Augmenting Augmented Reality with Non-Line-of-Sight Perception

Tara Boroushaki [1], Maisy Lam [1], Laura Dodds [1], Aline Eid [1,2], Fadel Adib [1]
[1] *Massachusetts Institute of Technology,* [2] *University of Michigan*
tarab@mit.edu, mllam@mit.edu, ldodds@mit.edu, alineeid@umich.edu, fadel@mit.edu

Abstract – We present the design, implementation, and evaluation of X-AR, an augmented reality (AR) system with non-line-of-sight perception. X-AR augments AR headsets with RF sensing to enable users to see things that are otherwise invisible to the human eye or to state-of-the-art AR systems. Our design introduces three main innovations: the first is an AR-conformal antenna that tightly matches the shape of the AR headset visor while providing excellent radiation and bandwidth capabilities for RF sensing. The second is an RF-visual synthetic aperture localization algorithm that leverages natural human mobility to localize RF-tagged objects in line-of-sight and non-line-of-sight settings. Finally, the third is an RF-visual verification primitive that fuses RF and vision to deliver actionable tasks to end users such as picking verification. We built an end-to-end prototype of our design by integrating it into a Microsoft Hololens 2 AR headset and evaluated it in line-of-sight and non-line-of-sight environments. Our results demonstrate that X-AR achieves decimeter-level RF localization (median of 9.8 cm) of fully-occluded items and can perform RF-visual picking verification with over 95% accuracy (F-Score) when extracting RFID-tagged items. These results show that X-AR is successful in extending AR systems to non-line-of-sight perception, with important implications to manufacturing, warehousing, and smart home applications. Demo video: y2u.be/bdUN21ft7G0

1 Introduction

The past few years have witnessed an increasing interest in augmented reality (AR) systems. Major tech companies - including Microsoft, Meta, Apple, and Google - have invested billions of dollars in developing AR technologies [7, 25, 52, 51]. A significant driver for these investments is the role that AR systems are expected to play in boosting efficiency across Industry 4.0 sectors including manufacturing, warehousing, logistics, and retail. For example, in e-commerce warehouses, AR headsets can boost labor efficiency by guiding workers in picking, sorting, and packing orders and returns [26]. Similarly, in manufacturing settings, AR headsets can guide employees by visualizing assembly tasks, automatically labeling tools in the environment, and helping users find parts they need [28]. More generally, AR headsets are expected to make workers more efficient by annotating their environments, visualizing their next tasks, and guiding them in executing these tasks [27, 30].

To realize their full potential, AR headsets need to deliver the above capabilities in real-world industrial environments, which are typically dense and highly cluttered. For example, a typical warehouse or dark store is dense with packages, and a standard manufacturing plant is dense with materials and compartments. In these environments, the majority of items are occluded due to being inside a box, under a pile, or behind other packages. Such occlusions make it difficult for existing headsets to perceive these items, which in turn prevents them from identifying and locating the items or guiding workers towards them. This limitation stems from the fact that today's AR headsets perceive their environment through cameras or other vision-based sensing systems which are inherently limited to line-of-sight (LOS) [6, 38]. Such line-of-sight restriction hinders AR systems from boosting worker efficiency where it is most needed, namely in cluttered and dense industrial environments.

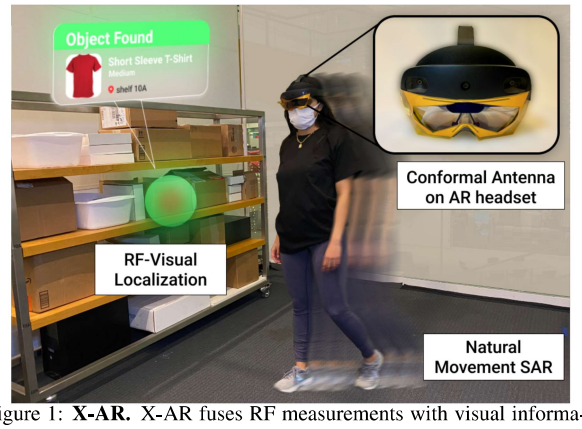


Figure 1: **X-AR.** X-AR fuses RF measurements with visual information and leverages natural human motion to localize RFID tagged items in the environment. The system uses a custom-designed, conformal, light-weight antenna mounted on an augmented reality headset and displays information to the user.

In this paper, we ask the following question: *Can we design and build an augmented reality system that can sense fully-occluded objects and expand the perception of humans beyond the line of sight?* With this capability, augmented reality would go beyond any natural human ability and truly augment the way we interact with the world, enabling significant advances in warehouse logistics, manufacturing, retail, and more. For example, AR headsets with non-line-of-sight (NLOS) perception could identify and localize specific items (e.g., customer orders, tools, materials) even when they are fully occluded, helping workers avoid a lengthy search process. Additionally, such AR headsets could be used to automate inventory control of items in warehouses or retail stores without needing to see all objects, and can alert

workers to misplaced items hidden behind occlusions.

To realize this vision, our approach is to leverage Radio Frequency (RF) signals, which, unlike visible light, can traverse everyday occlusions such as cardboard boxes, plastic containers, wooden dividers, and clothing fabric. Indeed, recent advances in RF sensing have demonstrated the potential to use RF signals to sense and accurately localize items in non-line-of-sight and highly cluttered environments [17, 40, 39]. Among existing RF sensing technologies, we are particularly interested in leveraging UHF RFID (Radio Frequency IDentification) tags due to their widespread adoption in supply chain industries (for example, over 93% of US retailers have adopted UHF RFIDs [5]). Our vision is to bring RFID sensing and localization to AR headsets to provide them with non-line-of-sight perception and augment human visual abilities for applications in warehouse automation, e-commerce fulfillment, and manufacturing.

We would like to build a system that realizes the above vision while satisfying the following requirements:

1. *AR-compatibility:* The system must seamlessly integrate with an AR headset without impacting the performance of its existing sensors and displays (i.e., without obstructing the headset cameras or the user field of view).

2. *Seamless Mobility:* The system must operate correctly with natural human mobility. Specifically, it must be able to accurately localize RFIDs (in LOS and NLOS settings) without requiring the user to perform unnatural movement patterns, which may hinder their productivity.

3. *Actionability:* The system should provide users with actionable tasks (e.g., guide the user where to search) and inform users of task success (e.g., verify to a warehouse picker whether or not they picked the right order).

4. *User-friendliness:* The system needs to be compact and lightweight, so that the user can easily wear the AR headset and move around to complete their tasks.

Satisfying the above requirements is challenging and cannot be realized by simply integrating a state-of-the-art RFID localization system with an AR headset. In particular, the majority of accurate RFID localization solutions require multiple antennas that are separated by meter-scale distances [39, 40], making them too bulky to mount onto an AR headset. Solutions that don't require such antenna arrays typically rely on robot-mounted antennas that need to be moved on predefined trajectories [58, 17, 15], making them incompatible with natural human mobility. In addition to these challenges, delivering AR-actionable tasks goes beyond simple RF localization and requires new mechanisms to fuse RF and vision perception under natural mobility and display the output on the headset.

In this paper, we present X-AR, an augmented reality headset with a built-in RF sensing system. A user wearing X-AR can freely walk in their environment (e.g., a warehouse or manufacturing plant), and the headset automatically identifies and localizes items in the environment, even when they are not the in line-of-sight. Using this information, X-AR guides the worker towards items of interest (tools, packages, etc.) and verifies whether or not they have picked up the correct item. Our system introduces multiple innovations that together allow it to satisfy the above requirements:

1. AR-Conformal Wideband Antenna: X-AR introduces the design of an ultra-lightweight and wideband antenna that is conformal to the headset (described in §3). Our unique antenna design matches the shape of the AR glasses visor, as depicted in Fig. 1, and does not block the user's view or any sensors. The antenna also achieves the radiation, bandwidth (BW), and gain properties required to perform accurate RFID localization.

2. RF-Visual Synthetic Aperture Radar: X-AR does not make restrictive assumptions about the user's motion pattern when localizing the RFID tags in the environment, and opportunistically leverages natural human mobility. To do this, X-AR first uses the visual information from the AR headset camera to self-localize in the environment. It then uses the RFID measurements collected during the user's motion to create a synthetic aperture radar (SAR) and localize RFID tagged items with high accuracy. In addition, X-AR introduces a number of techniques to handle localization artifacts and constraints that arise from natural human motions such as natural head tilts and RFID backscatter radiation properties. We describe this localization method in detail in §4 and show how the system guides the user to the item's location and displays it on the AR headset.

