TagRecon: Fine-Grained 3D Reconstruction of Multiple Tagged Packages via RFID Systems

ZIANG WANG, CHUNHUI DUAN*, JIAWEI XUE, FAN LI, and QIHUA FENG, School of Computer Science and Technology, Beijing Institute of Technology, China
YINAN ZHU, Department of Computer Science and Engineering, Hong Kong University of Science and Technology, China
ZIYANG ZHOU, School of Cyberspace Science and Technology, Beijing Institute of Technology, China

To meet the new requirements of Industry 4.0, the logistics field has introduced 3D reconstruction technology. Computer vision-based solutions face challenges like bad lighting conditions and line-of-sight constraints. Meanwhile, the widespread adoption of RFID tags in supply chains offers an opportunity to enhance current reconstruction methods. In this paper, we propose TagRecon, a fine-grained multi-object 3D reconstruction scheme utilizing well-deployed RFIDs. Specifically, TagRecon transforms the task of reconstruction into a problem of estimating 3D bounding boxes for tagged packages. By placing dual anchor tags on each target package, TagRecon enables accurate inference of the package's translation and rotation using RFID-based localization and orientation sensing. Our scheme introduces a novel method to estimate rotations and translations for tagged packages, utilizing the known geometric relationship of anchor tags. Besides, to achieve simultaneous reconstruction of multiple packages, we manage to match tags from various packages through the correlation between anchor tag pairs. As far as we know, this is the first RFID-based solution that can simultaneously realize 3D translation and rotation estimation of multiple objects to a fine granularity. Experiments validate TagRecon achieves a 28.0 cm translation error and 6.8°, 6.0°, and 7.5° rotation errors for roll, pitch, and yaw angles on average.

CCS Concepts: • **Human-centered computing** → Ubiquitous and mobile computing; • **Networks** → **Location based services**; **Mobile networks**.

Additional Key Words and Phrases: RFID, multi-object 3D reconstruction, 3D bounding box estimation

ACM Reference Format:
Ziang Wang, Chunhui Duan, Jiawei Xue, Fan Li, Qihua Feng, Yinan Zhu, and Ziyang Zhou. 2024. TagRecon: Fine-Grained 3D Reconstruction of Multiple Tagged Packages via RFID Systems. *ACM Trans. Sensor Netw.* 0, 0, Article 0 (2024), 24 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

*Corresponding author

Authors' addresses: Ziang Wang, ziangwang@bit.edu.com; Chunhui Duan, duanch@bit.edu.cn; Jiawei Xue, xuejw@bit.edu.cn; Fan Li, fli@bit.edu.cn; Qihua Feng, fengqh@bit.edu.cn, School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China; Yinan Zhu, yzhudf@cse.ust.hk, Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, China; Ziyang Zhou, zhouzy@bit.edu.cn, School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing, China.

ACM 1550-4859/2024/0-ART0
https://doi.org/10.1145/nnnnnnn.nnnnnnn

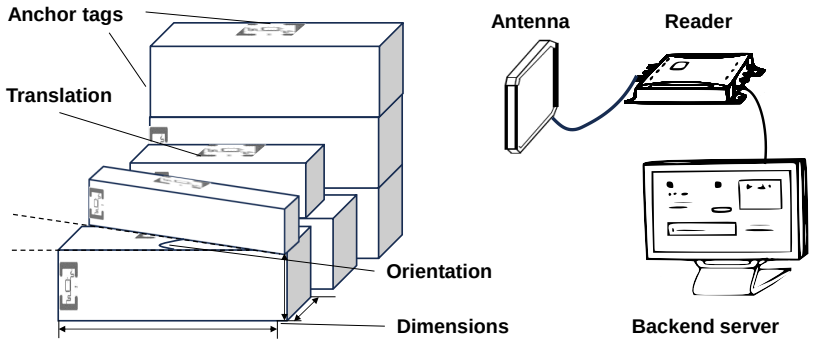


Fig. 1. Illustration of the TagRecon system.

1 INTRODUCTION

1.1 Background and Motivation

The pursuit of improved quality, efficiency, and safety in the era of Industry 4.0 brings new requirements for warehouse management and logistics. For example, in addition to locating targets, robots also need to appropriately interact with objects (*e.g.* packages with factory-produced goods). Understanding the precise orientation of items is crucial for these robots to pick, rotate, and place packages correctly, reducing the risk of damage and improving efficiency. As increasingly popular technologies in logistics, Augmented Reality (AR) and Virtual Reality (VR) [4, 5] also require a comprehensive understanding of an item's location and orientation to provide realistic and interactive simulations, critical for effective training and operational planning. To satisfy the above requirements, Three-Dimensional (3D) multi-object reconstruction has been introduced for monitoring the accurate placement of packages. This reconstruction process relies on capturing key parameters with high precision, including the dimensions, translations, and rotations of objects. Precise 9D pose information can not only elevate the quality of visualization in AR and VR, but also enhance efficiency and promote safety in automated warehouses and logistics.

Early computer vision (CV) methods, as detailed in [18] and [11], utilized multiple depth cameras placed at various locations to achieve object reconstruction. This approach resulted in significant deployment and hardware costs. However, recent advancements in CV technology have made it possible to perform 3D reconstruction of single or multiple objects within the field of view using just one camera, thereby reducing these costs. Research [39] indicates that state-of-the-art single-view object reconstruction methods [10, 13, 26, 34, 35] mainly perform recognition and retrieval rather than actual reconstruction. These visual recognition approaches have difficulty distinguishing between objects with similar shapes. Furthermore, recognition-dependent visual methods often lack an explicit localization step, resulting in inaccurate object positioning. These visual-based approaches face common challenges, such as dealing with occlusions, poor lighting conditions, and failing to achieve instance-level recognition. These limitations significantly affect the accuracy and reliability of visual 3D reconstruction solutions.

Radio Frequency Identification (RFID) technology has grown rapidly over the past few years and been widely used in all stages of the logistics and warehousing field. In recent years, the cost of RFID tags has dropped significantly, from several dollars to just a few cents [2]. Many modern warehouses now use passive RFID tags to identify items and packages [3]. As a key

enabler of automatic identification technology, RFID offers several benefits, including non-line-of-sight identification and the ability to read fast-moving and multiple objects simultaneously. These advantages present new opportunities to overcome the limitations of visual 3D multi-object reconstruction. The RFID-based method, RF-3DScan [7], allows 3D reconstruction by comparing angular profiles of different reference tags attached to packages to determine the orientation of packages and their stacking order. However, a limitation of RF-3DScan is that its antenna must move in a straight line at a constant speed and cannot be fixed in one position. Additionally, each package requires the deployment of at least four reference tags. The elevated deployment and hardware costs have restricted the widespread application of the system.

1.2 Our Scheme

To overcome the above limitations, we propose a fine-grained multi-object 3D reconstruction scheme named TagRecon. Unlike previous attempts, we aim to provide a more easy-to-deploy system using purely Commercial Off-The-Shelf (COTS) equipment. As demonstrated in Fig. 1, the TagRecon system consists of three main components of hardware devices: (i) an RFID reader, (ii) four circularly polarized antennas, which are connected to the reader, and (iii) a back-end server. In terms of functionality, TagRecon supports simultaneous reconstruction of multiple tagged packages, making it suitable for real-world supply scenarios where packages might be densely stacked in various arrangements.

To achieve this goal, TagRecon translates the reconstruction task to a 3D bounding box estimation problem for tagged packages. Specifically, we attach a pair of anchor tags to each package: one on the top and the other on the side. By retrieving the Electronic Product Code (EPC) of RFID tags from the database and estimating the locations and orientations of these anchor tags, TagRecon can determine three key parameters of each package: its dimensions, translation, and rotation. To further enable simultaneous reconstruction of multiple packages, TagRecon tries to match anchor tags from different packages by leveraging the correlation between anchor tag pairs.

1.3 Challenges and Solutions

We need to solve the following three technical challenges when implementing our TagRecon system:

The first challenge is how to infer a package's fine-grained 3D orientation from the orientations of the anchor tags. Inspired by state-of-the-art literature in RFID-based sensing domain [46], it is feasible to accomplish tag localization and orientation sensing concurrently with high precision. However, given the 3D, large-volume nature of packages, as opposed to planar and small-sized RFID tags, it is insufficient to solely rely on a single tag for accurate package rotation estimation, particularly in complex real-world scenarios. To tackle this challenge, we propose to attach a pair of tags on each package to serve as reference tags and devise an effective rule for easy deployment of the dual tags. Then the estimation of package rotation can be reduced to the absolute orientation problem and a Kabsch-based algorithm is suggested to solve it.

The second challenge is how to precisely determine the translation of a package in 3D space. Now that the position of each tag is acquired, simply approximating the centroid of a package by the location of the attached tag is theoretically reasonable yet too coarse-grained and may produce ordering inaccuracies especially when packages are densely stacked in practice. As a result, we come up with a translations estimation method for tagged packages utilizing the prior knowledge of the dual anchor tags' geometric relationship (*i.e.*, the dual tags are deployed on the top and side surfaces of a package, perpendicular to each other).

The third challenge is how to perfectly match the dual-tag pair for each object in tag-dense situations. In practical supply chain management applications, there could be a considerable

number of packages, making it difficult to achieve accurate pairing of two anchor tags for the same package. In addition, packages are usually placed very close to each other in real environments. The accompanying multipath effect and object occlusion may cause the problem of missing tags, which further degrades the matching precision. To address the issue, we reduce this matching problem to the stable roommates problem and present an improved solution based on the Gale-Shapley algorithm to figure out a most perfect match among all the tags effectively.

