Abstract—We present the design, implementation, and evaluation of RF-Grasp, a robotic system that can grasp fully-occluded objects in unknown and unstructured environments. Unlike prior systems that are constrained by the line-of-sight perception of vision and infrared sensors, RF-Grasp employs RF (Radio Frequency) perception to identify and locate target objects through occlusions, and perform efficient exploration and complex manipulation tasks in non-line-of-sight settings.

RF-Grasp relies on an eye-in-hand camera and batteryless RFID tags attached to objects of interest. It introduces two main innovations: (1) an RF-visual servoing controller that uses the RFID's location to selectively explore the environment and plan an efficient trajectory toward an occluded target, and (2) an RF-visual deep reinforcement learning network that can learn and execute efficient, complex policies for decluttering and grasping.

We implemented and evaluated an end-to-end physical prototype of RF-Grasp and a state-of-the-art baseline. We demonstrate it improves success rate and efficiency by up to 40-50% in cluttered settings. We also demonstrate RF-Grasp in novel tasks such as fully-occluded grasping in cluttered settings. We also demonstrate RF-Grasp in novel tasks such as fully-occluded grasping in cluttered settings.

I. INTRODUCTION

Mechanical search is a fundamental problem in robotics [1], [2], [3], [4]. It refers to the task of searching for and retrieving a partially or fully-occluded target object. This problem arises frequently in unstructured environments such as warehouses, hospitals, agile manufacturing plants, and homes. For example, a warehouse robot may need to retrieve an e-commerce customer's desired item from under a pile. Similarly, a robot may need to retrieve a desired tool (e.g., screwdriver) from behind an obstacle to perform a complex task such as furniture assembly [5].

To address this problem, the past few years have seen significant advances in learning models that can either recognize target objects through partial occlusions or actively explore the environment, searching for the object of interest. Recent proposals have also considered the geometry of obstacles or pile [1], [2], [6], [7], [8], demonstrating remarkable results in efficiently exploring and decluttering the environment.

However, existing mechanical search systems are inherently constrained by their vision sensors, which can only perceive objects in their direct line-of-sight. If the object of interest is behind an obstacle, they need to actively explore the environment searching for it, a process that can be very expensive and often fails [6]. Moreover, these systems are typically limited to a single pile or pick-up bin [1], [3], and cannot generalize to mechanical search problems with multiple piles or multiple obstacles. They also cannot perform tasks like prioritized sorting, where a robot needs to retrieve all objects belonging to a specific class (e.g., all plastic bottles from a box) then declare task completion.

In this paper, we draw on recent advances in RF (Radio Frequency) perception [9], [10] to enable novel and highly-efficient mechanical search tasks in unstructured environment searching for it, a process that can be very expensive and often fails [6]. Moreover, these systems are typically limited to a single pile or pick-up bin [1], [3], and cannot generalize to mechanical search problems with multiple piles or multiple obstacles. They also cannot perform tasks like prioritized sorting, where a robot needs to retrieve all objects belonging to a specific class (e.g., all plastic bottles from a box) then declare task completion.

In this paper, we draw on recent advances in RF (Radio Frequency) perception [9], [10] to enable novel and highly-efficient mechanical search tasks in unstructured environments. This paper provides three main contributions:

- It presents the first system that bridges RF and vision perception (RF+RGB-D) to enable efficient and novel mechanical search tasks across line-of-sight, non-line-of-sight, and highly cluttered environments. This paper provides three main contributions:

  - It presents the first system that bridges RF and vision perception (RF+RGB-D) to enable efficient and novel mechanical search tasks across line-of-sight, non-line-of-sight, and highly cluttered environments. This paper provides three main contributions:
  - It introduces two novel primitives: (1) RF-visual servoing,
RF-Grasp uses a customized reader that can identify and accurately localize RFID-tagged objects through occlusions. A novel controller that performs RF-guided active exploration and navigation to avoid obstacles and maneuver toward the target object, and (2) RF-visual grasping, a model-free deep reinforcement learning network that employs RF-based attention to learn and execute efficient and complex policies for decluttering and grasping.