3. RF-Visual Verification: The final component of X-AR's design is a mechanism that verifies when the user has picked up their desired RFID-tagged item. Such verification is important to avoid costly errors such as picking and shipping the wrong order to a customer in e-commerce warehouses. One might assume that such a capability can be simply realized by localizing the RFID-tagged target item to within a user's hand once they've picked it up (i.e., using the same localization mechanism described up). In practice, doing so is challenging because, unlike the above scenario where the user's walking emulates a synthetic aperture, a user picking an item stays in a relatively fixed location. To address this challenge, X-AR leverages the RFID tag's mobility instead. Specifically, it performs a *reverse* SAR to localize the headset with respect to the picked item's trajectory. In §5, we describe this technique in detail and show how X-AR fuses the AR-headset's hand-tracking capability with reverse SAR to perform the verification with high accuracy.

We implemented an end-to-end prototype of X-AR. We mounted a custom-designed conformal antenna on

Microsoft Hololens 2 headset [9]. The antenna is connected to bladeRF software defined radios [47] that communicate with the AR headset through an edge server. Our algorithms are written in the C driver and operate in real-time, and we program the Hololens through Unity to display item locations and labels, guide the user to the target items, and show the verification results.

We evaluated X-AR's performance over 230 experimental trials. Our evaluation demonstrated that:

1. X-AR's conformal antenna achieves all desired specifications in terms of weight ($<$1g), size (conformal), BW (200 MHz), and gain (around 0 dB).

2. X-AR accurately localizes RFID tagged objects in line of sight and non line of sight scenarios with a median accuracy of 9.8 cm. Even the 90^{th} percentile accuracy remains within a foot and a half (45 cm). In contrast, a standard SAR-based baseline has more than double the error, achieving a median accuracy of 24.8 cm and a 90^{th} percentile accuracy of 99.1 cm.

3. X-AR tracks the movement of the RFID tag and the user's hand to automatically verify what object has been picked up with over 95% accuracy (F-Score). Even if the user picks up a box with the RFID-tagged item inside it (rather than picking up the item itself), the F-Score remains over 91%.

Contributions: X-AR is the first augmented reality headset that can sense through occlusions and perceive fully occluded objects. This system introduces three key innovations: 1) A custom-designed, conformal, wideband, and light-weight antenna that can be integrated with a commercial AR headset, enabling RFID localization without obstructing the user's or cameras' view, 2) An RFID localization system that opportunistically leverages natural user motion to create a non uniform RF-Visual synthetic aperture radar and to localize and visualize the RFID tagged objects in 3D, 3) A verification mechanism that performs reverse RF-Visual localization to verify whether the user has picked the target item, in line-of-sight, non-line-of-sight, or occluded settings.

Although X-AR enables non-line-of-sight perception for augmented reality headsets, our current implementation still has a few limitations. First, X-AR is currently designed to operate on a single headset, and still has no mechanisms to extend to multiple coordinated headsets. Second, the range of the RF measurements are limited to 3m; however, future antenna design iterations can achieve an even longer range. Finally, X-AR only demonstrates two actionable tasks: guiding the user toward a target item, and verifying the target item is in the user's hand. As research evolves, it would be interesting to extend this system to other tasks. Despite these limitations, X-AR marks an important step in bringing RF sensing to AR and opens the door to future works bridging these fields.

2 System Overview

X-AR is a next-generation augmented reality system capable of perceiving objects in both LOS and NLOS conditions. The system can identify, locate, and label RFID-tagged items in the environment. It leverages an RF sensing module to read passive, off-the-shelf UHF RFID tags attached to items of interest. By combining this information with visual data from the AR headset's camera sensors, it locates RFID-tagged items with high accuracy.

X-AR is designed to be used in practical environments, such as warehouses, manufacturing plants, and retail stores. It opportunistically leverages human motion (i.e., as the user walks around and picks up items) in order to localize tagged items in the environment, guide the user towards them, and verify when the user has picked them up. For simplicity, the remainder of this paper discusses the system in a single tag scenario. sHowever, X-AR can easily extend to multiple RFID tags in the environment. Using the EPC Gen 2 protocol, X-AR can read each RFID tag separately, and perform the same localization and verification algorithms for each tag.

3 AR-Conformal Antenna

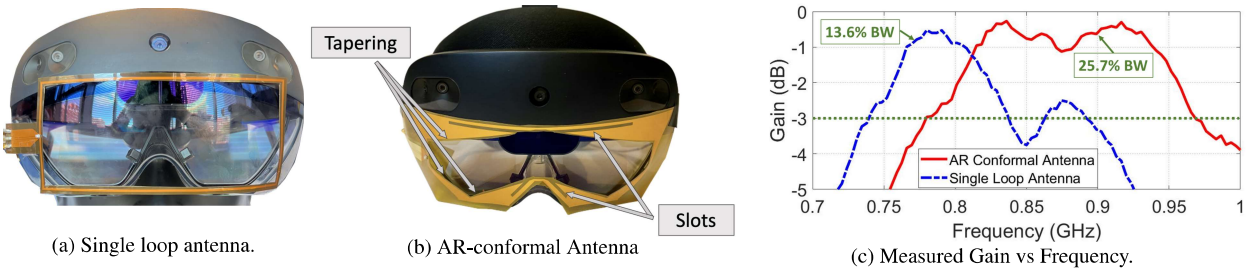
X-AR introduces a conformal antenna that can be mounted on the headset to identify, locate, and verify UHF RFID tags, without interfering with the headset's operation or constraining the user. Here, we describe our AR-conformal antenna design, its requirements, challenges, and the path describing its evolution from a conventional antenna structure to one satisfying all the desired needs. To perform RFID localization from the headset in the field of view of the user, the antenna needs to satisfy the following requirements:

• **Wideband operation around 900 MHz:** The antenna needs to maintain a matched operation and a good gain over a BW of at least 200 MHz to match the bandwidth requirements of state-of-the-art RFID localization systems [39, 40].

• **Conformal and unobstructive:** The antenna must be designed on a flexible substrate to easily conform to the Hololens' visor without obstructing a user's field of view or the cameras mounted on the front of the headset.

• **Lightweight and small form-factor:** The antenna must maintain an ultra-light weight ($<$ 1g) and be simple and easy to mount on the AR's visor.

Existing solutions in state-of-the-art wideband RFID localization systems do not satisfy these properties [16, 17, 40, 39, 14]. In particular, they rely on rigid, relatively-large, and often bulky antennas. For example, the majority of these systems leverage large patch antennas that are 26 $cm \times 26$ $cm \times 3.3$ cm and weigh approximately 1.04 kg, while others rely on log-periodic antennas that measure 15 $cm \times 13$ $cm \times 0.01$ cm. These solutions are too bulky and would obstruct the field of



(a) Single loop antenna. (b) AR-conformal Antenna (c) Measured Gain vs Frequency.

Figure 2: **Conformal Antenna Design.** (a) Fabricated single-loop antenna mounted on the headset (dimensions 122×51mm). (b) Fabricated conformal antenna (dimensions 165×64mm). (c) These plots show the measured gains of the single loop (blue) and AR conformal (red) antennas vs frequency while mounted on the headset and worn by the user. The horizontal green line is used to highlight the 3-dB bandwidth.

view of the AR headset, thus are ill-suited for our use case. While some RFID localization systems utilize compact and lightweight antennas [58, 57, 59], these antennas have a narrow band of operation, which makes them unsuitable for our use case.[1]

Below, we describe our investigation in designing the AR-conformal antenna to satisfy the above requirements.

3.1 Investigating a Single Loop Design

We first investigated whether we could achieve the above properties using a single loop antenna design. Our choice of a single loop was motivated by the fact that a loop can wrap around the outline of the visor, delivering a small form factor not obstructing the field of view. Also, the loop is a simple antenna that does not require a ground plane, making it easy to mount on an AR headset.

