, 2019.

[20] P. Kaiser, D. Kanoulas, M. Grotz, L. Muratore, A. Rocchi, E. M. Hoffman, N. G. Tsagarakis, and T. Asfour, "An affordance-based pilot interface for high-level control of humanoid robots in supervised autonomy," in *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, pp. 621–628, 2016.

[21] N. Vahrenkamp, M. Wächter, M. Kröhnert, K. Welke, and T. Asfour, "The Robot Software Framework ArmarX," *Information Technology*, vol. 57, no. 2, pp. 99–111, 2015.

[22] Y. Zhou, J. Gao, and T. Asfour, "Learning Via-Point Movement Primitives with Inter- and Extrapolation Capabilities," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4301–4308, Nov 2019.