- It presents an end-to-end prototype implementation and evaluation of RF-Grasp. The implementation is built using a UR5e robot, Intel RealSense D415 depth camera, and a customized RF localization system on software radios. The system is evaluated in over 100 physical experiments and compared to a baseline that combines two prior systems [6], [4]. The evaluation demonstrates that RF-Grasp improves success rate by up to 40% and efficiency by up to 50% in challenging environments. Moreover, it demonstrates that RF-Grasp can perform novel mechanical search tasks of fully-occluded objects across settings with obstacles, multiple piles, and multiple target objects.

In comparison to vision-only systems, RF-Grasp’s design requires target object(s) tagged with RFIDs. However, we believe that given the widespread adoption of RFIDs by many industries (with tens of billions deployed annually [12]), the system can already have significant practical impact. We also hope this work motivates further research bridging RF and vision for novel (and more efficient) robotic tasks.

II. BACKGROUND & RELATED WORK
RFIDs and their Applications in Robotics. Radio Frequency IDentification (RFID) is a mature technology, widely adopted by many industries as barcode replacement in retail, manufacturing, and warehousing [12]. Recent years have seen significant advances in RFID localization technologies [13], [14], [15], which not only use the tags for identification, but also locate them with centimeter-scale precision even in cluttered and occluded settings [16], [9], as in Fig. 2.

Prior work leveraged RFIDs as location markers for robotic navigation [17], [18], [19] and to guide mobile robots toward grasping [14], [20], [21], [22], [15], [23]. But because occlusions are transparent to RF, these systems could not perceive them to declutter or maneuver around them. In contrast, RF-Grasp demonstrates, for the first time, how RF-visual fusion enables complex tasks like mechanical search. Mechanical Search and Grasping in Clutter. Object manipulation in cluttered environments has received significant attention and achieved remarkable success via supervised and unsupervised methods [24], [25], [26], [27]. Our work is motivated by a similar desire to grasp in clutter and builds on this line of work but focuses on the problem of grasping a specific target object rather than any object or all objects.

Recognizing and locating target objects in occlusions has also received much attention [28]. Various techniques have been proposed including perceptual completion [29], [30] and active/interactive perception where a camera is moved to more desirable vantage points for recognizing objects [31], [32], [33]. In contrast to this past work, which requires a (partial) line-of-sight to an object to recognize it, our work uses RF perception to directly identify and locate objects through occlusions, and without requiring any prior training.

RF-Grasp is most related to recent prior work on mechanical search, where the goal is to search for and grasp a target object that may be partially or fully occluded [6], [1], [2], [34]; this includes both one-shot and multi-step procedures for search and retrieval [4], [3], [35], [36], [37]. Unlike RF-Grasp, this prior work first needs to search for the object to identify its location, which does not scale well with the number and size of piles, or the number of target objects; in contrast, by exploiting RF perception, RF-Grasp can recognize and locate RFID-tagged objects to perform highly efficient active exploration and manipulation.

III. SYSTEM OVERVIEW

We consider a generalized mechanical search problem where a robot needs to extract a target object in an unstructured environment. The object may be in line-of-sight or non-line-of-sight; it may be behind occlusions and/or under a pile, and the environment may have additional occlusions, piles, and clutter, similar to Fig. 1. Moreover, the robot may need to extract all target objects from a semantic class.

We assume that each target object (but not necessarily other objects) is tagged with a UHF RFID and kinematically reachable from a robotic arm on a fixed base. We also assume the environment is static. The robot is equipped with an eye-in-hand camera, mounted on a 6-DOF manipulator, which starts from a random initial location and orientation. The robot is aided by a fixed-mount RF perception module in the form of an RFID micro-location sensor with multiple antennas. The robot knows the target object(s) RFID number but no additional information about its geometry or location.