Fig. 2a shows the picture of our initial design of a loop antenna, fabricated on a 100 µm thin polyimide substrate, and mounted on the Hololens. Notice how our antenna (almost) follows the perimeter of the visor, thus not obstructing its view. In order to identify the appropriate dimensions corresponding to an operation around 900 MHz, we performed our antenna simulations in Ansys High Frequency Simulation Software (HFSS). In designing these antennas, we leveraged polyimide films because of their good electromagnetic properties, their common use for applications requiring flexible electronics, and their wide availability at low-cost. This antenna also weighs less than 1g, thus it satisfies our requirements of a small form factor while maintaining a light weight.

To investigate the bandwidth requirement, we mounted the antenna on the headset, worn by the user, and measured its gain over the frequency of interest. This was done by illuminating it with a transmitter antenna of a known gain and using a vector network analyzer (VNA) to extract the S parameters of the loop antenna (specifically the S21 parameter).

Fig. 2c plots the gain of the loop with respect to frequency, showing 3 dB BW of approximately 100 MHz around 780 MHz. This shows the loop antenna design would not allow us to achieve the desired 200 MHz of BW. It should be noted the loop antenna delivers a resonant frequency of 900 MHz and a gain of 3.8 dB when tested in air. However, its gain degraded by 3 dB and frequency detuned by 120 MHz when placed on the headset visor and worn by the user. This behavior is often observed with wearable antennas [32, 41], where the frequency of operation and antenna radiation properties degrade when mounted on a new material. Thus, while the loop antenna is conformal, unobstructive, lightweight, and small, it did not satisfy the BW requirements.

[1] As mentioned in §1, past systems that leverage these antennas require either bulky arrays or a robot to move the antenna on a pre-defined trajectory, neither of which are suitable for an AR localization system.

3.2 Wideband AR-Conformal Antenna

Motivated by a desire to increase the bandwidth of the single-loop conformal antenna, we investigated how strategies such as tapering (i.e., gradually changing the width of the loop) and slotting (i.e., adding slotted gaps in the loop) can help us achieve the desired bandwidth. Through an iterative design and simulation processing (whereby the simulation was performed using Ansys HFSS), we reached the design shown in Fig. 2b. Notice how we carefully chose the dimensions of the antenna to perfectly match the shape of the visor, without blocking any of the cameras. We also added tapers to the outline of the antenna and integrated slots on the top and bottom lines around the nose to achieve a wideband operation.

Similar to the loop antenna, we conducted gain measurements to assess the 3-dB bandwidth of our conformal antenna while mounted on the headset and worn by the user. The red plot in Fig. 2c shows the gain of our new antenna as a function of frequency. Notice how the 3 dB bandwidth of the gain is now 200 MHz - from 775 MHz to 975 MHz. This shows that the antenna achieves the desired gain pattern in the frequency range of interest. Note that the negative gain realized by these wearable antennas is normal with ultra-thin substrates due to close proximity with lossy material such as the headset and human tissues [48, 19]. In principle, it is possible to further optimize the gain of the antenna, however, the negative gain could be easily overcome by transmitting at a higher power, thus maintaining a constant effective radiation pattern (typically referred to as EIRP). It should be also noted that the detuned frequency observed with

the measured loop due to placement on headset was accounted for in the HFSS simulations for the new antenna, by simulating the structure on plexiglass that mimics the headset's visor, and thus resulted in the proper resonant frequency during measurements. Finally, we also simulated the radiation pattern of the conformal AR antenna on the headset as well as measured its gains across frequencies and elevation angles (see appendix).

It should be noted that while this antenna was designed to match this headset, the design could be adapted for different visor shapes, depending on the location of the cameras and other components that cannot be blocked.

4 RF-Visual Synthetic Aperture Radar

In the previous section, we described the custom, conformal, and lightweight antenna that X-AR uses to sense RFID tags in the environment. In this section, we describe how X-AR uses these RFID measurements, along with visual information from the AR headset's camera to locate RFID tags with high accuracy through RF-Visual Synthetic Aperture Radar (SAR). For ease of exposition, we discuss localizing a single tag, but the same approach generalizes to any number of tags in the environment.

4.1 Background on SAR

X-AR's localization builds on a technique called Synthetic Aperture Radar (SAR). At a high level, SAR leverages the same localization principle as antenna arrays, where measurements from multiple antenna locations are combined to localize a wireless device in 2D or 3D space. SAR differs from standard antenna arrays in that it moves a single antenna, collecting measurements from different physical locations to emulate an antenna array. Formally, we can estimate the power P received from every point in space using the following equation:

$$P(x,y,z) = \left|\left| \frac{1}{M}\frac{1}{N}\sum_{j=1}^{M}\sum_{i=1}^{N} h_{i,j} e^{\frac{4\pi d_i(x,y,z)}{\lambda_j}} \right|\right| \quad (1)$$

where M is the number of frequencies used, $h_{i,j}$ is the channel measurement of the i$^{\text{th}}$ location with the j$^{\text{th}}$ frequency, d_i is the distance from (x,y,z) to the i$^{\text{th}}$ location, and λ_j is the wavelength of j$^{\text{th}}$ frequency.

To localize the tag, we find the (x,y,z) location with the highest power. Formally, the location of the tag, p_{tag}:

$$p_{tag} = \texttt{argmax}_{(x,y,z)}(P(x,y,z)) \quad (2)$$

For more details on SAR please refer to the Appendix.

4.2 AR-Based SAR

Since it is infeasible to mount an antenna array on an AR headset, X-AR builds on SAR-based RFID localization. Specifically, X-AR opportunistically leverages natural human motion to collect wideband measurements from different locations and uses them to construct a synthetic aperture radar to localize RFID tagged items.

However, bringing SAR to an AR headset faces a number of challenges. Unlike prior systems that leverage SAR (e.g., robots or airplanes), X-AR cannot rely on a constant velocity or predictable trajectory. For example, humans naturally accelerate and decelerate and move slightly side-to-side as they walk, making it difficult to predict the exact antenna location. Moreover, recall that X-AR aims to opportunistically leverage human motion as opposed to controlling the user's trajectory, making it even more challenging to control the antenna's location.

Self-Tracking. To address these challenges and localize the antenna over time, X-AR leverages the AR headset's built-in self-tracking capability. Existing AR headsets can self localize by extracting feature points from their cameras' visual data and performing visual-inertial odometry (VIO). They then track these points over time to build a map of the environment and derive their 6D pose (i.e., location and rotation) within this map [38, 42].

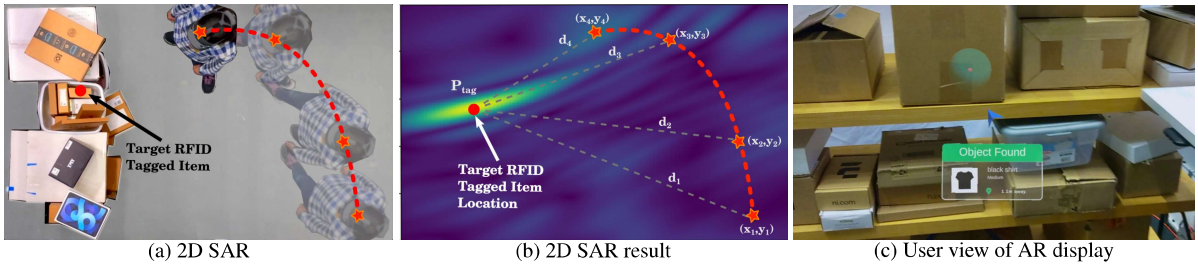
To leverage this built-in localization, X-AR requires an additional transformation. Specifically, the headset tracks its location as the center of the user's head, but the antenna is mounted on the front of the visor. This transform is essential since SAR relies on small changes in the RFID channel and therefore requires precise locations. This transform can be formulated as [55]:

$$\begin{aligned} {}^{W}P^{A} &= {}^{W}R^{H} \times {}^{H}P^{A} + {}^{W}P^{H} \\ {}^{W}R^{A} &= {}^{W}R^{H} \end{aligned} \quad (3)$$

where ${}^{W}P^{A}$ and ${}^{W}R^{A}$ are the position (x,y,z) and quaternion rotation of the antenna in the world frame W; ${}^{W}P^{H}$, ${}^{W}R^{H}$ are the position and quaternion rotation of the Hololens in the world frame. The x,y,z translation from the Hololens H to the antenna A is defined as ${}^{H}P^{A}$. The position and rotation of the Hololens are obtained from the vision-based AR self-tracking. We empirically measure the translation from the Hololens's center to antenna (${}^{H}P^{A}$) since this translation is fixed and results from mounting the antenna on the headset.

