Robotic Grasping of Fully-Occluded Objects using RF Perception

Tara Boroushaki, Junshan Leng, Ian Clester, Alberto Rodriguez, Fadel Adib
Massachusetts Institute of Technology

***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. We demonstrate it improves success rate and efficiency by up to 40-50% over a state-of-the-art baseline. We also demonstrate RF-Grasp in novel tasks such mechanical search of fully-occluded objects behind obstacles, opening up new possibilities for robotic manipulation. Qualitative results (videos) available at `rfgrasp.media.mit.edu`

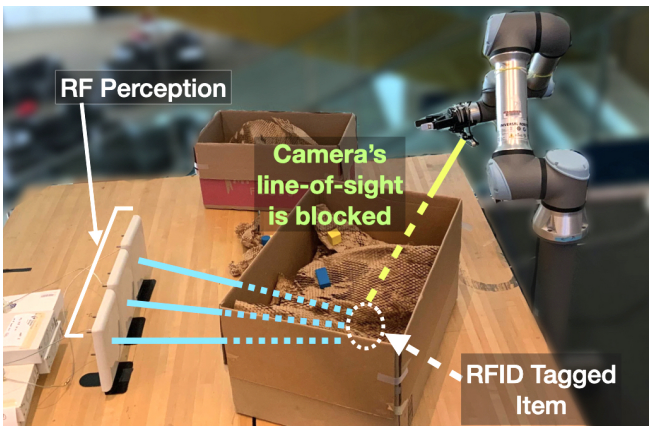
I. Introduction

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].

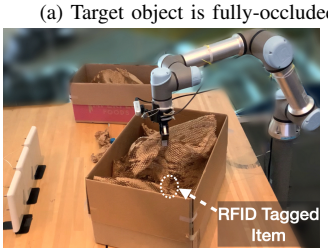
To address this problem, the past few years have seen significant advances in learning models that can either recognize target objects through partial occlusions or actively explore the environment, searching for the object of interest. Recent proposals have also considered the geometry of obstacles 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) and 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. Unlike visible light and infrared, RF signals can traverse everyday occlusions like cardboard boxes, wooden dividers (and walls), opaque plastic covers, and colored glass. This "see through" capability enables a robot to perceive objects tagged with passive 3-cent RF stickers (called RFIDs), even when they are fully occluded from its vision sensor. RFID systems can read and uniquely identify hundreds of tags per second from up to 9 m and through occlusions [11].



(a) Target object is fully-occluded in an unstructured environment.



(b) Decluttering

(c) Successful Grasping

Fig. 1: **RF-Grasp grasping fully-occluded objects.** (a) Even though the camera and RF sensor have no line-of-sight to the object, RF-Grasp uses RF perception to identify and locate the target object through occlusions. (b) It employs *RF-visual servoing* and *RF-visual manipulation* to efficiently maneuver toward the object and declutter its vicinity. (c) Successful pick-up.

The key challenge with RF perception is that unlike vision, it cannot produce high-resolution images with pixel-wise precision. Rather, it only obtains the 3D tag location with centimeter-scale precision [9]. Moreover, because standard (visual) occlusions are transparent to RF signals, a robot can neither perceive them using RF sensing nor reason about the visual exploration, collision avoidance, and decluttering steps that may be necessary prior to grasping the target object.

We present RF-Grasp (shown in Fig. 1), the first robotic system that fuses RF and vision information (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 mechanical search and extended mechanical search.

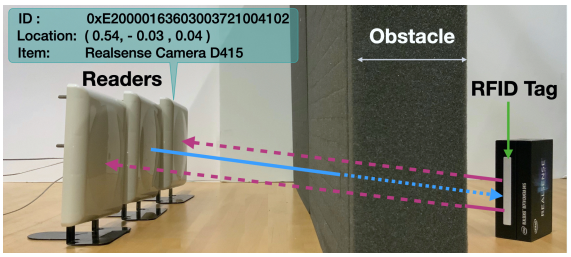


Fig. 2: **RFID-based Perception**. RF-Grasp uses a customized reader that can identify and accurately localize RFID-tagged objects through occlusions.