RF-Grasp’s objective is to extract the target(s) from the environment using the shortest travel distance and the minimum number of grasp attempts. It starts by querying the RFIDs in the environment and using its RF perception module to identify them and compute their 3D locations, even if they are behind occlusions [9]. It divides the mechanical search problem into two sub-problems as shown in Fig. 3 and addresses each of them using a separate subcomponent:

- **RF-Visual Servoing:** The first aims to maneuver the robotic arm toward the target object. It uses RGB-D images to create a 3D model of the environment and fuses the RFID’s location into it. It then performs RF-guided active exploration and trajectory optimization to efficiently maneuver around obstacles toward the object. Exploration stops when it can grasp the object or declutter its vicinity.

- **RF-Visual Grasping:** RF-Grasp’s second sub-component is a model-free deep-reinforcement learning network that
RF-Grasp aims to identify optimal grasping affordances from RGB-D, using the RFID’s location as an attention mechanism. The robot attempts to grasp and pick up the target object, and stops once the RF perception module determines that the RFID’s location has moved with the end-effector.

IV. RF-Visual Servoing

Given the RFID’s location and an RGB-D image, RF-Grasp needs to maneuver the robotic arm toward the target object. A key difficulty in this process is that the environment is not known a priori and the direct path may be occluded by obstacles. Below, we describe how RF-Grasp actively explores the environment as it tries to determine an optimal path around obstacles toward the target object.

A. Problem Definition

We frame the servoing problem as a Partially Observable Markov Decision Process (POMDP) where the robot needs to efficiently explore the environment while minimizing the trajectory toward the object of interest. The state of the environment \( S \) consists of the robot joint state \( (x^R_t ∈ R^6) \), RFID location \( (p = (x_p, y_p, z_p)) \), the obstacles, occlusions, and other objects. The control signal, \( u_t ∈ R^6 \), is applied to the joint state, and changes the robot pose. The observations \( Ω \) consist of the joint state, the RFID location, and RGB-D data from the wrist-mounted camera. The problem is partially observable because the robot has no prior knowledge of (nor observes) the entire 3D workspace.

Modeling Environmental Uncertainties. Similar to past work [6], RF-Grasp encodes uncertainty using a mixture of Gaussians. Each occluded region is modeled as a 3D Gaussian as shown in Fig. 4. The mean and covariance of the \( m \)-th Gaussian are denoted \( (x^m_t, Σ^m_t) \) at \( t = 0 \). The environment is assumed to be static; hence, the means remain the same over the planning horizon, but the covariances \( Σ^m_t \) get updated as the system explores the environment.

RF-Biased Objective Function. To efficiently maneuver toward the object, RF-Grasp aims to minimize its trajectory (control effort) while minimizing its uncertainty of the surrounding environment. Mathematically, the cost at time \( t \) is:

\[
C_{t=0:T}(x^R_t, Σ^{1:M}_t, u_t) = α\|u_t\|^2 + \sum_{m=1}^{M} β^m_tr(Σ^m_t)
\]

where \( M \) is the total number of occluded regions, \( tr \) is trace, \( α \) and \( β^m \) are scalar weighting parameters.

To bias the controller to explore the occluded region surrounding the RFID, we set that region’s corresponding weight, \( β^1 \), to be significantly larger than others. Moreover, to give the robot more flexibility to explore in the beginning, we start with a lower \( β^1 \) and increase it over time.

Given the above cost function, we can now formulate the trajectory optimization problem as a minimization function over the planning horizon \( T \) as follows:

\[
\min x^R_{0:T},u_{0:T} \quad \mathbb{E} \left[ \sum_{t=0}^{T} C_t(x^R_t, Σ^{1:M}_t, u_t) \right]
\]

s.t. \( x^R_{t+1} = f(x^R_t, u_t, 0) \), \( x^R_t ∈ X_{\text{feasible}} \), \( x^R_0 = x^R_{\text{init}} \), \( u^R_t ∈ U_{\text{feasible}} \), \( u_T = 0 \)

where \( X_{\text{feasible}} \) and \( U_{\text{feasible}} \) represent the set of feasible joint states and control signals of the robot arm. \( x^R_{\text{init}} \) is the initial joint state of the robot. The dynamics model for the robot is given by differentiable and stochastic function \( x^R_{t+1} = f(x^R_t, u_t, q_t) \), \( q_t ∼ N(0, I) \) where \( q_t \) is the dynamics noise.