1.4 Contributions

This paper presents a comprehensive study of RFID-based multi-object 3D reconstruction. In particular, it makes the following contributions.

- We propose an RFID-based scheme for the 3D reconstruction of tagged packages to enhance efficiency and promote safety within supply chain management field. This scheme operates by sensing the locations and orientations of the RFID tags attached to packages.
- We present a 3D bounding box estimation method for accurately capturing the 9D pose of tagged objects. To the best of our knowledge, this is the first work that achieves precise 3D translations, rotations, and dimensions estimation for multiple objects simultaneously through RFID technology.
- We introduce a matching algorithm based on the stable roommates problem that pairs anchor tags for the same package in scenarios where numerous packages exist. This algorithm successfully handles the negative impact of the missing tag phenomenon caused by indoor multipath effect and object occlusion.
- We implement a prototype system for TagRecon using COTS RFID devices and conduct extensive experiments to evaluate its performance. The system demonstrates average errors of 6.8°, 6.0°, and 7.5° for the roll, pitch, and yaw angles in rotation and 28.0 cm in translation, proving sufficient for most practical applications.

Roadmap. The remainder of this paper is structured as follows. We overview the main design of TagRecon in Section 3, and summarize related studies in Section 2. Section 4 delves into the technical details of tag localization and orientation sensing. A dual-tag based solution for 3D bounding box estimation of tagged packages is proposed in Section 5 to enhance TagRecon's granularity. Section 6 describes the implementation and evaluation of our system. Section 7 discusses the limitations of our system. Finally, we conclude this paper in Section 8.

2 RELATED WORK

We briefly review the literature related to our work in this section.

2.1 Multi-Object Reconstruction

Early CV-based solutions [11, 18] primarily rely on depth cameras to perform 3D reconstruction of multiple objects. A single depth camera can not fully capture all angles of a specific object, leading to "blind spots" in the reconstruction process. To address this issue, KinectFusion [18] employ multiple depth cameras positioned at different locations to achieve multi-view 3D reconstruction of the same object, significantly enhancing the completeness and accuracy of the reconstruction. Another work [11] involves moving the depth camera on mobile devices to implement 3D reconstruction of objects. 3D reconstruction of objects. This allows the camera to move in space, capturing multiple views of the object, thereby reducing blind spots and optimizing the quality of the reconstruction. However, these methods introduce significant hardware and deployment costs.

In recent years, single-view based multi-object reconstruction has gradually become the focus of research. A landmark work in this field is Total3DUnderstanding [34], which bridges object-centered reconstruction with 3D scene understanding, enabling joint learning of room layouts, camera poses, and object bounding boxes from a single image. This method significantly enhances the ability to reconstruct complex scenes from a single image. Another breakthrough research from Google's team develops the CoReNet [35], which jointly reconstructs all objects without relying on traditional detection stages. This approach allows for spatial exclusion constraints, achieving globally consistent scene reconstruction and improving overall reconstruction accuracy and reliability. [39] highlights that state-of-the-art single-view object reconstruction methods [10, 13, 26, 34, 35] mainly perform recognition rather than reconstruction. Visual recognition methods have difficulty in distinguishing objects of similar shapes and achieving instance-level recognition[38]. Not only that, visual methods that rely on recognition usually have low accuracy in object localization due to the absence of an explicit location step. Most importantly, they are more easily affected by bad lighting conditions and cannot deal with occlusions in the line of sight. These shortcomings limit the completeness and accuracy of CV-based reconstruction.

RFID technology, with its characteristics of low cost, passive tags, and non-line-of-sight identification, presents new opportunities to address challenges such as visual occlusion and sensitivity to lighting conditions in traditional visual methods. Specifically, RF-3DScan [7], which is the only known method that enables 3D reconstruction of multiple objects using RFID tags, achieves this by moving an antenna along straight lines within a confined 2-dimensional area and comparing angle profiles of distinct reference tags attached to packages. This innovative approach allows it to determine not only the package's orientation but also its stacking order in a multi-object environment. This capability distinguishes RF-3DScan from other RFID-based studies, which primarily focus on single-object location and tracking. However, it comes with the following constraints: the antenna must traverse in a straight line at a consistent speed; a minimum of four tags is required per package; and all tags must adhere to specific orientation and positioning rules. Such requirements increase the deployment cost. Additionally, RF-3DScan represents the package's rotation and position using a single angle and relative position only, without a detailed pose estimation. Such an incomplete representation greatly reduces the accuracy of the reconstruction. A more detailed comparison and discussion of our work and RF-3DScan be found at Section 6.4.

2.2 RFID-Based Location and Orientation Sensing

Location and orientation sensing are vital areas of research in the RFID domain and have been extensively studied.

Location Sensing: The localization problem can be mainly categorized into absolute positioning [6, 9, 22, 24, 29, 30, 32, 33, 41, 43, 45] and relative positioning [7, 8, 36]. As a representative work in absolute positioning, Tagoram [45] presents three RF hologram-based localization algorithms to achieve instant tracking of the mobile tag with high precision. OTrack [36], which is a pioneering work in RFID-based relative positioning, proposes a protocol to enable continuous detection of the order of tags, which also allows determining the order of goods on the conveyor belts.

Orientation Sensing: By deploying RFID tags onto a specified object, the change in orientation of the tag can be inferred from the change in RF signal, thus indirectly deriving the change in orientation of the object [14, 19–21, 25, 27, 28, 40, 44, 46]. Tagyro [44] presents a novel orientation tracking system that utilizes an array of passive RFID tags to accurately monitor the 3D orientation of objects, overcoming challenges related to imperfect radiation patterns and antenna polarity through advanced phase sampling and recovery algorithms. 3D-OmniTrack [19] introduces a polarization-sensitive phase model that adeptly incorporates both the distance and the 3D orientation of an object for precise tracking within 3D space. RF-Prism [46] develops a versatile RFID sensing system

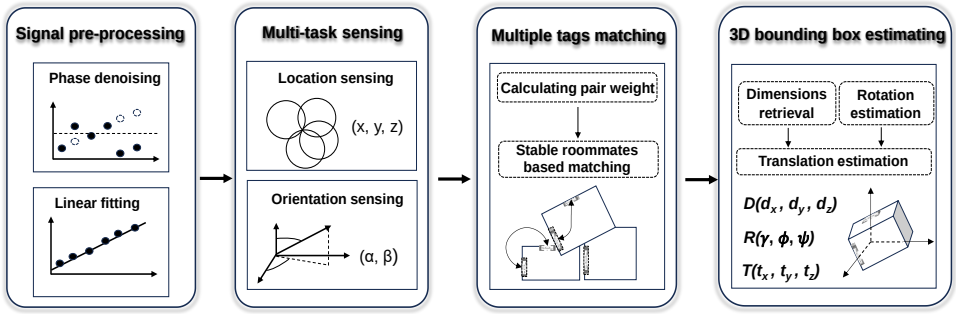


Fig. 2. Overview of the TagRecon system.

that can simultaneously infer multiple physical factors (*i.e.*, location, orientation, and material of targets) from the phase readings.

Our approach to estimate 3D package orientation and translation differs from previous methods in the RFID-based sensing domain, which relied on a single rotation angle to approximate object orientation, resulting in limited granularity and accuracy. We innovatively design a rotation estimation algorithm that leverages directional information of a pair of anchor tags to provide a more precise estimation of object's 3D orientation. Moreover, previous methods often use a single tag on one surface of the object to denote its position, which is not reliable for large objects *e.g.*, large-volume packages, due to the huge offset between the tag and the object's centroid. We introduce a novel translation estimation algorithm that minimizes such offset by employing dual anchor tags on different surfaces, significantly improving the accuracy of package positioning. These two estimation algorithms capture the subtle differences in the poses of the target object and the attached tags, enhancing the overall granularity and accuracy of the estimation.

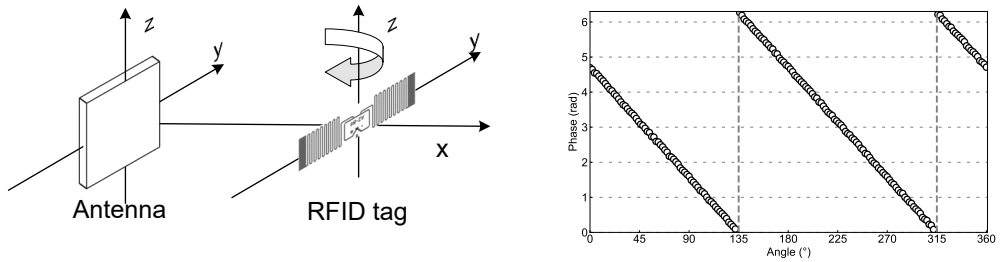
3 SYSTEM OVERVIEW OF TAGRECON

TagRecon is an RFID-based system designed for multi-object 3D reconstruction that accurately estimates the 3D bounding boxes of the tagged packages. This system deduces the dimensions, translation, and rotation of each package using only the phase readings of the RFID tags attached to it. TagRecon incorporates a single COTS RFID reader and multiple tags to identify the target objects. The reader is equipped with four circularly-polarized antennas. The precise coordinates and orientations of these antennas are recorded at the time of deployment. During sensing, the reader hops through different frequency channels. In this process, the response of the tags to the reader and the phase sequences at different frequencies can be observed and utilized for sensing.