After applying the transformation, X-AR uses them as the antenna array locations. This allows it to then exploit wideband measurements as per Eq. 1 to opportunistically apply SAR along the user's trajectory.

RFID Localization. Fig. 3 shows an example of X-AR's RFID localization (shown in 2D for simplicity). Fig. 3a shows an overhead view of a user walking through the environment. RFID measurements are taken during the user's trajectory, resulting in the measurement locations shown by the red stars. These measurements are then used to compute the power at each point in the workspace using Eq. 1 to estimate the tag's location. Fig. 3b shows this power as a heatmap with yellow indicating areas of higher power and blue indicating areas of lower power. The tag's location (red dot) overlaps with the area of highest power, showing that the localization was successful. While the above description focused on 2D localiza-



(a) 2D SAR (b) 2D SAR result (c) User view of AR display

Figure 3: **AR-Based SAR.** (a) As the user moves naturally, X-AR collects RF measurements. (b) Using RF-Visual SAR, X-AR creates a heatmap of the RFID tag's possible location. The target RFID location overlaps with the area of highest power (yellow), indicating a successful localization. (c) The user's view from the Hololens application. The sphere shows the estimated tag location and the arrow points to it.

tion, the same method extends to 3D as per Eq. 2, enabling X-AR to localize items in 3D space.

Holographic Visualization. Once the item has been localized, X-AR leverages holographic visualization to display its location to the user and guide them towards it. To do this, X-AR leverages the transforms described in Eq.3 to compute the tag's location in the world frame. Fig. 3c shows an example from the user's perspective as displayed on the AR headset. In this example, X-AR places a spherical hologram around the estimated location, and a floating arrow appears in order to guide the user towards the localized tag for object retrieval. The arrow is programmed to float slightly above the user's eye level at a fixed distance in front of them. For every frame update, the application queries the location and rotation of the user in the world space. It then computes their directionality to update the pointing vector of the arrow to properly guide the user towards the target object.

4.3 Practical Considerations

Standard wideband SAR systems typically design their antennas to have uniform gain across the entire frequency band. However, off-the-shelf UHF RFID tags are not designed to be wideband and therefore have significant variability in their antenna gain across frequency. In general, measurements with frequencies further from the tag's resonant frequency (typically 900MHz) will be weaker and therefore more susceptible to noise. These weak measurements can introduce significant error in the location estimate. To overcome this, we introduce a weighted SAR formulation that biases the estimation towards confident measurements to improve the accuracy.

To do this, X-AR starts by quantifying its confidence in each of its measurements using the signal-to-noise ratio (SNR). For any wideband measurement with an average SNR below a certain threshold, X-AR is unlikely to be able to accurately estimate the RFID channel and it therefore removes the measurement from the SAR formulation entirely. The remaining measurements all contain useful information, however, as described above, certain frequencies in each wideband measurement may have weaker responses due to the tag's frequency dependent response. To prioritize frequencies with stronger responses, X-AR applies an SNR-based weighting function to each frequency in a measurement.

This is formalized in the following equation:

$$P(x,y,z) = \left|\left| \sum_{j=1}^{M} \sum_{i=1}^{N} \begin{cases} w_{i,j} h_{i,j} e^{\frac{4\pi d_i}{\lambda_j}} & \overline{SNR_i} > \tau \\ 0 & \overline{SNR_i} < \tau \end{cases} \right|\right| \quad (4)$$

$$w_{i,j} = \frac{SNR_{i,j}}{\max_{k \in [1,M]}(SNR_{i,k})} \quad (5)$$

where $w_{i,j}$ is the weight for the ith location and jth frequency, and τ is the SNR threshold for removing poor measurements. $SNR_{i,j}$ is the SNR of the ith location with the jth frequency, and $\overline{SNR_i}$ is the average SNR across all frequencies for the ith location. [2]

A few additional points are worth noting:

• In practice, the self-localization frame rate is different from that of the RFID channel measurements. To overcome this, X-AR linearly interpolates between Hololens self-tracked locations to find the corresponding location of the mounted antenna for any given measurement.

• X-AR continues to collect measurements until it has become confident in the tag's location. To determine its confidence, it finds all (x,y,z) locations whose power is within 0.75dB of the peak power.[3] It then computes a bounding box around these locations. When this bounding box's size falls below a threshold, X-AR declares the localization complete and visualizes the location.

5 RF-Visual Verification

So far, we explained how X-AR opportunistically leverages human motion to localize RFID-tagged target items and visualize them on the AR headset for retrieval. In principle, this visualization should be sufficient to indicate to the user to pick up the item within the holographic sphere shown in Fig. 3c. In practice, however, the user may still pick up an incorrect item. For example, multiple items may lie within the glowing sphere.[4] Even if the user knows what they're looking for (e.g., red shirt), there might be several items that are visually similar to each other or in similar packaging in the region. More generally, the picked item may be incorrect because the

[2] When computing $w_{i,j}$ in our implementation, we offset all of the SNR values and clip them at 0 to avoid negative weights.

[3] In practice, other thresholds are possible, but a looser threshold would reduce the confidence and hence the localization accuracy.

[4] The size of the sphere is determined by the confidence interval from RFID localization accuracy which is around 10-20 cm.

picker is prone to human error.

To ensure that the user has picked up the correct item, X-AR incorporates a mechanism for picking verification. We describe *RF-Visual Verification*, which enables an accurate and seamless verification of grasped items.

5.1 RF-Visual R-SAR

At the most basic level, the goal of X-AR's RF-Visual verification primitive is to verify whether the correct item is *in* the user's hand after they have picked up an object. Said differently, it aims to localize the RFID-tagged target item to within the user's palm. At first blush, one might assume that such a capability is trivial given that X-AR already has a mechanism to localize RFIDs, as described in the previous section on RF-Visual SAR. However, the two localization problems are fundamentally different. Unlike the earlier scenario where the user's walking emulates synthetic aperture, a user picking an item is in a relatively fixed location. Hence, one cannot leverage the user's movements to localize the item.

To localize the item despite the user's stationary position, X-AR leverages the RFID tag's mobility instead. Fig. 4a shows a sample scenario, demonstrating how the tag itself traces an antenna array. X-AR leverages this emulated array in order to localize the AR-headset (more specifically, the antenna on the headset) with respect to the array. This formulation is the reverse of the SAR described in §4, where the RFID tag was stationary, and the AR conformal antenna on the headset was moving with the user. Notably, in §4, we could leverage the AR headset's self-tracking capability to track the antenna locations. Here, we still need a mechanism to track the RFID locations in order to properly apply the antenna array equations.[5] To track the RFID's location as it moves, our idea is to leverage the *hand-tracking capability* of the AR headset. Specifically, AR headsets like the Microsoft Hololens 2 can detect and track multiple feature points on a user's hand, including their palm [8]. Thus, if the user picks up the correct RFID-tagged item, then the RFID traces a similar trajectory to the user's palm as shown in Fig. 4a.

X-AR leverages the above observation and applies the antenna array equations on the hand's trajectory in order to localize the headset. If the headset's estimated location using this method coincides with the headset's visual-inertial odometry-based location, that indicates that the target RFID tagged item was accurately retrieved and is indeed in the user's hand. On the other hand, if the headset localization fails, the failure indicates that the target RFID tag is not in the user's hand. Below, we formalize the above intuition by describing scenarios where the user picks the correct item and compare it to a scenario where the user picks an incorrect item.

[5] In principle, one could use ISAR [11]. It is less desirable than SAR because the former suffers from a larger direction location ambiguity.