- It introduces two novel primitives: (1) *RF-visual servoing*, 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 of partially or fully occluded target objects [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

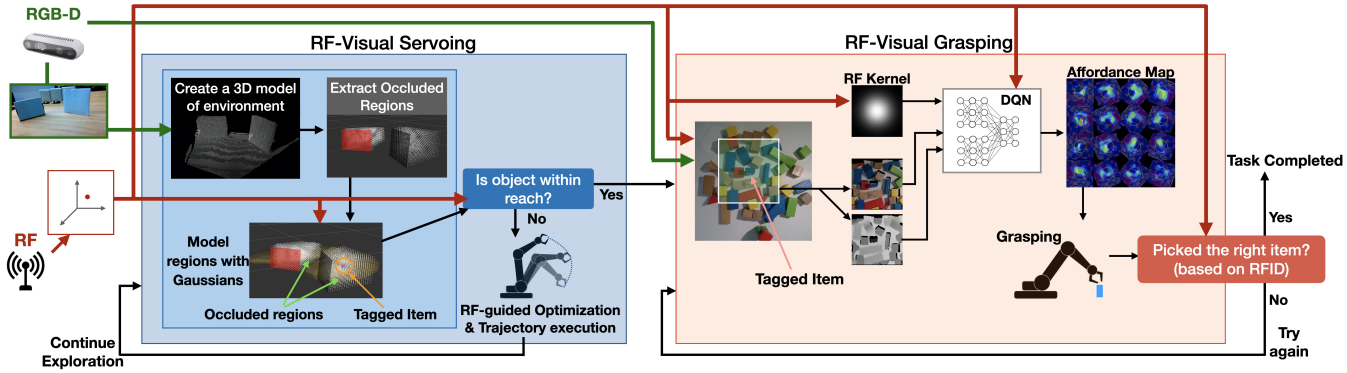


Fig. 3: **System Overview.** RF-Grasp uses RGB-D (green arrows) to create a 3D model of the environment and extract occluded regions, and it plugs the location of the RFID-tagged object (red arrows) into it. If the object is behind obstacles, it performs RF-guided active exploration and trajectory optimization to obtain a better view of the object by maneuvering around obstacles. Once a stopping criterion is met, it proceeds to grasping. Here, it applies RF-based attention to the RGB-D information and relies on a model-free deep-reinforcement network to discover optimal grasping affordances. RF-Grasp uses the RFID's locationto close the loop on the grasping task and determine whether it has been successful or whether it needs to make another grasping attempt.

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 ($\mathcal{S}$) consists of the robot joint state ($\mathbf{x}_t^R \in \mathbb{R}^6$), RFID location ($\mathbf{p} = (x_p, y_p, z_p)$), the obstacles, occlusions, and other objects. The control signal, $\mathbf{u}_t \in \mathbb{R}^6$, 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 $(\mathbf{x}^m, \mathbf{\Sigma}_0^m)$ at $t = 0$. The environment is assumed to be static; hence, the means remain the same over the planning horizon, but the covariances $\mathbf{\Sigma}_t^m$ 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:

$$C_{t=0:T}(\mathbf{x}_t^R, \Sigma_t^{1:M}, \mathbf{u}_t) = \alpha||\mathbf{u}_t||_2^2 + \sum_{m=1}^{M} \beta_t^m \text{tr}(\Sigma_t^m) \quad (1)$$

where M is the total number of occluded regions, tr is trace, α and β_t^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_t^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_t^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:

$$\begin{aligned} \min_{\mathbf{x}_{0:T}^R, \mathbf{u}_{0:T}} \quad & \mathbb{E}[\sum_{t=0}^{T} C_t(\mathbf{x}_t^R, \Sigma_t^{1:M}, \mathbf{u}_t)] \\ \text{s.t.} \quad & \mathbf{x}_{t+1}^R = \mathbf{f}(\mathbf{x}_t^R, \mathbf{u_t}, 0),\ \mathbf{x}_t^R \in \mathbb{X}_{\text{feasible}},\ \mathbf{x}_0^R = \mathbf{x}_{\text{init}}^R \\ & \mathbf{u}_t^R \in \mathbb{U}_{\text{feasible}},\ \mathbf{u}_T = 0 \end{aligned}$$

where $\mathbb{X}_{\text{feasible}}$ and $\mathbb{U}_{\text{feasible}}$ represent the set of feasible joint states and control signals of the robot arm, $\mathbf{x}_{\text{init}}^R$ is the initial joint state of the robot. The dynamics model for the robot is given by differentiable and stochastic function $\mathbf{x}_{t+1}^R = f(\mathbf{x}_t^R, \mathbf{u}_t, \mathbf{q}_t), \mathbf{q}_t \sim N(0, I)$ where $\mathbf{q}_t$ 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,[1] 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

[1]It is stored in a TSDF (Truncated Signed Distance Function) volume.