Fig. 2 provides an overview of TagRecon's architecture, which consists of four main components:

Signal pre-processing module. In this module, we initiate by denoising the raw phase readings from the RFID reader, addressing the 2π folding issue and rectifying the abrupt π change. In order to eliminate the influence of multipath phenomena, we identify the channels with severe multipath interferences and eliminate the influence of these "dirty" channels. Finally, a linear fit of the "clean" phase after denoising is performed as a function of frequency.

Multi-task sensing module. Leveraging the multi-frequency and multi-antenna phase model, this module separates phase changes induced by varying factors (*i.e.*, propagation distance, tag orientation, and target material). The disentangled phase signals from the four antennas are then used for simultaneous 3D location and orientation sensing. This allows us to roughly determine



(a) Tag rotations along the X-axis (b) Phase changing with tag rotations

Fig. 3. Impact of tag rotations.

the orientation of the package and the stacking order of multiple packages in 3D space, using the position and orientation information of the single anchor tag.

Multiple tags matching module. In this module, we manage to pair the anchor tags designated to the same package when multiple packages exist simultaneously. We first calculate the pair weight between any two tags based on their orientation and distance. Then, we apply a stable roommates based matching algorithm to match the anchor tags on each package.

3D bounding box estimating module. In this module, with the acquired tag matching results from the previous module, we perform dimensions retrieval, 3D translation, and rotation estimation to accomplish the 3D bounding box estimation for the package.

4 PHASE-BASED 3D LOCATION AND ORIENTATION SENSING

In this section, we first introduce how TagRecon obtains the 3D position and orientation of a tag based on phase modeling and then describe the signal pre-processing procedures.

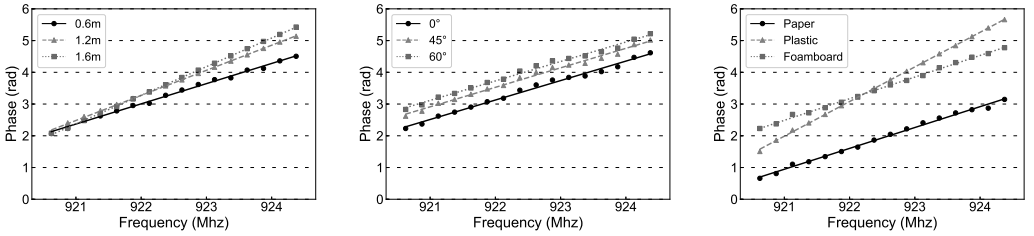
The RF phase is a common signal parameter reported by commercial RFID devices. Let d be the distance between one tag and one antenna. Since the signal traverses a total distance of $2d$ in backscatter communication, the phase rotation output by the reader in a traditional phase model can be expressed as [16]:

$$\theta = \left(2\pi \frac{2d}{\lambda} + \theta_{\text{dev}}\right) \bmod 2\pi, \tag{1}$$

where λ is the wavelength, and the term θ_{dev} describes the constant phase shift caused by devices' hardware characteristics. Ideally, changes in the receiving phase are thought to be affected only by distance changes.

4.1 Polarization-Sensitive Phase Model

In fact, we observe that the orientation of the tag also influences the phase value. To validate this, the following empirical study has been conducted. We fix both the reader and tag's positions, and ensure a consistent distance of 1 m between them. We place the tag parallel to the reader's antenna plane. Subsequently, the tag undergoes a 360° rotation around the X-axis, which is defined by the line connecting the antenna's center to the tag's center, as depicted in Fig. 3(a). Fig. 3(b) shows that the measured phase correlates linearly with both the polarization direction of the antenna and the tag's orientation. This suggests that tracking or positioning methods based on the traditional phase model may yield inaccurate results if the tag rotates.



(a) Propagation (b) Orientation (c) Material

Fig. 4. Phase changing with frequency hopping.

The polarization-sensitive phase model, which explicitly quantifies the respective impact of the 3D orientation of the tag on the phase change, is first proposed by [19]. According to its findings, when a signal propagates from a circularly-polarized antenna to a linearly-polarized tag, the tag's polarization vector w will affect the phase value θ_{ori} as:

$$\tan(\theta_{\text{ori}}) = \frac{2(u \cdot w)(v \cdot w)}{(u \cdot w)^2 - (v \cdot w)^2}, \tag{2}$$

where u and v are the horizontal and vertical unit direction vectors of the reader antenna (u and v also correspond to the polarization directions of two linearly polarized antennas, which are equivalent to circularly polarized antenna). The vector w denotes the direction of the tag. Eqn. 2 reveals that θ_{ori} relies only on the directions of the antenna and the tag (*i.e.*, u, v, and w). This makes it possible to estimate the tag's orientation from the phase value.

4.2 Sensing Methodology

RFID readers use frequency hopping to avoid interference. A typical UHF reader (*e.g.*, Impinj R420) hops between channels in the 920.375 ∼ 924.375 MHz ISM band. By extending the polarization-sensitive tag phase model to the multi-frequency scenario, we obtain:

$$\theta(f) = \left(\theta_{\text{prop}}(f) + \theta_{\text{ori}}(f) + \theta_{\text{dev}}(f)\right) \bmod 2\pi, \tag{3}$$

where $f \in f_1, \cdots, f_n$ ($n = 16$) is the center frequency of the CW signal. θ_{prop} is the phase rotation introduced by signal propagation, while θ_{dev} is caused by the hardware characteristics of the device.

Through the frequency hopping operation of the reader, RF-Prism [46] proposes the multi-frequency phase model as:

$$\begin{cases} \theta(f) = (k \cdot f + b) \bmod 2\pi, \\ k = \frac{4\pi d}{c} + k_t, \\ b = \theta_{\text{ori}} + b_t. \end{cases} \tag{4}$$

Eqn. 4 shows that the phase reading θ varies linearly with the frequency f. The parameters k_t and b_t are influenced by the material to which the tag is attached. The slope of the line (*i.e.*, k) depends on both the antenna-tag distance and the material, and the intercept of the line (*i.e.*, b) depends on both the orientation of the tag and the material.

For verification, Fig. 4(a) shows the phase readings acquired at different frequencies when the distance d between the antenna and the tag is set to be 0.6 m, 1.2 m, and 1.6 m, while other factors remain unchanged. A clear linear relationship is observed between the phase θ and the frequency f, with the slope varying based on the distance. Fig. 4(b) illustrates the phase readings obtained when the orientation of the tag is adjusted to 0°, 45°, and 60°, while other factors being fixed.

We find the θ_{ori} does not vary with frequency hopping and any change in the orientation of the tag causes the same phase shift at all frequencies. And the slope of the line is the same across different tag orientations. Fig. 4(c) depicts phase readings for tags with different attached materials: paper, plastic, and foamboard, while other factors remain unchanged. We can see that the slope and intercept of the lines obtained by different materials are varied. Therefore, we can obtain b and k by performing linear fitting on the phase readings θ collected under different frequencies f. Now, we have two equations and four unknown parameters, namely d, θ_{ori}, b_t and k_t. The number of unknowns exceeding the number of equations makes it impossible to directly solve for the key parameters of each influencing factor from the phase readings.

RF-Prism [46] further separates different factors in Eqn. 4 by adding equations from multiple antennas and introduces the multi-antenna phase model. However, adding new antennas introduces new unknown parameters (*i.e.*, the antenna-tag distance d and the tag orientation θ_{ori}). Specifically, there will be $2N$ equations and $2N + 2$ unknown parameters in a system with N antennas. The number of unknown parameters can be reduced by substituting the location and orientation of the tag for d and θ_{ori}. We denote the position of the tag as (x, y, z) in 3D coordinates and the directional vector of the tag as (α, β) in the spherical coordinate system. Eqn. 5 is rewritten by replacing variables d and θ_{ori} with $\text{dist}(\text{coordinate}(A_i), (x, y, z))$ and $\theta_{\text{ori}}(\text{direction}(A_i), \alpha, \beta)$ as follows:

$$\begin{cases} k_i = \frac{4\pi \cdot \text{dist}(\text{coordinate}(A_i), (x,y,z))}{c} + k_t, \\ b_i = \theta_{\text{ori}}\left(\text{direction}(A_i), (\alpha, \beta)\right) + b_t. \end{cases} \tag{5}$$

Let dist() denote the Euclidean distance function. A_i denotes the i-th antenna, whose coordinate and direction are measured during deployment. This makes it possible to solve the equation for the unknown parameters (x, y, z) and (α, β) by calculating the two parameters (k and b) of the fitted line of phase readings, in order to achieve tag location and orientation sensing. Moreover, this does not introduce any additional variables. It can be seen that the number of unknown parameters is always seven, irrespective of how many antennas are included. The four antennas located at different positions can generate eight independent equations, enough to determine these unknown variables and obtain a unique solution. By solving the above eight equations, we can separate the effect of attached materials and achieve tag localization and orientation sensing in 3D space.