(a) Scenario where the User Picks the Correct Item. Fig. 4a shows an example where the target item is in the user's hand. Here, the palm location (P_{palm}) and the tag location (P_{tag}) are similar. As the user's hand moves, P_{palm} and P_{tag} change similarly together. As a result, the target tag's location can be accurately approximated with the palm location over time for applying SAR and estimating the AR conformal antenna's location according to the following equation:

$$P(x,y,z) = || \sum_{j=1}^{M} \sum_{i=1}^{N_v} h_{ij} e^{\frac{4\pi d(t_i)}{\lambda_j}} || \tag{6}$$

$$d(t_i) = |(x,y,z) - P_{\text{palm}}(t_i)| \tag{7}$$

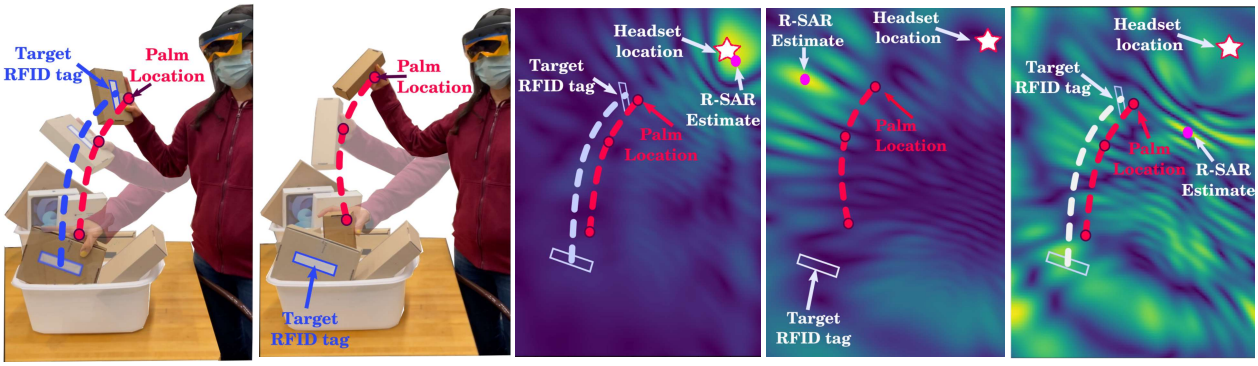
$$(x_h, y_h, z_h) = \max_{x,y,z} P(x,y,z) \tag{8}$$

where N_v is the number of measurements, t_i is the time of i$^{\text{th}}$ measurement, $d(t_i)$ represents the distance at time t_i from the (x,y,z) position to the user's palm location, $P_{\text{palm}}(t_i)$. X-AR obtains $P_{palm}(t_i)$ through vision based hand tracking. The SAR estimated headset location, (x_h, y_h, z_h), is the position that emanated the maximum power. Remember that when the user has the target item in their hand, $P_{\text{palm}}(t_i)$ is similar to the target RFID location at time t_i.

Fig. 4c shows the result of applying SAR to localize the headset in the form of a 2D heatmap from a side view. For simplicity, the result of antenna array projections is sliced in the plane that coincides with the real-world plane containing the user's body and the RFID-tagged item. In this heatmap, yellow indicates higher probability of the headset location, while navy blue indicates low probability. As the figure shows, the location of highest power (the pink dot) is very close to the actual location of the headset antenna (the white star), indicating that the headset has been accurately localized.

(b) Scenario where the User Picks an Incorrect Item. Next, consider a scenario where the user picks up an incorrect item, as shown in Fig. 4b. Here, the user's palm location (P_{palm}) changes as the user's hand moves, but the target RFID tag location (P_{tag}) does not change. In this case, when X-AR uses the user's palm location to estimate the tag location for the SAR, it will fail to accurately locate the AR conformal antenna location.

Figure 4d shows the result of applying SAR in this scenario. Notice how the heatmap displays multiple high probability regions that are far from the actual headset location. In this case, the highest probability location (depicted by the pink dot) which corresponds to the SAR-based estimate of AR conformal antenna's location is far from the actual location of the headset antenna (depicted by the white star). Thus, the SAR based headset localization fails because of large error. Since the headset knows its actual location (using the self-tracking via visual-inertial odometry as described in §4), it can determine that the reverse localization has failed, and use this information to determine that the target RFID tag is not



(a) Target is in hand (b) Target is not in hand (c) In-hand heatmap (d) Not in-hand heatmap (e) Without compensation

Figure 4: **RF-Visual In-Hand Verification.**(a) The RFID trajectory (blue dashed line) is similar to the palm trajectory (red dashed line) when it is in-hand. (b) The RFID's location (blue rectangle) differs significantly from palms trajectory(red line) when not in-hand. (c) When the tag is in hand, RF-Visual R-SAR accurately estimates the headset location (pink dot) relative to the actual headset location (white star). (d) The R-SAR estimation of the headset location (pink dot) is not accurate when the tag is not in hand. (e) Without compensating for natural head movement, RF-Visual R-SAR cannot locate the headset accurately even when the target RFID is in the user's hand.

in the user's hand.

The criteria for declaring that the target tag is in the user's hand is that the headset localization error should be within an acceptable range and can be formulated as follows:

$$||(x_h, y_h, z_h) - (x_G, y_G, z_G)|| < \tau \quad (9)$$

where (x_G, y_G, z_G) is the headset's location based on built-in odometry, and τ is threshold for localization error.[6]

5.2 Compensating for Natural Tilts

Our above description assumes that the user's head is perfectly still as they are picking an item. In practice, a user's head naturally tilts during picking, and its important to compensate for these tilts in the reverse SAR localization.[7]As a result of head movement throughout the retrieval process, the distance of user's palm to headset's initial location can be different from the actual distance from the user's palm to the headset's antenna location.

In Fig. 4a, we had shown the result of SAR after compensation. For comparison, Fig. 4e shows the result of applying SAR without compensating for the user's natural head movements. Multiple high probability regions are visible in the heatmap showing that if the natural head movements are not accounted for, the SAR estimated head location may have a large error and the item in the user's hand may be incorrectly classified.

To address this issue, X-AR tracks these natural head movements through the visual-inertial odometry and compensates for them in the RF-Visual SAR formulation. Specifically, X-AR translates the palm position from current headset coordinate to the initial headset coordinate. This can be formulated by replacing $d(t_i)$ in Eq. 6 with $\hat{d}(t_i)$ as follows:

$$\hat{d}(t_i) = |(x, y, z) - (P_{palm}(t_i) - [P_{head}(0) - P_{head}(t_i)])|$$

where $P_{head}(t_i)$ is the visual-inertial odometry-based head location at time t_i. In this new formulation, $\hat{d}(t_i)$ represents the compensated distance from head's initial position to the palm location at time t_i. The headset's estimated initial location, $P_{head}(0)$, is the same as (x_G, y_G, z_G) in Eq. 9. X-AR uses the same criteria as Eq.9 for the headset's initial location to determine if the target item is accurately retrieved by the user. In the system evaluation, we demonstrate how much this compensation is critical for RF-Visual Verification. We also note:

- X-AR can also use the camera visual data to determine if and when the user grasps an item by tracking her hands and fingers. It can use this information to trigger the RF-Visual verification module.
- The retrieval process often includes grasping and removing items to declutter the surroundings of the target item before the user actually grasp the target item. As a result, X-AR uses the latest received N_v RFID measurements[8] at each point of time for the RF-Visual verification. When the latest N_v satisfy the Eq.9's criteria, X-AR notifies the user that the target item is in her hand by showing text stating that target item is retrieved.

6 Implementation & Evaluation

Physical Setup: We implemented X-AR on a Microsoft Hololens 2. We mounted our custom conformal antenna on the front visor of the AR headset and connected it to two Nuand BladeRF 2.0 Micrsoftware radios [47]. Our device was tested using standard off-the-shelf UHF RFID tags [4](3-5 cent each). We tagged common items such as office supplies or clothes and placed them in boxes of different arrangements.

RFID Reader: To obtain wideband RFID channel measurements for localization, we implemented the EPC Gen 2 protocol [31] on a wideband RFID reader design similar to [17]. In order to transmit and receive signals from a single antenna, we introduced a CS-0.900 circulator to the reader. To cancel self-interference and extend

[6]In our implementation, τ is 0.3m. Note that the length of the AR-conformal antenna is 0.165m. We experimented with different τ's and found this achieves a good balance between precision and recall.

[7]Note that these tilts remain too subtle to perform SAR on the head movement itself, but are sufficiently large to make reverse SAR inaccurate if they are not accounted for.

[8]In our implementation, N_v is 35.

the range, we implemented over-the-wire nulling through the BladeRF's MIMO capability[47] and a ZAPD-2-21-3W-S+ 2-Way Pass DC Splitter. We connected the reader to a Raspberry Pi to collect and process RFID measurements from the software defined radios.