B. RF-Guided Trajectory Optimization

RF-Grasp’s approach for solving the above problem follows prior work in Gaussian Belief Space Planning (GBSP), including modeling the environment using a 3D voxel grid map, extracting frontiers and occluded regions, dealing with discontinuities in the RGB-D observation model, modeling the observation and dynamics using Gaussians, and propagating beliefs using Extended Kalman Filter (EKF). We refer the interested reader to prior work for details [6], and focus below on two unique features of our solution, aside from the RF-biased objective function described above:

(a) RF-based Initial Trajectory. To aid the optimization solver and enable faster convergence, we seed the optimization function with a straight-line initial trajectory in Cartesian space from the end-effector to the RFID location.

(b) Exploration Termination Criteria. In principle, RF-Grasp should stop exploring when it determines that no major obstacles remain, and it can proceed to grasping. But, such reasoning is challenging because the target may still...
be occluded by distractor objects (e.g., under a pile), and RF-Grasp needs to declutter before grasping.

We formulate the exploration termination criteria as a function of the uncertainty region around the target object. Such uncertainty is encoded both in the covariance of the 3D Gaussian around the object $\Sigma^2$ and in visibility of the voxels $v$ in the vicinity of $\nu$ the target object’s location $p$.

Mathematically, the termination criteria can be expressed as:

$$\text{trace}(\Sigma^2) < \rho_\Sigma \lor \sum_{v \in \nu(p)} F(v)/|\nu(p)| > \rho_v$$

(2)

where $F(v) = 1$ if voxel $v$ has been seen by camera and 0 otherwise. The criteria imply the uncertainty around the object is smaller than threshold $\rho_\Sigma$, or the visible fraction of the region around the RFID is larger than threshold $\rho_v$.

The combination of the above criteria is necessary to deal with the diversity of clutter scenarios. For example, when the target item is under a soft cover, the trace of covariance won’t be below $\rho_\Sigma$ since the cover may occlude a large region. However, enough voxels in the vicinity of the item will be visible to the camera such that the second criterion is met and RF-Grasp will proceed to grasping.

Finally, it is worth noting that RF-Grasp’s active exploration formulation’s objective function pushes the robot end-effector away from collisions. This is because if the camera is too close to a large occlusion, the covariances $\Sigma^2_{\nu}$ become larger, thus penalizing the expected cost and biasing the optimal trajectory away from the large obstruction.

V. Radio-Visual Learning of Grasping Policies

The above primitive enables RF-Grasp to intelligently explore the environment and servo the robotic arm around obstacles, closer to the object of interest, but the robot still needs to grasp it. Below, we describe how RF-Grasp exploits RF-based attention to learn efficient grasping policies.

A. The Grasping Sub-problem

We formulate the grasping sub-problem as a Markov Decision Process (MDP). Here, the action $a_t$ is grasping with a parallel jaw gripper at position $g_t = (x_{a_t}, y_{a_t}, z_{a_t})$ with gripper rotation of $\theta_{a_t}$. The goal is to learn the optimal policy $\pi^*$ to grasp the item of interest either directly or by manipulating the environment.