4.3 Signal Pre-processing

TagRecon senses the position and orientation of the tag based on the multi-antenna phase model introduced before, which requires a linear fitting of the phase measurements with frequency hopping. But the following three phenomena: 2π folding, half-wave hopping, and multipath effect will disrupt the linear growth of phase with increasing frequency, thereby compromising the accuracy of linear fitting. In order to deal with these three cases, we summarize Eqn. 6:

$$\theta'_{i+1} = \begin{cases} \theta_{i+1}, & |\theta_{i+1} - \theta_i| \le \tau \\ \frac{\theta_i + \theta_{i+2}}{2}, & \tau < |\theta_{i+1} - \theta_i| \le \frac{\pi}{2} \\ \theta_{i+1} - \pi, & \frac{\pi}{2} < \theta_{i+1} - \theta_i \le \pi \\ \theta_{i+1} - 2\pi, & \pi < \theta_{i+1} - \theta_i \le 2\pi \\ \theta_{i+1} + \pi, & -\pi < \theta_{i+1} - \theta_i \le -\frac{\pi}{2} \\ \theta_{i+1} + 2\pi, & -2\pi < \theta_{i+1} - \theta_i \le -\pi \end{cases}, \tag{6}$$

where $i \in 1, \cdots, n$ ($n = 16$) is the channel sequence number, θ_i defines the raw phase from reader, θ'_i defines the corrected phase, and τ is a pre-defined threshold (we choose τ to be 0.5 in our experimentation).

Table 1. An example of the RFID database.

EPC	Item Type	Model	Length (l)	Width (w)	Height (h)
1234	Box	Model-1	10 cm	20 cm	30 cm
5678	Package	Model-2	15 cm	25 cm	35 cm
9012	Envelope	Model-3	20 cm	10 cm	2 cm
3456	Tube	Model-4	30 cm	5 cm	5 cm
7890	Pallet	Model-5	100 cm	120 cm	80 cm

Upon detecting the occurrence of 2π folding or half-wave jumps, we add or subtract 2π or π correspondingly to each phase value after the jump. To deal with multipath, we take advantage of the fact that only a few channels are anomalous, even in multipath-rich environments. This means that we can simply detect anomalous phase values and replace them with the average of phase measurements from neighboring 'clean' channels. With the corrected phase values after phase denoising from one antenna for a certain tag, we can use the linear model in Eqn. 4 to fit them, and then derive the two important parameters (*i.e.*, k and b) of that antenna-tag pair for the later tag localization and orientation sensing.

4.4 Package Representation

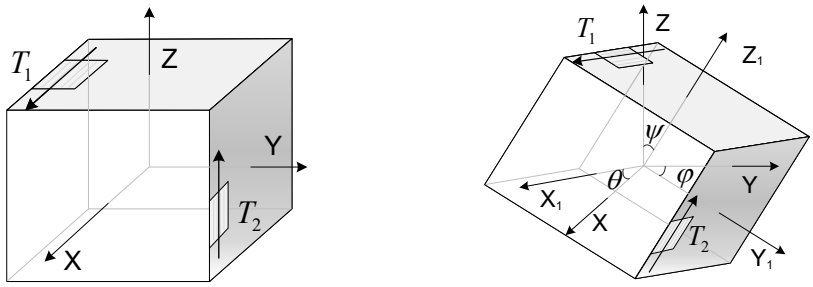
Intuitively, we can represent the tagged package's rotation by the single tag's orientation. A package's rotation can be characterized by three Euler angles: the roll γ, the pitch ϕ, and the yaw ψ [37]. These angles comprehensively capture the package's rotation. However, RF-3DScan [7] just describes the rotation of a package using the angle of the perpendicular plane of the package to the ground based its idealized assumptions of packages rotation. This merely corresponds to the yaw angle ψ and does not completely capture the package's rotations. Also, it is not possible to map a single tag's orientation to the three Euler angles. Assuming no roll occurs in the package, we can map the tag orientation to the yaw angle ψ. However, it is not feasible to provide a fine-grained representation of the rotation of an object with only the orientation of a single tag. Similarly, we can approximate the location of package's centroid using the tag's location, aiding in predicting package stacking order along various directions. Although this naive idea is convenient and cost-effective, it may produce ordering inaccuracies, particularly when tags are closely placed or packages are densely stacked. For those seeking a more precise representation and arrangement of multiple packages, we design a dual-tag solution, which we will elaborate in the following section.

5 3D BOUNDING BOX ESTIMATION OF TAGGED PACKAGES

High-precision package reconstruction requires more than just determining the stacking order and yaw angle. To achieve the goal, we design a 3D bounding box estimation method for packages in this section.

5.1 Package Dimensions Retrieval

3D bounding box estimation is a key technical component in multi-object 3D reconstruction tasks. Such a box is defined by its dimensions $D = (d_x, d_y, d_z)$, representing the length, width, and height; its translation $T = (t_x, t_y, t_z)$, indicating the box center's position; and its rotation $R = (\gamma, \phi, \psi)$, describing the rotation using Euler angles [31]. By accurately determining the translation, rotation, and dimensions of a package in the world coordinate system, it is ensured that each package is



(a) Canonical space (b) Rotated case

Fig. 5. Rotation estimation of a package.

placed in the correct location, thus distinguishing it from other packages. In the rest of this section, we will delve into estimating these parameters for a certain package.

In bulk commodities transportation scenarios, it usually involves a large number of packages of the same type and size, so package dimensions are often standardized. As shown in Table 1, the RFID database typically includes detailed category-level information about the item associated with the tag, such as its type, model number, and dimensions. For packages, the database often records their length l, width w, and height h. For any tag detected by the RFID system, its unique EPC can be read by the reader. Using this EPC, the RFID database can be queried to retrieve the specifications of the tagged package. These dimensions can be represented as $D' = (l, w, h)$.

5.2 Rotation Estimation

5.2.1 Package Rotation Estimation. In 3D space, any two non-collinear directional vectors can be used to derive a third, orthogonal vector, and together, these three vectors determine a coordinate system. In the case of a rectangular package with three mutually perpendicular sides, it is sufficient to know the direction of two of the sides, and we can calculate the direction of the third side via the right-hand rule. Inspired by this point, we get a key intuition that we can deploy two anchor tags and use the direction vectors of the two anchor tags to determine the coordinate system.

5.2.2 Anchor Tags Deployment Rules . To accurately estimate the translation T and the rotation R of each package, but with as few tags as possible and with minimum deployment cost, we formulate the following two rules for the deployment of anchor tags:

1. The tags are not in the same plane. For rectangular packages, we can place the two tags on different faces. This deployment reduces the possibility that the reference tags are all in the signal-blind zone due to the complete occlusion of one face of the package.

2. The tags are oriented perpendicular to each other. Under this rule, at most one of the tags is in the blind direction of the antenna signal. Moreover, employing two mutually perpendicular vectors would be very convenient and efficient when determining the rotation matrix with the algorithm described later. More importantly, this makes tag deployment easier, simply sticking the tags along the sides of the package that are perpendicular to each other.

Given the above rules, Fig. 5(a) depicts a viable tag deployment strategy in canonical space. One reference tag can be placed on the top surface along the X-axis, and the other on the side surface along the Z-axis, resulting in the direction vectors of two tags: $w_1 = (0°, 90°)$ and $w_2 = (0°, 0°)$. After the package rotates, as shown in Fig. 5(b), the direction vectors become $w_1^{'} = (\alpha_1, \beta_1)$ and

Algorithm 1 Algorithm for Rotation Computing

Require: Source vectors: $\{w_1, w_2\}$; Target vectors: $\{w_1^{'}, w_2^{'}\}$;
Ensure: The package's rotation matrix: R_m; Euler angles of the package's rotation: R;
1: Convert directional vectors in spherical coordinates to 3D Cartesian points: $\{c_1, c_2\}$ and $\{c_1^{'}, c_2^{'}\}$;
2: Define the source matrix $c^T = \begin{bmatrix} c_1 & c_2 \end{bmatrix}$and target matrix $c'^T = \begin{bmatrix} c_1^{'} & c_2^{'} \end{bmatrix}$
3: Compute the covariance matrix: $H = c^T \cdot c^{'}$;
4: Decompose H using SVD: $U, V = \text{SVD}(H)$;
5: Compute the rotation matrix: $R_m = V \cdot U^T$;
6: Convert the rotation matrix R_m to Euler angles R;
7: **return** R_m, R;

$w_2^{'} = (\alpha_2, \beta_2)$. The angle between the new coordinate system $X_1Y_1Z_1$ and the canonical system XYZ represents our desired rotation $R = (\gamma, \phi, \psi)$.

By utilizing the phase-based localization and orientation sensing method detailed in Section 4, we obtain the direction vectors $w_1^{'}$ and $w_2^{'}$ of the two anchor tags after rotation. In canonical space, these direction vectors are represented as w_1 and w_2. Given both the direction vectors before and after the object's rotation, our objective is to identify the rotation $R = (\gamma, \phi, \psi)$ by establishing the transformations between these two vector sets.

In order to obtain R, we simplify the task to the absolute orientation problem [15]. The absolute orientation problem is well-known in computer vision, and its goal is to find a transformation that maps one set of points to another such that the distance between the two sets of points is minimized. The basic intuition for using the idea of absolute orientation is that we can consider the direction vector as a point so that we have two point sets, described as 'source' and 'target'. The transformation from the target set to the source set is the rotation R we aim to determine. More precisely, our goal is to find a rotation matrix that maps the source vectors w_1 and w_2 to the target vectors $w_1^{'}$ and $w_2^{'}$.