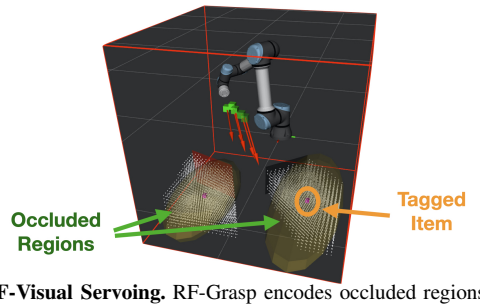


Fig. 4: **RF-Visual Servoing.** RF-Grasp encodes occluded regions as Gaussians and biases its active exploration toward the RFID-tagged target item.

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 Σ^1 and in visibility of the voxels υ in the vicinity of ν the target object's location $\mathbf{p}$. Mathematically, the termination criteria can be expressed as:

$$\text{trace}(\Sigma^1) < \rho_\Sigma \quad \text{or} \quad \sum_{\upsilon \in \nu(\mathbf{p})} F(\upsilon)/|\nu(\mathbf{p})| > \rho_\upsilon \qquad (2)$$

where $F(\upsilon) = 1$ if voxel υ has been seen by camera and 0 otherwise. The criteria imply the uncertainty around the object is smaller than threshold ρ_σ, or the visible fraction of the region around the RFID is larger than threshold ρ_υ.[2]

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 cover that occludes a large region, the trace of covariance won't be below ρ_Σ. However, enough voxels in the vicinity of the item will be visible to meet the second criterion 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 Σ_t^m 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 problem as a Markov Decision Process. Here, the action a_t is grasping with a parallel jaw gripper at position $\mathbf{g_t} = (x_{a_t}, y_{a_t}, z_{a_t})$ with gripper rotation of θ_{a_t}. The goal is to learn the optimal policy π^* to grasp the target item (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).

B. RF-based Attention and Rewards

RF-Grasp trains a deep reinforcement learning network in simulation, using RF information in the reward and attention

[2]In our implementation, $\rho_\Sigma = 0.005$ and $\rho_\upsilon = 0.1$, and the vicinity, $\nu(\mathbf{p})$, is set to a 5cm × 5cm × 10cm cube centered at RFID location.

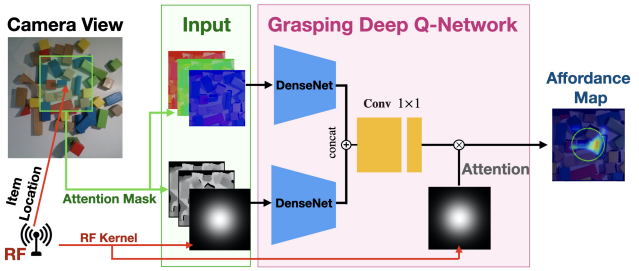


Fig. 5: **RF-Visual Grasping.** Using the RFID's location, RF-Grasp crops out the RGB-D data near the tagged object location, and feeds the cropped RGB-D data and RFID location probability as input to the Deep-NN that outputs the Q-value estimations. The rotated inputs in 16 different directions are separately fed to DNN to estimate Q-values of different grasping rotations.

mechanisms. Fig. 5 shows the overall architecture, consisting of feed-forward fully convolutional networks. The networks takes RGB-D and RFID position as input, and output pixel-wise map of Q values. The optimal policy is selected 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 $r(s_t, s_{t+1}, a_t)$ is 1 if the robot grasps the target object; $\min\left(\frac{\varrho}{||\mathbf{p}_t - \mathbf{g}_t||}, 1\right)$ if it grasps another item; and 0 otherwise. ϱ is chosen such that the maximum reward is given to any grasp point within the resolution of RFID positioning.[3] 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[4] are cropped to a square around the RFID's location.[5] 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 $\mathbf{p}$, whose standard deviation accounts for RF localization errors. The kernel is fed to the network, and is multiplied by 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 1×1 come after, producing the 2D map of the Q function estimation. The output contains 16

[3]In our implementation, ϱ is set to 0.007.

[4]Heightmaps are computed by extracting 3D point cloud from RGB-D images and projecting the images orthographically, parallel to the table top.

[5]The square is 11cm×11cm in our implementation.

maps for grasp, from which RF-Grasp chooses the position and rotation with highest probability of success.

Training Details. 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. We use Huber loss function and stochastic gradient descent with learning rates 10^{-4}, momentum 0.9, and weight decay 2^{-5} for training. 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 Real-Sense 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 RFIDs (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 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.