This can be cast as a deep reinforcement learning problem where the robot aims to maximize the future reward (by grasping the target object). Similar to prior work on unsupervised grasping [24], we use a Deep Q-Network (DQN, an off-policy Q-learning approach) to solve this problem. The action-value function $Q(s_t, a_t)$ estimates the expected reward at the state $s_t$ when taking the action $a_t$ according to the policy $\pi(s_t)$. Q-learning can be defined as:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)]$$

where $r_t$ is the reward that system receives after action $a_t$, $\alpha$ is the learning rate, and $\gamma$ is the discount factor. The aim of the network is to learn Q-function by minimizing:

$$\delta_t = [r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)]$$

B. RF-based Attention and Rewards

RF-Grasp trains a deep reinforcement learning network in simulation, and folds in RF information via reward and attention mechanisms. Fig. 5 shows the overall architecture, consisting of feed-forward fully convolutional networks (FCNs). The networks take state representations (RGB-D and RFID position) as input, and output pixel-wise map of Q values. RF-Grasp selects the optimal grasping policy as the one with the highest Q across the output affordance map.

Spatio-Temporal Reward Function. We construct a spatio-temporal reward function to encourage the robot to grasp the target item or those in its near vicinity. The reward function $r_t(s_t, s_{t+1}, a_t)$ is 1 if the robot grasps the target object; maximum ($\rho_\rho$, $\rho_\delta$) if it grasps another item; and 0 otherwise. In our implementation, $\rho$ is a constant set to 0.007. Since RF perception tracks the RFID’s location, it can determine whenever the grasp is successful (it is picked up).

RF-based Attention. RF-Grasp incorporates RF-based attention through two main strategies (and at three layers):

- **RF-based Binary Mask.** The RGB and depth heightmaps are cropped to a square around the RFID’s location (11cm x 11cm). This pre-processing attention mechanism allows the network to focus on the vicinity of the target object and compute the affordance map with higher resolution and/or less computational complexity.

- **RF Kernel.** The RFID location is also used to construct an RF kernel, a 2D Gaussian centered around $p$, whose standard deviation accounts for RF localization errors. The kernel is fed to the network, and is multiplied by

2In our implementation, $\rho_\Sigma = 0.005$ and $\rho_v = 0.1$, and the vicinity, $\nu(p)$, is set to a 5cm x 5cm x 10cm cube centered at RFID location.

3Heightmaps are computed by extracting 3D point cloud from RGB-D images and projecting the images orthographically, parallel to the table top.
DQN’s last layer to compute the final affordance map. This increases the probability of grasping the target item.

RF-Grasp uses the above RF attention mechanisms to extend a state-of-the-art deep-reinforcement learning grasping network [24], as shown in Fig. 5. It consists of two 121-layer DenseNets: the first takes as input three channels of cropped RGB heightmaps; the second’s input is the RF kernel plus two copies of cropped depth heightmap. To discover optimal grasping affordances, the inputs are rotated in 16 directions and fed separately to the network. Then, the outputs of both streams are concatenated along channels. Two convolution layers with kernel size of $1 \times 1$ come after, producing the 2D map of the Q function estimation. The output contains 16 maps for grasp, from which RF-Grasp chooses the position and rotation with highest probability of success.

Training Details. We use Huber loss function for training our network to reduce sensitivity to outliers. The DenseNet initial weights were taken from the pre-trained model in [24], and fine-tuned by training for 500 iterations in simulation. The gradient is only backpropagated through the pixel of affordance map that we executed grasping according to its predicted value. The network is trained by stochastic gradient descent with learning rates $10^{-4}$, momentum 0.9, and weight decay $2^{-5}$. The reward discount factor is 0.2.

In training, we used prioritized experience replay [38] to improve sample efficiency and stability. We define threshold $\rho_i = 0.05 + \min(0.05 \times \#\text{iteration}, 4)$, and performed stochastic rank-based prioritization among experiences with rewards $\geq \rho_i$. Prioritization is estimated by a power-law distribution.

VI. IMPLEMENTATION & EVALUATION

Physical Prototype. Our setup consists of a UR5e robot arm, 2F-85 Robotiq gripper, and an Intel RealSense D415 depth camera mounted on the gripper. The RF perception module is implemented as an RFID localization system on USRP software radios (we refer the reader to [9] for implementation details). The RFID localization system is set up on a table in front of the robot arm 16cm below the robot base level. The robot is connected through Ethernet to a PC that runs Ubuntu 16.04 and has an Intel Core i9-9900K processor; RTX 2080 Ti, 11 GB graphic card; and 32 GB RAM. We also used a USB to RS485 interface to control the gripper from the PC.