We leverage the Kabsch-Umeyama algorithm [23] to solve this problem, as shown in Algorithm 1. Initially, we convert directional vectors in spherical coordinates to points in 3D Cartesian coordinates. These vectors can be conceived as originating from the origin and pointing towards a specific point in space. The conversion can be accomplished using the formula Eqn. 7:

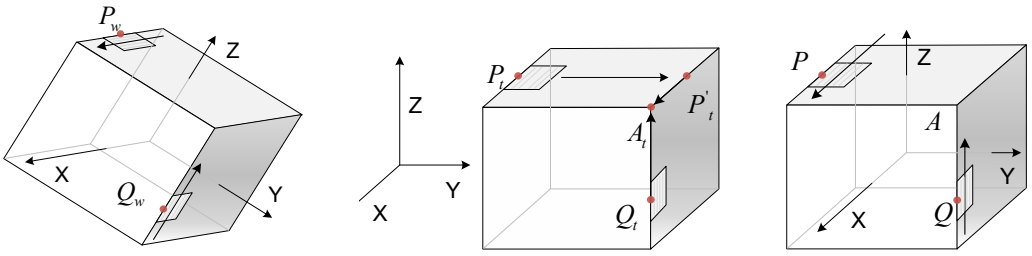
$$\begin{cases} x = r \cdot \sin(\beta) \cdot \cos(\alpha), \\ y = r \cdot \sin(\beta) \cdot \sin(\alpha), \\ z = r \cdot \cos(\beta), \end{cases} \tag{7}$$

where r denotes the radial distance, *i.e.*, the distance from the origin to the point. By setting $r = 1$, we transform these vectors into unit vectors in 3D space, maintaining only their directionality. After this transformation, we arrange the direction vectors $\{c_1, c_2\}$ and $\{c_1^{'}, c_2^{'}\}$ into the corresponding matrices: $c^T = \begin{bmatrix} c_1 & c_2 \end{bmatrix}$ and $c'^T = \begin{bmatrix} c_1^{'} & c_2^{'} \end{bmatrix}$, where c is the matrix of source vectors and $c^{'}$ is that of the target vectors. Given two sets of corresponding points, their relationship can be captured by the cross-covariance matrix, expressed in Eqn. 8:

$$H = c^T \cdot c^{'}, \tag{8}$$

where H denotes the cross-covariance matrix, reflecting the geometric relationship between the two sets of points. One method to extract the rotation matrix from this relationship is to use the Singular Value Decomposition (SVD). We then perform a SVD of H as below:

$$H = U\Sigma V^T, \tag{9}$$



(a) World coordinate (b) Intermediate state (c) Canonical space

Fig. 6. Translation estimation of a package.

where U is the left singular vector matrix of H, Σ is the diagonal matrix containing the singular values of H, and V^T is the right singular vector matrix of H. The optimal rotation matrix R_m that aligns the two sets can be calculated as Eqn. 10:

$$R_m = V \cdot U^T. \tag{10}$$

With this, we get the rotation matrix R_m, but it may not be intuitive to directly read the exact angle of rotation from the rotation matrix. Then we can convert the rotation matrix to the corresponding Euler angles $R = (\gamma, \phi, \psi)$. The conversion is given by Eqn. 11 as follows:

$$\begin{cases} \gamma = \arctan 2(r_{32}, r_{33}), \\ \phi = \arctan 2(-r_{31}, \sqrt{r_{11}^2 + r_{21}^2}), \\ \psi = \arctan 2(r_{21}, r_{11}), \end{cases} \tag{11}$$

where r_{ij} denotes the element of row i and column j of the rotation matrix. Note that these formulas only apply to the rotation order XYZ. If other rotation orders are used, the formulas need to be adjusted accordingly. So far, we have successfully estimated the rotation $R = (\gamma, \phi, \psi)$ for a single package.

5.3 Translation Estimation

After acquiring the rotation R and the dimensions D of the package, the remaining component is the translation T. We can establish the correspondence between the points in the world coordinate system shown in Fig. 6(a) and the canonical space coordinate system shown in Fig. 6(c). The relation is captured by Eqn. 12:

$$K_w = D \cdot R_m \cdot K + T, \tag{12}$$

where K and K_w represent the coordinates of a point in canonical space and the world coordinate system, respectively. Thus, once we know the locations of a single point in both coordinate systems, we can solve for T. Although we can determine the positions of the two anchor tags, P and Q, in the real world using the RFID-based localization method, we do not have the exact positions of these tags in the canonical space. Consequently, a direct solution for T using these anchor tags remains elusive. However, our deployment strategy places the dual anchor tags on the package edges, providing us with some prior knowledge. By utilizing the known deployment positions of the two anchor tags, we can infer the position of at least one endpoint in the world coordinate system. For a clearer understanding, we introduce an intermediate state situated between the canonical space and the translated-rotated state. The package in this state only translates relative to canonical space and lacks the rotation and dimension scaling relative to the world coordinate system. The

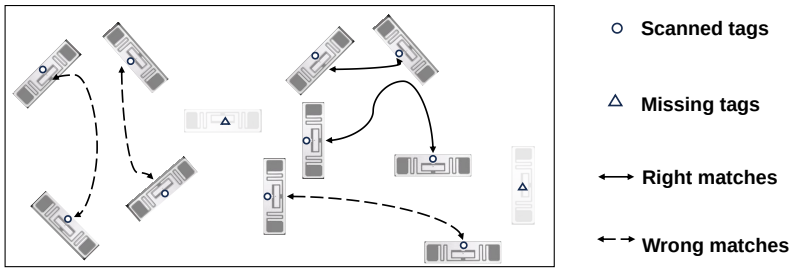


Fig. 7. Multiple tags matching.

intermediate state is demonstrated in Fig. 6(b), where the edges of the package are unit length, and each side is parallel to the corresponding side in canonical space. We can describe the relationship between the point K_t in this intermediate state and its corresponding points K and K_w in the other two coordinate systems as follows:

$$\begin{cases} K_w = D \cdot R_m \cdot K_t, \\ K_t = K + \dfrac{T}{D \cdot R_m}. \end{cases} \tag{13}$$

With Eqn. 13, we can determine the estimated coordinates of P_t and Q_t as (x_1, y_1, z_1) and (x_2, y_2, z_2), respectively. We define $P_t^{'}$ as the point on the top surface's edge, parallel to the edge where P_t is located. $P_t^{'}$ is located at the position $(x_1, y_1 + 1, z_1)$ where P_t is shifted along the direction $(0, 1, 0)$ by length 1. Ideally, the point A_t, defined by the intersection of the line passing through P_t in the direction $(1, 0, 0)$ and the line passing through Q_t in the direction $(0, 0, 1)$, is one of the endpoints of the rectangle. However, due to potential location errors, these two lines might not intersect within a plane.

To deal with this situation, we can find two points on the above two lines such that their distance is minimal, and consider the midpoint of the line connecting these two points as the approximate intersection of the two lines. Using this mechanism, we can obtain an estimate of the coordinates of A_t as $(x_2, \frac{y_1+y_2+1}{2}, z_1)$. In canonical space, its corresponding point A has coordinates $(0.5, 0.5, 0.5)$. Now that A_t and A are acquired, we can compute T by Eqn. 14 as below:

$$T = D \cdot R_m \cdot (A_t - A) = D \cdot R_m \cdot (x_2 - 0.5, \frac{y_1 + y_2}{2}, z_1 - 0.5). \tag{14}$$

5.4 Multiple Tags Matching

In the previous part of this section, we implement 3D bounding box estimation for a single package utilizing two anchor tags. However, in scenarios where multiple packages exist simultaneously, we must match the anchor tag pair for each package. In high-efficiency assembly line operations, recording the pairing relationship of tags is impractical due to significant operational overhead. This process requires additional manpower, leading to increased labor costs in practical applications. Additionally, the time required to record these relationships extends the processing time for each package, which is counterproductive in terms of efficiency. Therefore, instead of recording tag pair relationships, we adopt a method of dynamically matching the anchor tag pair for each package. Based on the following observation that the challenge of pairing tags is similar to finding matches in a set, we can reduce the problem to the stable roommates problem [17]. The stable roommates problem involves members of a set where each member ranks the others according to preference.

Algorithm 2 Algorithm for Tag Matching

Require: Tags $O = \{o_1, o_2, ..., o_n\}$; The weight matrices $w_{i,j}$;
Ensure: The matching result of tags res;

```
for o_i in O do
    Sort {o_j | j ≠ i} by w_{i,j} in ascending order as o_i's preference list;
    Remove tags with w_{ij} ≥ 1;
end for
Initialize all tags as free;
while there exists a free tag o do
    e ← the first tag in o's preference list that has not been proposed;
    if e is free then
        Add (o, e) to res;
    else if e prefers o to its current partner o' then
        o' becomes free and remove (o', e) from res;
        Add (o, e) to res;
    else
        e rejects o;
    end if
end while
return res;
```

The goal is to find a stable pairing such that no two unpaired participants would prefer each other over their current pairings.