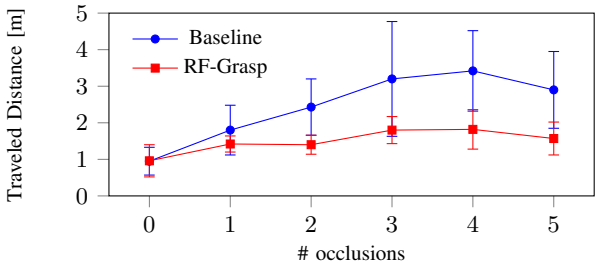


Fig. 6: **Traveled Distance.** The figure plots the average traveled distance of the gripper's tip for baseline (blue) and RF-Grasp (red) in successful trials across 0-5 occlusion. Error bars represent the standard deviation.

Evaluation Metrics. The robot workspace is atop a table with dimensions of 0.8m×1.2m. We consider three metrics: *1) Average Traveled Distance:* the distance that the robot's end-effector moves until grasping the target item. *2) Task Completion Rate:* the percentage of trials that successfully grasped the tagged item (or green item for baseline) before 10 grasping attempts and 5 meter of traveled distance in the exploration phase. *3) Grasping Efficiency:* defined as $\frac{\text{\# successful grasps}}{\text{\# total grasps}}$ over trials where exploration succeeded.

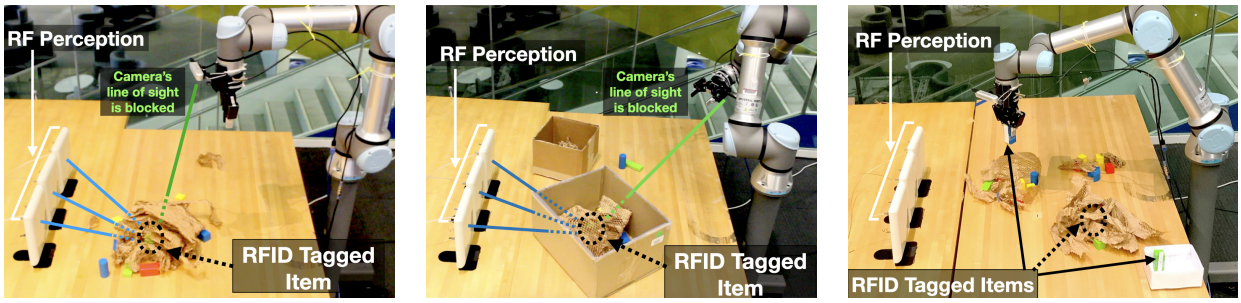
VII. Results

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 camera sees one or more occluded regions in its initial observation. Other frontiers and occluded regions were discovered during exploration.

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.



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

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

(b) Task Completion Rate: Next, we evaluated RF-Grasp's completion rate (defined in VI) 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.

	Scenario (M=#occlusions, N=#objects)					
System	M=0 N=5	M=1 N=1	M=2 N=5	M=3 N=10	M=4 N=13	M=5 N=15
Baseline	$10/10$	$10/10$	$10/10$	$8/10$	$7/10$	$8/10$
RF-Grasp	$10/10$	$10/10$	$10/10$	$10/10$	$10/10$	$10/10$

TABLE I: **Completion Rate.** In each scenario, M denotes number of occlusions, N number of blocks. Results are reported as the number of successfully completed trials out of the total number of experimental trials.

(c) Grasping Efficiency. We measured the grasping efficiency across successful trials. RF-Grasp has 78% efficiency, while the baseline has 68% efficiency. The improved grasping efficiency for RF-Grasp likely results from variations in lighting which impact color but not RF measurements.

B. Generalized Mechanical Search

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 the target item and more than 20 distractor objects 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. RF-Grasp was successful in all 5 trials we performed. The average travelled distance was 2.68m with standard deviation of 0.98.

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 with large obstacles and a cover (Fig. 7(b)). We used two boxes (with 5 items each) creating two occluded regions. We placed a tagged item in one box. RF-Grasp was successful in exploring the environment, maneuvering toward the target, removing the cover, and grasping the object. RF-Grasp was successful in all 3 trials we performed. The average travelled distance was 4.13 with standard deviation of 1.91. 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 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. Discussion & Conclusion

RF-Grasp fuses RF and visual information to enable robotic grasping of fully-occluded objects. RF-Grasp can be extended in multiple ways to overcome its current limitations: 1) By applying the RF attention to networks capable of grasping more complex items at different angles, this system could be extended to more complex setups. 2) The current system requires separate components for RF localization and grasping. This necessitates a calibration phase for RF-eye-hand coordination. Exploring fully integrated robotic systems could remove this requirement. 3) The underlying approach of this paper can be extended beyond mechanical search to other robotic tasks using RFIDs, including semantic grasping, scene understanding, and human-robot interaction.