The robot is controlled using Universal Robots ROS Driver on ROS kinetic. We used Moveit! [39] and OMPL [40] for planning scene and inverse kinematic solver. TSDF volume is created using Yak [41]. We used PCL [42] for extracting clusters and occlusion from TSDF volume. To solve the SQP for trajectory optimization, we used FORCES [43]. The code is implemented in both C++ and Python. Objects of interest are tagged with off-the-shelf UHF RFID tags (e.g., Alien tag [44]), whose dimensions typically range from 1-12 cm.

Simulation. We built a full simulation for RF-Grasp by combining three environments: 1) VREP [45] simulates the depth camera visual data, and robot and gripper’s physical interactions with objects in the environment. We used Bullet Physics 2.83 as the physics engine. 2) Gazebo [46] predicts the robot state. This module takes into account gravity, inertia, and crashing into the ground and other obstacles. 3) Rviz [47] visualizes the robot trajectory, obstacles in the environment, occluded areas, and their mean and covariances. In the simulated system, we used the VREP remote API to get each object’s location to simulate the RFID localization system. Note that only VREP was used to train the DQN, while all three were used for RF-Visual servoing.

Baseline. Since prior work does not deal with the generalized mechanical search problem where the target objects are both behind an occlusion and dense clutter, we built a baseline by combining two state-of-the-art past systems. The first performs efficient active exploration and trajectory optimization using GBSP [6], but without RF-biasing and guidance. The second is a DQN with color-based attention [4] which estimates the location of the item using a unique color. Without loss of generality, the baseline aims to grasp a green item in a workspace with non-green objects, and it switches from exploration to grasping when it detects more than 100 green pixels in its field of view. For fairness to the baseline, we only compare it to our system in scenarios where the object is only partially occluded, but not fully occluded.

Evaluation Metrics. The robot workspace is atop a table with dimensions of 0.8m×1.2m. We consider two metrics: 1) Average Traveled Distance: the distance that the robot’s end-effector moves from the initial position until the item of interest is grasped. 2) Task Completion Rate: the percentage of trials that resulted in successful grasping of tagged items (or green item for baseline) before 10 grasping attempts and 5 meter of traveled distance in the exploration phase.

VII. RESULTS

A. Performance Results

We evaluated RF-Grasp and the baseline quantitatively by varying the number of large occlusions ($M$) and distractor objects ($N$). Each large occlusion hides a different region of the workspace from the camera. We also varied the initial position of the robot across experimental trials, but ensured the baseline and RF-Grasp shared the same initial position.

(a) Traveled Distance: We recorded the robot’s end-effector position and computed its traveled distance for RF-Grasp and the baseline. We tested six scenarios with 0-5 occluded regions (each with 1-15 distractor objects) and ran 10 trials per system and scenario. We placed the robot in an initial pose where the wrist-mounted distractor objects partially sees a frontier, i.e., with one or more occluded regions in its initial observation. As the robot explored the environment to minimize its cost function, it found other frontiers and occluded regions.
Next, we evaluated RF-Grasp’s (b) Task Completion Rate: enabling efficient exploration and trajectory optimization. Of RF-Grasp’s first core component (RF-visual servoing) in within that workspace). This result demonstrates the value by its arm size and mobility (and all occluded regions are Note that the average traveled distance plateaus beyond four occlusions because the workspace of the robot is limited by its arm size and mobility (and all occluded regions are within that workspace). This result demonstrates the value of RF-Grasp’s first core component (RF-visual servoing) in enabling efficient exploration and trajectory optimization.

(b) Task Completion Rate: Next, we evaluated RF-Grasp’s completion rate (defined in [7]) across the same experiments described above. Table I shows the results. RF-Grasp was able to complete the task in 60 out of 60 trials, while the baseline was unable to complete the task in 7 out of 60 trials. The baseline completion rate decreases as the number of occlusions and complexity in experiments increases.