Software: We implemented the processing described in §4 and §5 on an edge server running Ubuntu 20.04 on an Intel(R) Core(TM) i9-10900X CPU @ 3.70GHz. The code is developed in Python and C++ and uses ROS [50] to enable multicore processing. We developed our own application for the Hololens to stream device transforms and tracked hand locations to the edge server via TCP protocol and present the designed UI to the user. The application was developed in C# in Unity3D [56] and Visual Studio IDE [43]. On the Rasberry Pi, we implemented code in Python to stream the processed RFID channel estimates to the edge server.

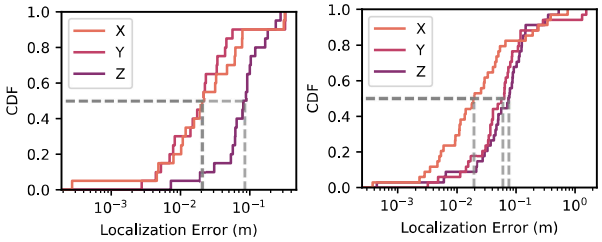
Evaluation Environment: We evaluated X-AR in multipath-rich indoor settings that mimic warehouses, retail stockrooms, and dark stores, which are cluttered with boxes. Fig. 3c shows a sample evaluation environment. Across experimental trials, we changed the arrangement of stacked boxes, moving them near metal shelving and/or wooden bench tops. Since our evaluation setups were created in a lab, they were also surrounded by furniture including chairs, desks, and computers. These environments also had typical wireless interference from various technologies, as well as multipath interference from building occupants who walked around the environment while going about their daily activities. Across our experimental trials, a user wears the X-AR headset and walks around to find and pick up an RFID-tagged target item. To evaluate localization with various human trajectories, we asked users to walk in several different patterns. These patterns included walking towards the target object, in a diagonal path approaching the target, and in 2D "L" or "V" shaped trajectories with respect to the target. We tested objects of different sizes/shapes across both LOS and NLOS settings. In LOS, the tagged object was not occluded from the AR headset's field of view. For NLOS settings, the RFID-tagged target was hidden inside a box or behind clutter. Across trials, we varied the target RFID-tagged object's location to cover various potential scenarios.

Baselines. We implemented 2 state-of-the-art baselines:

SAR baseline: Our first baseline is SAR-based localization algorithm (similar to [58, 65]). In this baseline, we used the AR-based VIO similar to X-AR to obtain antenna locations. However, we limited the localization to only frequencies within the UHF ISM band (around 22 MHz), and did not implement X-AR's weighting optimizations as described in §4.3.

Time-of-Flight baseline: Our second baseline implements state-of-the-art time-of-flight(ToF) estimation using wideband RFID measurements (similar to [40, 17]). For this baseline, we selected 6 measurements from the user's trajectory (similar to [17], spaced evenly in time), computed the ToF-based distance estimates (using the algorithm from [40]), then performed robust trilateration to compute the final 3D location (as in [17]).

For fairness of comparison, in both baselines, we applied the same initial SNR filter as X-AR to remove low-confidence measurements.



(a) Line-of-sight (LOS) (b) Non-line-of-sight (NLOS)

Figure 5: **3D Localization Accuracy.** CDF plots of X-AR's RF-visual SAR localization accuracy in the x/y/z dimensions for LOS and NLOS.

Ground Truth: To measure the localization accuracy of our system, we used the AR headset's built-in spatial awareness to determine the origin of the coordinate system in each trial. In each experimental trial, we aligned the tag's location with the Hololens's origin. This was done by manually moving the RFID tag to the origin (displayed as a hologram by the AR advice) so that it aligns with the Hololens origin at the beginning of each trial. Subsequently, the localization accuracy was computed as the difference between X-AR's RFID tag estimated location and the Hololens' origin. This was repeated for each experimental trial.

7 Results

We ran 234 trials to evaluate the performance of X-AR.

7.1 3D Localization Accuracy

We first evaluated the accuracy of our system in localizing target RFID-tagged items in the environment. We define the localization error to be the euclidean distance between the system's estimated location and the ground-truth location. We ran 54 experimental trials to measure the performance of RFID localization. In each trial, the user walked in a different motion pattern and X-AR automatically localized the target item via RF-visual SAR as described in §4.

Fig. 5 plots the CDF of the localization error across the experimental trials in both line-of-sight and non-line-of-sight scenarios. We plot the localization error along the x(orange), y(pink), and z(purple) dimensions. We note:

- In LOS settings, the median errors are 2.1 cm, 2.1 cm, and 8.4 cm along the x, y, and z dimensions, respectively. In NLOS settings, the median errors are 1.9 cm, 6 cm, and 7.7 cm along the x, y, and z dimensions, respectively. These results demonstrate that X-AR can achieve

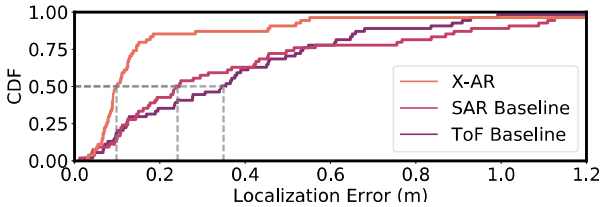


Figure 6: **Comparison to Baseline Localization Accuracy.** CDF plots of L2-norm error for X-AR(orange), *SAR*(pink), and *ToF*(purple).

centimeter-level localization accuracy in each dimension while opportunistically leveraging human motion that is not known a priori or directed in a particular way.

• The median L2 norm of localization error for the LOS and NLOS scenario are 9.6 cm and 10.6 cm. Therefore, there is no significant difference between localization error for NLOS and LOS, showing that X-AR's is able to augment the AR device with perception capabilities for both LOS and NLOS conditions.

• The localization accuracy along the x-axis is generally better than along y and z (especially in the NLOS scenario). This is because in our experimental setup, the object is located on a shelf against a wall. The user walks toward the shelf but not past it, meaning the RFID measurements are only on one side of the RFID in the y direction. On the other hand, the user walks parallel to the shelf and measurements are taken on both sides of the RFID along the x-axis leading to a better accuracy in the x direction than in the y direction. Note that the aperture in z direction (vertical direction) is very small since the user's head does not move vertically.

Baseline Comparison: We compare the performance of our system to the two baselines described in §6. We used the same experimental trials for X-AR and the baselines.

Fig. 6 plots the CDF of the total localization error for X-AR (orange), *SAR Baseline* (pink), and *Time-of-Flight Baseline* (purple). For simplicity, we show the L2-norm distance error (rather than the error along each of the x/y/z dimensions). We make the following remarks:

• For X-AR, the median and 90^{th} percentile localization errors are 9.8 cm and 45 cm, respectively. These results are in-line with those reported above (as L2-norm in 3D).

• For *SAR Baseline*, the median and 90^{th} percentile localization errors are 24.8 cm and 99.1 cm, respectively. This shows that by leveraging our system's custom wideband antenna and wideband RF-visual SAR techniques, X-AR can achieve over $2\times$ performance improvement in both the median, and 90^{th} percentile over a system that is limited to the UHF ISM band, thus demonstrating the value of our customized wideband conformal antenna design and RF-visual SAR localization scheme.

• The *Time-of-Flight* baseline has a median and 90^{th} percentile localization errors of 34.9 cm and 78.8 cm. This shows that X-AR has an improvement of over $3\times$ in the median and almost $2\times$ in the 90^{th} percentile. We note that the baseline's performance is worse than that reported in prior work [40, 17]. This is because that prior work had control over the aperture of measurements (i.e., through physical antenna placement or controlling robotic motion). In contrast, when applying these techniques to an AR system with natural human motion, the aperture cannot be optimized and the resulting accuracy is poor. This demonstrates the benefit of our AR-based SAR techniques when utilizing natural human motion.