Fundamentally, this work creates a new bridge between robotic manipulation and RF sensing, and we hope it encourages researchers to explore the intersection of these fields.

Acknowledgments. We thank the anonymous reviewers, the Signal Kinetics group, and Laura Dodds for their feedback. The research is sponsored by the National Science Foundation, NTT DATA, Toppan, Toppan Forms, and Abdul Latif Jameel Water and Food Systems Lab (J-WAFS).

REFERENCES

[1] M. Danielczuk, A. Angelova, V. Vanhoucke, and K. Goldberg, "X-ray: Mechanical search for an occluded object by minimizing support of learned occupancy distributions," *IROS*, 2020.

[2] M. Danielczuk, A. Kurenkov, A. Balakrishna, M. Matl, D. Wang, R. Martín-Martín, A. Garg, S. Savarese, and K. Goldberg, "Mechanical search: Multi-step retrieval of a target object occluded by clutter," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 1614–1621.

[3] T. Novkovic, R. Pautrat, F. Furrer, M. Breyer, R. Siegwart, and J. Nieto, "Object finding in cluttered scenes using interactive perception," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 8338–8344.

[4] Y. Yang, H. Liang, and C. Choi, "A deep learning approach to grasping the invisible," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2232–2239, 2020.

[5] R. A. Knepper, T. Layton, J. Romanishin, and D. Rus, "Ikeabot: An autonomous multi-robot coordinated furniture assembly system," in *2013 IEEE International conference on robotics and automation*. IEEE, 2013, pp. 855–862.

[6] G. Kahn, P. Sujan, S. Patil, S. Bopardikar, J. Ryde, K. Goldberg, and P. Abbeel, "Active exploration using trajectory optimization for robotic grasping in the presence of occlusions," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 4783–4790.

[7] D. Katz, A. Venkatraman, M. Kazemi, J. A. Bagnell, and A. Stentz, "Perceiving, learning, and exploiting object affordances for autonomous pile manipulation," *Autonomous Robots*, vol. 37, no. 4, pp. 369–382, 2014.

[8] M. Dogar, K. Hsiao, M. Ciocarlie, and S. Srinivasa, "Physics-based grasp planning through clutter," in *Proceedings of Robotics: Science and Systems VIII*, July 2012.

[9] Z. Luo, Q. Zhang, Y. Ma, M. Singh, and F. Adib, "3d backscatter localization for fine-grained robotics," in *16th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 19)*, 2019, pp. 765–782.

[10] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand, "Capturing the human figure through a wall," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 6, p. 219, 2015.

[11] "Impinj R420," http://www.imping.com, Imping Inc.

[12] R. Das, "RFID Forecasts, Players and Opportunities 2019-2029," IDTechx, 2019.

[13] J. Wang and D. Katabi, "Dude, where's my card? rfid positioning that works with multipath and non-line of sight," in *ACM SIGCOMM*, 2013.

[14] J. Wang, F. Adib, R. Knepper, D. Katabi, and D. Rus, "RF-Compass: Robot Object Manipulation Using RFIDs," in *ACM MobiCom*, 2013.

[15] L. Shangguan and K. Jamieson, "The design and implementation of a mobile rfid tag sorting robot," in *Proceedings of the 14th annual international conference on mobile systems, applications, and services*, 2016, pp. 31–42.

[16] Y. Ma, N. Selby, and F. Adib, "Minding the billions: Ultrawideband localization for deployed rfid tags," *ACM MobiCom*, 2017.

[17] S. Park and S. Hashimoto, "Autonomous mobile robot navigation using passive rfid in indoor environment," *IEEE Transactions on industrial electronics*, vol. 56, no. 7, pp. 2366–2373, 2009.

[18] W. Gueaieb and M. S. Miah, "An intelligent mobile robot navigation technique using rfid technology," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 9, pp. 1908–1917, 2008.

[19] M. Kim and N. Y. Chong, "Direction sensing rfid reader for mobile robot navigation," *IEEE Transactions on Automation Science and Engineering*, vol. 6, no. 1, pp. 44–54, 2008.