![RF Perception](image1)
![RF Perception](image2)
![RF Perception](image3)

(a) Under Soft Cover  
(b) Multi-obstacle and Multi-pile  
(c) Prioritized Sorting

Fig. 7: Generalized Mechanical Search. We demonstrate RF-Grasp in 3 challenging tasks which the baseline cannot perform.

Fig. 6 plots the average traveled distance of the gripper for both RF-Grasp and the baseline. For all except $M = 0$, our results show that RF-Grasp travels shorter distances (by up to 50%); moreover, RF-Grasp significantly outperforms the baseline as the number of occlusions increases. This result is expected because RF-Grasp is guided by the RFID’s location (in the cost function), which reduces the traveled distance and the time spent on exploring the environment searching for the tagged item in comparison to the baseline. Note that the average traveled distance plateaus beyond four occlusions because the workspace of the robot is limited by its arm size and mobility (and all occluded regions are within that workspace). This result demonstrates the value of RF-Grasp’s first core component (RF-visual servoing) in enabling efficient exploration and trajectory optimization.

Next, we show that RF-Grasp can successfully perform mechanical search in challenging scenarios where the state-of-the-art baseline is unsuccessful. We consider three such scenarios, shown in Fig. 7 (see video for demonstration).

Scenario 1: Under Soft Cover: We consider scenarios with more than 20 distractor objects along with the target item, covered with soft package filling sheets (Fig. 7(a)). There is no line-of-sight from the wrist-mounted camera or antennas to the item. RF-Grasp localizes the RFID and moves directly above the covered item. Because the target is occluded, the robot’s attention mechanism biases it to first pick up the cover, but realizes it hasn’t picked the target object since the RFID’s location has not changed. It puts the cover aside and makes a second grasping attempt. This time, the tracked RFID location changes with the robot’s end-effector, and RF-Grasp confirms it has grasped the requested item. We tested RF-Grasp in this scenario 5 times. The average travelled distance was 2.68m with standard deviation of 0.98. RF-Grasp was successful in all 5 trials.

In contrast to RF-Grasp, our baseline is incapable of successfully completing this task because the green-tagged object would remain fully-occluded; thus, it can neither complete its exploration nor efficiently grasp the target object. It is also worth noting that some recent mechanical search system could, in principle, succeed in such a scenario [2].

Scenario 2: Multiple piles and multiple obstacles: Next, we tested RF-Grasp in a scenario with large obstacles and a cover (Fig. 7(b)). We used two boxes, occluding two regions from camera. We placed 5 items in each box, and a tagged item in one of them. Similar to the above scenario, RF-Grasp was successful in exploring the environment using RF-guided optimization, maneuvering toward the target, removing the cover, and successfully grasping. We tested RF-Grasp in this scenario 3 times. The average travelled distance was 4.13 with standard deviation of 1.91. RF-Grasp was successful across all 3 trials. To the best of our knowledge, this is a novel task that existing robotic systems cannot perform.

Scenario 3: Multiple piles and multiple target objects: Our final scenario is involves mechanical search for all objects belonging to a semantic class. An example is shown in Fig. 7(c), where the robot needs to extract all RFID-tagged items that have a certain feature (e.g., the same dimensions) and sort them into a separate bin. The RF perception module reads and locates all RFID-tagged items, and determines which it needs to grasp. Subsequently, the robot performs mechanical search for each of the items, picks them up, and drops them in the white bin to the bottom right. RF-Grasp succeeded and declared task completion once it localized all target objects to the final bin. The total travelled distance (for 3 objects) was 16.18m. To the best of our knowledge, this is also a novel task that existing systems cannot perform.

VIII. CONCLUSION

RF-Grasp is a system that fuses RF-visual information to enable robotic grasping of fully-occluded objects. As the work evolves, it can be extended to grasping at different angles and grippers, novel tasks that use RFID’s for semantic grasping and scene understanding, and more complex manipulation and assembly including human-robot interaction.