The underlying rationale for applying the stable matching concept is that, within any tag pair, the orientations of the two tags are usually perpendicular to each other approximately. Moreover, the distances between the positions of such paired tags are typically much smaller than the distances between unrelated tags. Given these observations, we can prioritize the matching of specific tag pairs based on their orientation and distance preferences. One key aspect of our method is the emphasis on the orientation relationship between two tags. If the angle between their orientations deviates significantly from 90°, we assign a higher weight to the pair. Conversely, if the angle closely approximates 90°, we then factor in the distance between the tags. A higher weight suggests a lower likelihood of the two tags being a suitable pair. Considering two tags, G and S, with (α_g, β_g) and (α_s, β_s) denoting their respective orientations and (x_g, y_g, z_g) and (x_s, y_s, z_s) representing their positions, the weight $w_{g,s}$ can be computed as:

$$w_{g,s} = \begin{cases} 1, & \text{if } |w_{\text{ori}}| > k \\ w_{\text{ori}} \cdot \text{dist}\left((x_g, y_g, z_g), (x_s, y_s, z_s)\right), & \text{if } |w_{\text{ori}}| \leq k \end{cases} \tag{15}$$

where

$$w_{\text{ori}} = \cos((\alpha_g, \beta_g), (\alpha_s, \beta_s)) = \cos(\alpha_g - \alpha_s) \cdot \sin(\beta_g) \cdot \sin(\beta_s) + \cos(\beta_g) \cdot \cos(\beta_s).$$

Here, w_{ori} is defined by the cosine of the angle between two tags' orientation vectors. As the angle approaches 90°, w_{ori} approaches 0. And k is a pre-defined threshold which we set to 0.34, *i.e.*, the cosine of 20°. If w_{ori} surpasses k, we deem the tags are mismatched and assign a weight of 1. When w_{ori} is below k, the weight $w_{a,b}$ factors in both the distance and the angle between the tags.

By computing w_{ori}, we can find a stable match for all tags by determining the preference list for each tag. But our scenario differs from the standard stable roommates problem. Not every tag is expected to find a match because some tags might go undetected. In logistics and warehousing

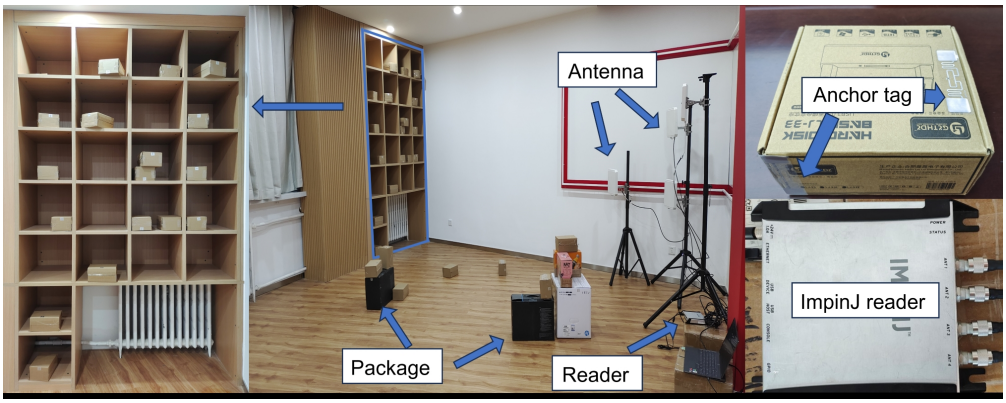


Fig. 8. Experiment setup. We build a prototype and evaluate TagRecon using commercial RFID devices.

environments, where packages are densely stacked, objects might obscure RFID tags. If a tag is obscured by objects, the RF signal emitted by the reader may not reach the tag. Additionally, the indoor multipath effect can lead to direct and reflected RF signals nullifying each other at certain tags' locations. These challenges can prevent the effective activation of some tags, leading to their non-detection. As illustrated in Fig. 7, this "missing tags" problem causes that not every scanned tag has a corresponding match. Such tags, when engaged in the matching process, lead to wrong matches.

To address this issue, we employ the Gale-Shapley algorithm [12], as detailed in Algorithm 2. Initially, the algorithm iterates through all tags to construct a preference list for each tag o_i, based on the weight matrix $w_{i,j}$. Specifically, the algorithm sorts all the other tags o_j in ascending order of their weights $w_{i,j}$ as o_i's preference list. Tags with weights greater than or equal to 1 are disregarded as irrelevant. Subsequently, every tag's status is set to "free". As the algorithm processes the free tags in each iteration, it selects an arbitrary free tag o. It then identifies the most highly preferred tag e that has not been proposed yet in its preference list. If e is also free, a match between o and e is established. If e is currently paired with another tag o', but favors o over o', then the match with o' is dissolved, and e is paired with o. This process continues until all tags are matched or deemed unmatchable, at which point the algorithm returns the final matching results.

6 IMPLEMENTATION AND EVALUATION

We implement TagRecon using COTS RFID devices and conduct performance evaluation in a typical indoor environment as shown in Fig. 8.

6.1 Prototype

Hardware: TagRecon employs an Impinj Speedway Revolution R420 reader, compatible with the EPC Gen2 standard. This reader operates in the UHF frequency band spanning 920.5 $\sim$ 924.5 MHz, hopping over 16 frequency channels. It connects to our host system via Ethernet. Our setup incorporates four antennas with circular polarization and 8 dBi gain, each measuring 22.5 cm $\times$ 22.5 cm $\times$ 4 cm. We utilize four types of tags from Alien Corp, modeled "Squiggle", "Sguig", "Square" and "3D Func". On average, these tags are economically priced at about 7 cents each [1], which is cost-competitive in the manufacture of commodities.

Software: Communication with the reader is facilitated using the Low-Level Reader Protocol (LLRP), which the Impinj reader extends to support phase reporting. We adjust the reader's configuration to instantly report its readings upon detecting tags. The software of TagRecon is implemented

using Java language. All our programs run on a Lenovo R9000X laptop, which interfaces with the reader through LLRP. This laptop is equipped with an AMD Ryzen 7 5800H CPU and 24GB of memory.

System Settings: We conduct the experiment in an indoor environment with a spatial dimensions of 8 m × 5 m × 3 m. We place four antennas at the corners of the room with the height from the range of [1 m, 2 m]. The antennas are arranged with a simple yet effective rule: they should not be co-planar, or oriented in the same direction. This rule ensures comprehensive coverage with the RFID tags throughout the space and capture sufficient spatial information for accurate readings. Before starting experiments, we measure the coordinates and orientations of the antennas. The ground truth for locations is measured using a laser range finder with an expected error of 0.15 mm, while the orientations are determined by an Inertial Measurement Unit (IMU) with an accuracy of 0.2°. The packages are placed in various positions on the floor, with some closely stacked to emulate dense storage conditions. Additionally, a subset of packages is distributed across the floor with different orientations, randomized to mimic the natural variability found in real-world scenarios. Furthermore, other packages are intentionally placed on shelves positioned along the room's periphery, with the shelves varying in height from 0.2 m to 3 m. This tiered arrangement complements the floor placement by adding another dimension to the system's challenge, testing its ability to read tags from different heights and distances. Such comprehensive setup, with diversity in package placement and orientation, is designed to thoroughly evaluate our system's performance in a wide range of conditions. Other system settings are tunable parameters, which we will evaluate later.

6.2 Location and Orientation Sensing Accuracy

The location and orientation sensing accuracy of tags play a key role in TagRecon's performance. We compare TagRecon with 3D-OmniTrack [19], a system designed to track objects in 3D space by considering both the distance and the orientation. The comparison under three different configurations is illustrated in Fig. 9 and Fig. 10. In the x-label, a minus symbol signifies that the associated factor remains constant in the experiment, while a plus symbol indicates a variable factor.

As shown in Fig. 9, when both the tag's orientation and the material of the attached object are invariant, TagRecon and 3D-OmniTrack achieve similar mean location accuracies of 8.1 cm and 8.4 cm, respectively. Nonetheless, when different materials are introduced without changing the position, TagRecon maintains a location error of 8.5 cm, whereas 3D-OmniTrack's error swells to 15.1 cm. This discrepancy can be attributed to TagRecon's approach: it factors in the effect of the target material into its phase model, rather than dismissing it as merely general hardware noise. We also find that rotating the tag does not further enlarge the performance gap. The reason lies in that both TagRecon and 3D-OmniTrack's phase models allow for the 3D orientation's effect of objects.

Similar trends in orientation accuracy are discernible in Fig. 10. With consistent tag-antenna distances and object materials, TagRecon achieves slightly better results with 5.0° error, compared to 3D-OmniTrack's 5.4°. However, as the material varies, TagRecon retains an error of 5.1°, while 3D-OmniTrack's error surges to 8.5°. Adjustments in the tag-antenna distance exert minimal influence on orientation accuracy, yielding errors of 5.1° and 8.6°, respectively. These results illustrate that TagRecon improves the location and orientation sensing accuracy by modeling the influence of various factors on the phase value. Therefore, it can be applied to many other more complex scenarios.

6.3 Multiple Tags Matching Accuracy

Aiming for accurate 3D bounding box estimation of tagged packages, it is vital to match anchor tags with high precision. To evaluate TagRecon's matching algorithm, we deploy several packages on

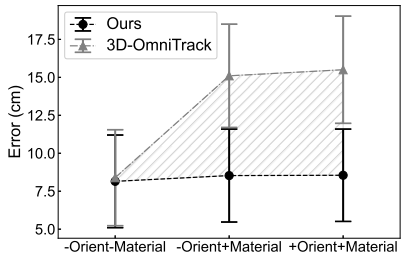


Fig. 9. Comparion of location error.