[20] T. Deyle, C. J. Tralie, M. S. Reynolds, and C. C. Kemp, "In-hand radio frequency identification (rfid) for robotic manipulation," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 1234–1241.

[21] T. Deyle, C. Anderson, C. C. Kemp, and M. S. Reynolds, "A foveated passive uhf rfid system for mobile manipulation," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2008, pp. 3711–3716.

[22] T. Deyle, M. S. Reynolds, and C. C. Kemp, "Finding and navigating to household objects with uhf rfid tags by optimizing rf signal strength," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 2579–2586.

[23] T. Deyle, H. Nguyen, M. Reynolds, and C. C. Kemp, "Rf vision: Rfid receive signal strength indicator (rssi) images for sensor fusion and mobile manipulation," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 5553–5560.

[24] A. Zeng, S. Song, S. Welker, J. Lee, A. Rodriguez, and T. Funkhouser, "Learning synergies between pushing and grasping with self-supervised deep reinforcement learning," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 4238–4245.

[25] D. Morrison, P. Corke, and J. Leitner, "Multi-view picking: Next-best-view reaching for improved grasping in clutter," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 8762–8768.

[26] A. Zeng, S. Song, K.-T. Yu, E. Donlon, F. R. Hogan, M. Bauza, D. Ma, O. Taylor, M. Liu, E. Romo, *et al.*, "Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 1–8.

[27] S.-K. Kim and M. Likhachev, "Planning for grasp selection of partially occluded objects," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 3971–3978.

[28] E. Jang, S. Vijayanarasimhan, P. Pastor, J. Ibarz, and S. Levine, "End-to-end learning of semantic grasping," in *Conference on Robot Learning*, 2017, pp. 119–132.

[29] X. Huang, I. Walker, and S. Birchfield, "Occlusion-aware reconstruction and manipulation of 3d articulated objects," in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 1365–1371.

[30] A. Price, L. Jin, and D. Berenson, "Inferring occluded geometry improves performance when retrieving an object from dense clutter," *International Symposium on Robotics Research (ISRR)*, 2019.

[31] A. Aydemir, K. Sjöö, J. Folkesson, A. Pronobis, and P. Jensfelt, "Search in the real world: Active visual object search based on spatial relations," in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 2818–2824.

[32] R. Bajcsy, "Active perception," *Proceedings of the IEEE*, vol. 76, no. 8, pp. 966–1005, 1988.

[33] J. Bohg, K. Hausman, B. Sankaran, O. Brock, D. Kragic, S. Schaal, and G. S. Sukhatme, "Interactive perception: Leveraging action in perception and perception in action," *IEEE Transactions on Robotics*, vol. 33, no. 6, pp. 1273–1291, 2017.

[34] C. Nam, J. Lee, Y. Cho, J. Lee, D. H. Kim, and C. Kim, "Planning for target retrieval using a robotic manipulator in cluttered and occluded environments," *arXiv preprint arXiv:1907.03956*, 2019.

[35] Y. Cui, J. Ooga, A. Ogawa, and T. Matsubara, "Probabilistic active filtering with gaussian processes for occluded object search in clutter," *Applied Intelligence*, pp. 1–15, 2020.

[36] K. Wada, K. Okada, and M. Inaba, "Joint learning of instance and semantic segmentation for robotic pick-and-place with heavy occlusions in clutter," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 9558–9564.

[37] C. Chen, H.-Y. Li, X. Zhang, X. Liu, and U.-X. Tan, "Towards robotic picking of targets with background distractors using deep reinforcement learning," in *2019 WRC Symposium on Advanced Robotics and Automation (WRC SARA)*. IEEE, 2019, pp. 166–171.

[38] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," *arXiv preprint arXiv:1511.05952*, 2015.

[39] Ettus Research, CDA-2990, https://moveit.ros.org/.

[40] The Open Motion Planning Library, https://ompl.kavrakilab.org/.

[41] Yak, https://github.com/ros-industrial/yak.

[42] The Point Cloud Library (PCL), https://pointclouds.org/.

[43] Forces Pro, https://www.embotech.com/.

[44] Alien Technology Inc., "ALN-9640 Squiggle Inlay," www.alientechnology.com.

[45] VREP, https://www.coppeliarobotics.com/.

[46] Gazebo, http://gazebosim.org/.

[47] Rviz, http://wiki.ros.org/rviz.