Impact of Walking Pattern: Next, we investigated the impact of different walking patterns on X-AR's localization accuracy. Recall that we asked users to walk in different patterns: vertically toward the tag's plane, diagonally toward the tag, as well as L-shaped and V-shaped trajectories. To understand the impact of different motion patterns on localization accuracy, we measured the 10^{th}, median and 90^{th} percentile for each of these patterns and reported them in Table 1.We note:

	Vertical	**Diagonal**	**L-shape**	**V-shape**
10^{th} percentile	5.7 cm	3.8 cm	7.9 cm	6.3 cm
50^{th} percentile	10.8 cm	12.5 cm	9.8 cm	8.4 cm
90^{th} percentile	47.7 cm	51.0 cm	14.9 cm	13.3 cm

Table 1: **Trajectory Impact.** Location error for different trajectories.

• All walking patterns have a similar median localization error, between 8.4 to 12.5 cm. This shows that X-AR works well in different motion patterns and is generally robust to different trajectories. It also suggests that X-AR does not need to constrain the user to a pre-defined 2D trajectory to achieve good localization performance.

• Interestingly, we noticed that 90^{th} percentile accuracy is markedly different across these motion patterns. In particular, while the L-shaped and V-shaped patterns have a 90^{th} percentile around 15 cm, this error increases to around 50 cm for linear motion patterns (diagonal & vertical). This is likely due to the differences in spatial diversity and aperture variability across these motion patterns. In particular, L-shaped and-V-shaped trajectories involve independent mobility in two dimensions, while the diagonal and vertical trajectories involve mostly linear motion patterns, giving less overall aperture.

Impact of Aperture. We investigated the impact of the trajectory's aperture size on localization accuracy through a micro-benchmark evaluation. To do this, rather than providing the RF-visual SAR algorithm with the entire trajectory for localization, we trimmed the trajectory of each trial to a certain maximum aperture. For example, to evaluate an aperture of 0.6 m, we only provided the first 0.6 m of the user's trajectory to the localization algorithm.[9] We repeated the same process for apertures of different lengths, and computed the localization accuracy for each of them across all the experimental trials.

[9]The aperture of a trajectory is defined by the diagonal of the bounding box encompassing the measurements in that trajectory.

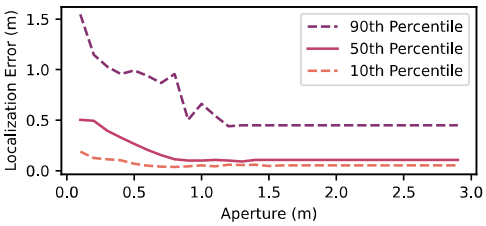


Figure 7: Impact of Aperture. Localization error vs the aperture of the user's trajectory. The plots show the 10^{th}, 50^{th}, and 90^{th} percentiles.

Fig. 7 plots the L2-norm localization error as a function of the aperture. The plot shows the 10$^{\text{th}}$ (orange), 50$^{\text{th}}$ (pink), and 90$^{\text{th}}$ (purple) percentile errors across all experimental trials. We make the following remarks:

• When limiting the aperture to 0.1 m, the 50$^{\text{th}}$ and 90$^{\text{th}}$ percentile errors are 0.5 m and 1.5 m respectively. As the aperture increases to 0.8 m, these errors drop to 0.11 m and 0.96 m. This shows that for small apertures, X-AR's performance greatly improves as the user walks further.

• After the aperture reaches 0.8 m, the median errors become relatively constant. For example, expanding the aperture to 1.2 m only decreases the median error by 2 cm. This shows that increasing aperture after 0.8 m does not improve the median localization accuracy.

• The 90$^{\text{th}}$ percentile continues to improve as the aperture is increased from 0.8 m to 1.2 m, dropping by 0.43m. This shows that larger apertures improve reliability.

• X-AR visualizes the RFID tag on the AR device once the user's walking trajectory allows for adequate RF measurement aperture, such that X-AR is confident about the RFID tag's location, as described in §4.3. As a result, the time it takes X-AR to find the requested item is dependent on the user's walking speed and trajectory.

Impact of SNR-based Weighting Function. We investigated how weighting measurements based on the received SNR impacts the localization accuracy of X-AR. We processed the experiments with SNR-based weighting (Eq.4) and with uniform weighting (Eq.1) and calculated the L2 norm of RFID localization error. Our results showed that the SNR-based weighting improves the robustness of the system, specifically in the 90$^{\text{th}}$ percentile localization accuracy. While the uniform weighting and SNR-based weighting have a similar median errors (around 10 cm), the 90$^{\text{th}}$ percentile in our SNR-based weighting approach is 45 cm, while the uniform weighting approach has 71 cm error.

7.2 In-Hand Verification

Next, we evaluated X-AR's ability to successfully determine if the correct RFID tagged object was retrieved by the user. We conducted 180 trials in total. In each trial, the user grasped a tagged or non-tagged item and moved their hand in a pick and place motion. In each trial, X-AR predicted whether the RFID tag was in the user's hand or not, (i.e., the correct item being picked or not). We define a successful trial as one in which X-AR correctly determines whether or not the tag was in the user's hand.

	Precision	**Recall**	**F-score**
Extracting RFID-tagged item (LOS)	98%	100%	98.9%
(without compensation)	98%	98%	98%
Picking boxed RFID-tagged item (NLOS)	100%	85.1%	91.9%
(without compensation)	100%	74.3%	85%
Large Object (LOS+NLOS)	100%	87.5%	93%
Small Object (LOS+NLOS)	98%	93%	95.4%

Table 2: **In-hand Verification Accuracy.** The table reports the results for in-hand verification across different evaluation scenarios. The results are reported as percentages for precision, recall, and F-measure.

Table 2 reports the results for X-AR's in-hand verification algorithm. Here, *Precision* indicates the number of trials where the target item was correctly classified in-hand divided by the overall number of trials that systems classified the target item as in-hand. *Recall* indicates the number of trials where target item was correctly classified in-hand divided by the overall number of trials where the target RFID tagged item was actually in the user's hand. We make the following remarks:

• X-AR achieves a 98% precision rate, and 100% recall rate. These values demonstrate that when the user retrieves an item, X-AR can reliably and correctly predict whether the target item has been picked up.

• The system has 98% precision rate, which indicates 2% of the trials when the system registered it as a potential retrieval, the user has picked up an incorrect item (e.g., non-tagged item, or potentially an item that is tagged with a different RFID). We suspect the reason for some trials being mistakenly registered as positive arises from multipath. Specifically, even though the user did not pick up the target RFID-tagged item in these trials, the wireless signal reflecting off the user's hand during motion creates an array of the multipath reflections. Such multipath arrays may have inadvertently allowed localizing the headset, resulting in false positives.

Picking Boxed RFID-tagged items (NLOS): We also evaluated whether X-AR can accurately verify when a user picks up an RFID-tagged item that remains *inside* a box during the picking process. While such scenarios are less likely in practice (e.g., in warehousing or retail), they may arise and serve to test the limit of our system in performing RF-visual verification of RFIDs in NLOS.

The results for this experiment are shown in the third row in Table 2. The results show that even though the recall rate drops, X-AR remains largely successful in performing the verification, achieving a precision and recall rate of 100% and 85.1%. This change in performance can be attributed to the fact that when target tags are not in line-of-sight (and are inside a box) their distance to the user's palm is markedly higher. This offset between the tag location and visually extracted palm location impacts the reverse SAR calculation. In the future, this may

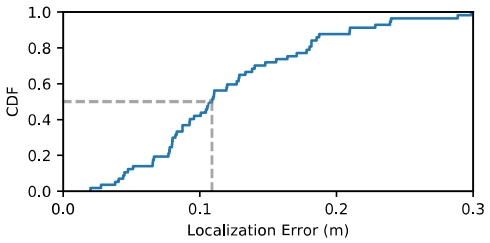


Figure 8: **CDF of Reverse SAR.** CDF plots of the headset's localization accuracy by applying R-SAR on the trajectory of the target item.

be compensated for by estimating the potential location of the RFID inside the box and/or investigating different features from the AR headset's built-in hand-tracking.

Impact of Motion Compensation: Recall from §5 that X-AR's RF-visual verification primitive compensates for head tilts in the reverse SAR localization process. To investigate the impact of such compensation, we processed the same experimental trials as above (for both LOS and NLOS scenarios) without performing motion compensation and reported the results in Table 2. The table shows that the recall rate drops for both LOS (from 100% to 98%) and NLOS scenarios (from 85.1% to 74.3%). This demonstrates that by accounting for head tilts, X-AR's accuracy in prediction markedly improves.