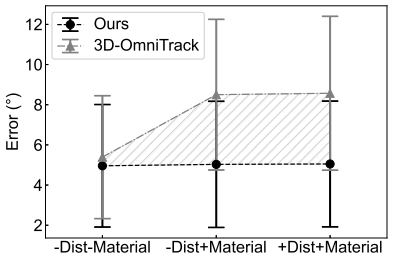


Fig. 10. Comparion of orientation error.

the experimental environment. Most of these packages are attached with two anchor tags, while a small minority of packages are deployed with only one tag. The distance between adjacent packages is randomly chosen from the range of [10 cm, 50 cm]. For each case, the result is averaged from 50 experiments with the same setting. We run our matching algorithm mentioned in Section 5 to pair tags belonging to the same package and compute the corresponding matching accuracy.

Missing ratio of tags: To evaluate the performance of our matching algorithm in handling the "missing tags" problem, we further test our matching algorithm under different missing ratios of tags as shown in Fig. 11. We vary the missing ratio of tags by changing the number of packages with one tag. It can be seen that as the missing ratio increases from 0 to 40%, the matching accuracy decreases gradually from 98%, but remains above 85%. This implies that our matching algorithm effectively tackles the outliers resulting from missing tags. In other experiments, we adopt a default missing ratio of 20%.

Number of packages: In practical deployments, there could be multiple tagged packages in a logistics scenario. Hence, we assess our matching algorithm in such multi-package situations. We change the number of packages with a pair of tags from five to thirty on the floor or on shelves of different levels. The results are plotted in Fig. 12. We have the following observations from the figure: a) with up to five dual-tagged packages in the reading zone, TagRecon performs excellently with a matching error below 10%; b) the error slightly increases as more packages are added, rising to 19% with thirty packages; c) even with a significant number of tags (*e.g.*, sixty), TagRecon maintains a moderate accuracy of about 80%. For our subsequent experiments, we use five packages as default.

Type of tags: To examine TagRecon's robustness with different tag types, we test four models: "Squiggle", "Sguig", "Square", and "3D Func". Each of these tag models has distinct antenna size and shape, as depicted in Fig. 13. We find that although the errors of all models maintain a small value (less than 15%), there are nuances in their performances. Specifically, "Squiggle" and "Sguig" exhibit similar accuracies, 89% and 90%, respectively. In contrast, the "Square" model has a slightly reduced accuracy of 85%, and the "3D Func" model outperforms the others with an accuracy of 93%. This difference can be explained by the antenna size of the tags. For instance, the "Square" tag, being more compact at 22.5 mm × 22.5 mm, contrasts with the larger "3D Func" tag at 50 mm × 50 mm. Generally, tags with larger antennas can harness more energy from the reader, resulting in a stronger backscattered signal (*i.e.*, higher SNR), thereby boosting the matching accuracy. In our subsequent tests, the "Sguig" model is used as the standard.

6.4 Overall Performance

We compare TagRecon with the sole RFID-based package reconstruction system: RF-3DScan [7]. We implement two schemes: the single-tag and the dual-tag setup. In the single-tag setup, an anchor

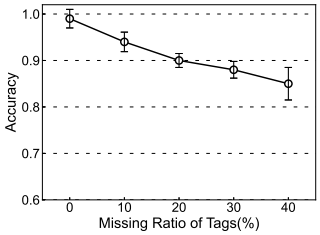


Fig. 11. Matching accuracy vs. missing ratio of tags.

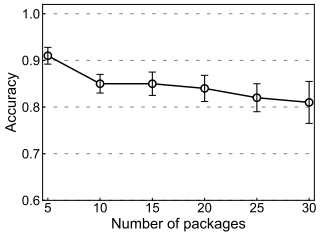


Fig. 12. Matching accuracy vs. number of packages.

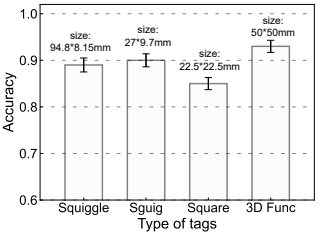


Fig. 13. Matching accuracy vs. type of tags.

tag is placed on the top surface's center of the package. And the dual-tag setup is deployed as shown in Fig. 5(a). RF-3DScan compares the angle profiles of the different reference tags deployed around a package in 1D scanning or 2D scanning mode to determine the package orientation and stacking for 3D reconstruction. The main differences between the two systems are summarized in Table 2.

Deployment cost: While TagRecon involves the use of more antennas compared to RF-3DScan, it significantly reduces the number of required tags and the complexity of tag deployment. TagRecon requires only one tag for the single-tag setup and two for the dual-tag setup, whereas RF-3DScan necessitates four tags for 1D scanning and six for 2D scanning. The primary long-term cost of such systems is driven by the complexity involved in deploying these tags. By minimizing the number of tags required, TagRecon not only lowers the direct expenses of acquiring additional tags but also streamlines the deployment process. Consequently, even though TagRecon's initial setup might involve higher antenna overhead, its approach ultimately results in significant long-term savings by substantially decreasing the complexity and costs related to tag deployment and maintenance.

Antenna scanning mode: TagRecon operates in a static antenna scanning mode, which is generally less complex and low-cost compared to RF-3DScan's mobile antenna scanning approach. The latter requires the antenna to traverse in a straight line at a constant speed, typically achieved by mounting the antenna on a platform such as a cart. This introduces additional hardware costs associated with the cart. Moreover, in complex environments, obstacles can prevent the cart from maintaining a strictly constant and linear motion, further complicating the operation and reducing the accuracy. To avoid using the mobile antenna mode, TagRecon introduces a more advanced multi-antenna phase model to achieve accurate tag position and orientation sensing.

Package reconstruction granularity: In contrast to RF-3DScan's approach of just coarsely estimating package orientation by focusing primarily on the yaw angle and stacking order, TagRecon

Table 2. Comparison of TagRecon and RF-3Dscan methods.

	TagRecon		**RF-3Dscan** [7]	
	Single-tag	**Dual-tag**	**1D scanning**	**2D scanning**
Number of deployed tags	1	2	4	6
Antenna scanning mode	static	static	mobile	mobile
Translation of packages	no	yes	no	no
Rotation of packages	no	yes	no	no
Package stacking in 3D space	yes	yes	no	yes

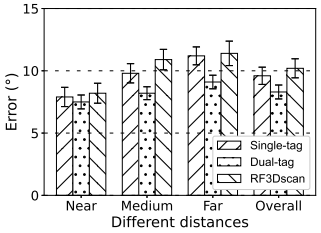


Fig. 14. Angle accuracy vs. distance.

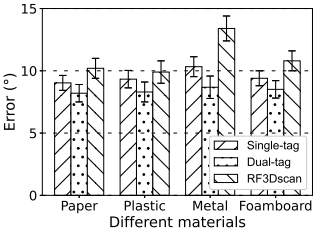


Fig. 15. Angle accuracy vs. material.

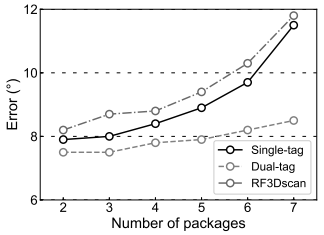


Fig. 16. Angle accuracy vs. number of packages.

significantly elevates package 3D reconstruction's granularity. It pioneers in providing detailed translation and rotation estimations for packages by leveraging dual tags. This substantial advancement not only improves the accuracy of package placement, but also increases the efficiency and accuracy of automated systems, making TagRecon a more advanced tool in package handling and management.

3D stacking capability: Unlike RF-3DScan, which limits its capability of estimating 3D stacking order to its 2D scanning mode, TagRecon supports package stacking in 3D space in both single-tag setup and dual-tag setup. This ability in 3D stacking allows TagRecon to adapt more efficiently to varying spatial requirements and package arrangements, offering a more versatile solution for complex stacking scenarios.

6.4.1 Yaw Angle Accuracy. In this section, we test the performance of TagRecon on yaw angle, stacking, translation, and rotation accuracies of tagged packages. Since the 1D scanning is not as accurate, robust, and comprehensive as the 2D scanning, we implement 2D scanning to represent RF-3DScan in comparison to our TagRecon. The yaw angle accuracy of TagRecon is evaluated under varying distances, materials, and number of tagged packages.

Different distances: Fig. 14 shows the yaw angle of TagRecon when the paper package is located at different positions. We divide all the positions into three different regions (*i.e.*, near, medium, and far), based on the tag-antenna distances. Both configurations of TagRecon exhibit competitive accuracy in comparison to RF-3DScan. However, the dual-tag setup's accuracy is notably superior to the other two. Specifically, the dual-tag setup achieves accuracy results of 7.5°, 8.2°, and 9.1° for the near, medium, and far regions, respectively. We think this discrepancy arises because, in complex scenarios where packages deviate from RF-3DScan's ideal assumptions, only the dual-tag setup can yield accurate yaw angles. In contrast, the other two configurations can merely operate under idealized assumptions.

Different materials: Fig. 15 plots the yaw angle errors of TagRecon when the tags are attached to packages with different materials. TagRecon achieves a mean angle error of 9.5° for single-tag setup and 8.4° for dual-tag setup across four materials, both lower than the 11.0° of RF-3DScan. Metal exhibits a slightly higher error than the other materials with TagRecon; but even so, its error is still below 11°, notably better than RF-3DScan's 13.4°. Another important observation in Fig. 15 is that the material of the target does not have an apparent impact on the performance. These results validate TagRecon's ability to handle material diversity.