Impact of Target Object Size: Next, we investigated the impact of target object size on the verification accuracy. Specifically, we divided our experimental trials between objects of smaller and larger sizes. The object size was determined by their longest dimension, using 10 cm as a divider between objects that we referred to as small and large - i.e, objects with largest dimension $<$ 10 cm classified as small, and those with a dimension $>$10 cm classified as large. In practice, choosing a different threshold does not make a significant difference, as the primary goal of this experiment was to micro-benchmark the impact of object size on verification accuracy.

The last two rows of Table 2 show the results comparing the accuracy for different object sizes, covering both LOS and NLOS scenarios. The table shows that smaller items have higher recall rate (93%) than larger items (87.5%). This can be attributed to the fact that for larger items, there is a larger offset between the tag location and palm location. Specifically, recall from §5 that X-AR approximates an RFID's location as the user's palm location (extracted from the AR-headset's hand-tracking module). As a result, the smaller the object is, the more accurate this approximation is, leading to higher accuracy for smaller objects. In the future, it would be interesting to explore mechanisms that adapt the threshold to the object size, or alternatively leverage the RFID location inside the box and apply a transformation to the user's palm to compensate for these differences and achieve higher accuracy for larger objects.

Reverse SAR Localization Accuracy. Our final result looks into the reverse SAR localization accuracy. Recall from §5 that X-AR's verification component relies on the ability to correctly localize the headset (specifically the AR-conformal antenna) by applying SAR on the mobile tag. To investigate this primitive, we evaluated the method's ability to correctly locate the position of the user's head. Here, we defined the ground truth of the location of the user's head to be the visual-inertial odometry-based location and estimated the error by calculating the euclidean distance between the ground truth and X-AR's predicted location. We computed the localization error for all scenarios where the user picks up an RFID tagged item.[10] Here, we included experimental trials from LOS scenarios described above.

The CDF of the localization error is plotted in Fig. 8. The figure shows that the method allows localizing the headset using SAR with a median accuracy of 11 cm and a 90^{th} percentile accuracy of 19.6 cm. These results show that even with simple pick and place movements, X-AR can accurately locate a user's head using reverse SAR techniques, while compensating for head movements. This high localization accuracy is why the system can accurately verify picking RFID-tagged items.

8 Related Work

RFID Localization. RFID localization is a well-studied problem in the networking community with researchers exploring various techniques including received signal strength (RSS) [20, 46], angle of arrival (AOA) [13, 36, 71], and wide-band sensing [40, 16, 39]. The closest to X-AR is past work that leverages motion for RFID localization, which falls in two main categories. The first places an antenna on robots that move along *predefined* trajectories and leverage these trajectories to localize the tags [57, 29, 45, 53, 17, 16, 44, 70, 14]. Our system does not require users to move along specific (unnatural/robotic) trajectories, yet can still localize accurately by leveraging natural movement. The second category tracks RFIDs that are already in motion, e.g., for gesture recognition [59, 65, 21, 66]. Our work differs from these in that it can also localize stationary tags by using an AR mounted antenna. Thus, our work is the first to bring fine-grained RFID localization to AR headsets, addressing challenges that span antenna design, natural human mobility, and various localization artifacts.

Augmented Reality. Augmented Reality (AR) refers to systems that overlay a virtual world on top of the physical world to enable new experiences and interactions [12, 35]. Most prior work that leverages RF in AR systems does not involve headsets altogether and simply visualizes tagged items on a screen or a smartphone [49, 37, 62, 63]. This includes past work that

[10] Note that the error for non-tagged items is much higher since the formulation does not hold. Empirically, the median localization error for those scenarios is over a meter.

deploys an RFID localization infrastructure in the environment and uses it to localize tags and visualize their locations on a screen [49, 37]. It also includes robotic systems mounted with RFID readers and cameras to scan the environment and send the result for visualization on a screen[62, 63]. X-AR builds on this area and brings RFID localization to AR headsets, addressing the associated challenges in antenna design, human mobility, and headset-based localization. X-AR is also related to past work that involves users wearing RFID readers on their hands or in their backpacks to detect objects in the environment [69, 54] or self-localize [64, 37]. X-AR differs from these systems in directly integrating the localization and sensor fusion into the headset itself, resulting in a more natural and seamless AR experience.

Conformal Antennas. Antenna design is a mature field that targets satisfying multiple requirements such as compactness, robustness to flexing, radiation pattern, and weight. The closest to our work are Bluetooth headset antennas desired to radiate outwards while close to a human head, and designed to be mounted around the ear or on glasses handles [23, 22, 34, 33]. These past designs differ from our work in their bandwidth requirements, desired radiation pattern, and form factor. Other wearable antennas were designed for safety helmets [24] or smart glasses [60], but were either too bulky and obstructive or lacking the wideband operation desired for wideband RFID localization. Loop antennas are simple, and do not require a ground plane, but are inherently narrowband. Past techniques such as tapering and slots help improve their bandwidth, but none of the existing wideband loop antenna designs can simultaneously operate at the desired frequency range while matching the dimensions of the visor [68, 61, 67]. Our proposed design takes advantage of wideband antenna techniques to deliver a broadband, compact, and conformal loop antenna that perfectly fits on the headset's visor without covering the cameras or blocking the user's view.

9 Discussion and Limitations

Antenna Placement: In principle, one could design alternate versions of our X-AR system by placing the antenna on top of the headset, on the user's shoulder, or even in the user's hand. However, these alternative approaches are suboptimal in most scenarios compared to X-AR's design. For example, in warehouses, pickers are more efficient when they can use both hands (rather than carring an antenna with one hand all the time). Similarly, mounting large and heavy antennas on their shoulders or heads would create undesirable additional weight which may impact their balance. That said, it is possible that such alternate implementations may be useful in certain use-cases and can be explored as the research evolves.

Transmission Power: X-AR's transmit power is lower than that of existing wrist-worn RFID readers [10] since bladeRF software-defined radios transmit less than 8dBm. This is also lower than the power transmitted by Apple AirMax headphones, which use Bluetooth 5.0 technology and have a maximum transmission power of 20 dBm [1, 2]. In production systems, X-AR could leverage a deployed RFID reader infrastructure to power RFID tags in the environment, and an X-AR headset for wideband measurements for localization, UI, etc.

RFID reliability: Our implementation of X-AR inherits the typical limitations of RF/RFID signals. For example, it cannot detect or localize items inside closed metallic boxes. However, it can still read RFIDs on metal or liquid bottles if proper tags are used. Moreover, due to its wideband sensing capabilities, it can work in multipath-rich environments, including those with metal shelving, as demonstrated in our evaluation.

Form factor: As X-AR moves closer to commercial deployments, we envision that the entire RF sensing hardware can be integrated into the headset. In particular, while our proof-of-concept prototype was implemented using software radios and a Raspberry Pi, future versions may be designed in form factors similar to existing RFID reader chips (e.g., Lepton3 [18] that are around 1"x1"x0.1"), thus small enough to fit into AR headsets.

Range: The operation range of X-AR is approximately 3-4 meters which is similar to mobile (portable) handheld RFID readers on the market [3]. While this range is lower than stationary readers (which can reach around 10 m), that is primarily because stationary ones typically transmit much higher power. In contrast, handheld readers usually transmit lower power to conserve their battery life, and we envision the same would be desired for future readers integrated in headsets like X-AR.

10 Conclusion

The past few years have witnessed remarkable advances in augmented reality and its metaverse applications. Motivated by these advances, this paper brings a new sensing modality to AR systems through networked RF sensing, giving them the ability to perceive what used to be invisible to the human eye and to existing AR headsets. In doing so, the paper opens the door to more exciting capabilites and applications at the intersection of RF sensing and AR systems. As the research evolves it would be interesting to explore how various networked wireless sensing modalities and sensor fusion techniques - spanning RFID, WiFi, mmWave, and THz - can further augment augmented reality and open new possibilities in visualization and interaction.

Acknowledgments We thank the anonymous reviewers, our shepherd Dr. Behnaz Arzani, and the Signal Kinetics group for their help and feedback. We also thank Yuechen Wang for her help with UI design and implementation. This research is sponsored by NSF (Awards #1844280 and #2044711), the Sloan Research Fellowship, and MIT Media Lab.