Number of packages: The yaw angle errors of TagRecon, corresponding to the number of packages ranging from two to seven, are illustrated in Fig. 16. The angle accuracies of RF-3DScan and the single-tag setup of TagRecon decrease with with the increasing number of packages. The presence of more packages intensifies the multipath effect, leading to a reduction in the number

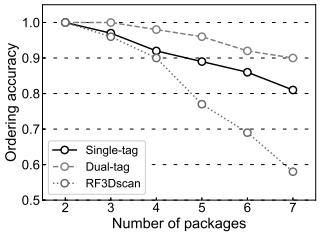


Fig. 17. Stacking accuracy vs. number of packages.

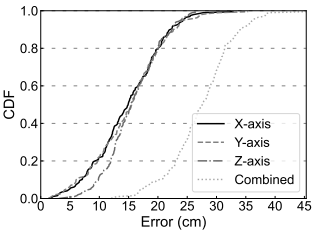


Fig. 18. CDF of translation error.

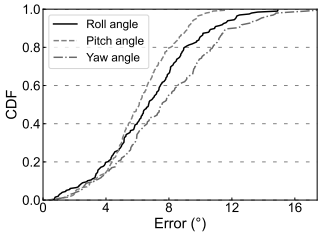


Fig. 19. CDF of rotation error.

of tags being activated, and consequently, a decrease in accuracy. Contrarily, the dual-tag setup's angle error remains relatively stable regardless of the number of stacked boxes, hovering around 2°.

6.4.2 Stacking Accuracy. To evaluate the performance of package stacking, we change the number of packages from two to seven, in front of the antennas at a distance of about 2 m. The packages are close to each other. Fig. 17 plots the ordering accuracy with different types of schemes. When the number of packages increases, the difference in ordering accuracy between TagRecon and RF-3DScan increases. We find that using three approaches can accurately determine the package stacking with accuracies above 96% when there are two or three packages. However, with more packages in the stack, the ordering accuracy of TagRecon decreases but is still above 80%. Among the two setups of TagRecon, the dual-tag setup's accuracy is even higher than 90%. The main reason why the single-tag setup is less robust than the dual-tag setup is that the position of the tag is not a perfect substitute for the position of the center of the tagged package. As the number of packages increases, this imperfect substitution creates an increasing inaccuracy.

7 DISCUSSIONS

We discuss the limitations and the corresponding challenges for system practicality in this section.

Deployment cost. Despite reducing the number of passive RFID tags needed per package, we have increased the number of reader antennas required. Each reader antenna we use costs approximately $100, and the system requires at least four antennas. Thus, designing low-cost RFID reader antennas remains a significant challenge in current RFID-based applications. The high deployment cost could be a barrier for widespread adoption, especially in cost-sensitive environments.

Priori knowledge of anchor tag placement and package dimensions. For the system's effectiveness, we require prior knowledge of anchor tag placement, *i.e.*, the dual tags are deployed perpendicular to each other on the top and side surfaces of each package, and package dimensions can also be retrieved through its EPC. Such requirement presents a challenge, particularly for applications involving non-standardized packages where dimensions cannot be predetermined. In practical scenarios, variability in package sizes and tag placements can hinder the system's usability and scalability.

Limited tag reading rate. Due to this limitation, all antennas may not simultaneously capture data from all tags, especially in dense-tag environments. This restriction affects the system's performance in scenarios where numerous tags are present and need to be read concurrently. Enhancing the tag reading rate is essential to improve system efficiency and reliability in high-density tagging situations.

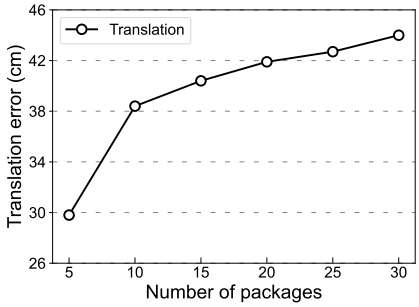


Fig. 20. Translation error vs. number of packages.

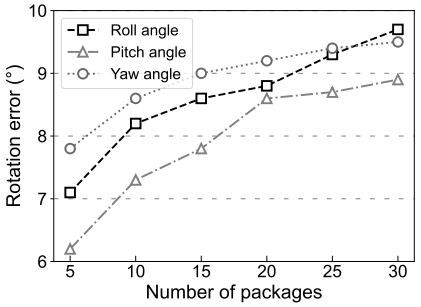


Fig. 21. Rotation angle error vs. number of packages.

In conclusion, while our system demonstrates significant potential for fine-grained multi-object 3D reconstruction using RFID technology, addressing the above limitations is crucial for its practical implementation and widespread adoption.

8 CONCLUSION

In this work, we design and implement a fine-grained 3D reconstruction system for multiple tagged packages based on commercial RFID devices. A key innovation is the use of two anchor tags on each package, allowing us to infer its accurate pose from the absolute locations and orientations of these tags. We propose a 3D bounding box estimation method to obtain accurate poses of the tagged packages. Additionally, we devise a solution based on the stable roommates algorithm to match the dual anchor tags attached to the same package. Experimental results demonstrate that our system can achieve accurate 3D reconstruction of multiple packages while maintaining robustness in various settings. We believe that our work will enhance the quality, efficiency, and safety of logistics and warehousing applications and further expand the possibilities of RFID technology in Industry 4.0 sectors.

ACKNOWLEDGMENTS

This research is supported in part by the National Natural Science Foundation of China (Grants No. 62372045 and No. 62072040) and the Beijing Institute of Technology Research Fund Program for Young Scholars.

REFERENCES

[1] 2019. Alien Tag Family Product Brief. http://www.alientechnology.com/wp-content/uploads/Tag.
[2] 2021. RFID's Renaissance in Retail. https://www.mckinsey.com/industries/retail/our-insights/rfids-renaissance-in-retail
[3] 2023. Amazon Just Walk Out RFID Technology. https://www.aboutamazon.com/news/retail/amazon-just-walk-out-rfid-technology
[4] 2023. The AR Warehouse – Augmented Reality in Logistics. https://www.teamviewer.com/en/ar-warehouse/
[5] 2023. Improving Logistics and Warehousing: Vision Pro's Role in Efficient Inventory Management and Product Tracking. https://gtcsys.com/improving-logistics-and-warehousing-vision-pros-role-in-efficient-inventory-management-and-product-tracking/
[6] Mathieu Bouet and Aldri L Dos Santos. 2008. RFID Tags: Positioning Principles and Localization Techniques. In *Proceedings of the 1st IFIP Wireless Days*. 1–5.
[7] Yanling Bu, Lei Xie, Yinyin Gong, Jia Liu, Bingbing He, Jiannong Cao, Baoliu Ye, and Sanglu Lu. 2019. RF-3DScan: RFID-based 3D Reconstruction on Tagged Packages. *IEEE Transactions on Mobile Computing* 20, 2 (2019), 722–738.

[8] Chunhui Duan, Jiajun Liu, Xuan Ding, Zhenhua Li, and Yunhao Liu. 2021. Full-Dimension Relative Positioning for RFID-Enabled Self-Checkout Services. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (UBiComp)* 5, 1, 1–23.

[9] Chunhui Duan, Xing Rao, Lei Yang, and Yunhao Liu. 2017. Fusing RFID and Computer Vision for Fine-grained Object Tracking. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*. 1–9.

[10] Francis Engelmann, Konstantinos Rematas, Bastian Leibe, and Vittorio Ferrari. 2021. From Points to Multi-object 3D Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 4588–4597.

[11] Michael Firman, Oisin Mac Aodha, Simon Julier, and Gabriel J Brostow. 2016. Structured Prediction of Unobserved Voxels from a Single Depth Image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 5431–5440.

[12] David Gale and Lloyd S Shapley. 1962. College Admissions and the Stability of Marriage. *The American Mathematical Monthly* 69, 1 (1962), 9–15.

[13] Can Gumeli, Angela Dai, and Matthias Niebner. 2022. ROCA: Robust CAD Model Retrieval and Alignment from a Single Image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 4012–4021.

[14] Yuan He, Yilun Zheng, Meng Jin, Songzhen Yang, Xiaolong Zheng, and Yunhao Liu. 2019. Red: Rfid-based Eccentricity Detection for High-speed Rotating Machinery. *IEEE Transactions on Mobile Computing* 20, 4 (2019), 1590–1601.

[15] Berthold KP Horn. 1987. Closed-form Solution of Absolute Orientation Using Unit Quaternions. *Journal of the Optical Society of America A* 4, 4 (1987), 629–642.

[16] ImpinJ. 2010. Speedway Revolution Reader Application Note: Low Level User Data Support. (2010).

[17] Robert W Irving. 1985. An Efficient Algorithm for the "stable roommates" Problem. *Journal of Algorithms* 6, 4 (1985), 577–595.

[18] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, et al. 2011. Kinectfusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera. In *Proceedings of the ACM symposium on User interface software and technology*. 559–568.

[19] Chengkun Jiang, Yuan He, Songzhen Yang, Junchen Guo, and Yunhao Liu. 2019. 3D-OmniTrack: 3D Tracking with COTS RFID Systems. In *In Proceedings of the International Conference on Information Processing in Sensor Networks (IPSN)*. 290–301.

[20] Chengkun Jiang, Yuan He, Xiaolong Zheng, and Yunhao Liu. 2019. OmniTrack: Orientation-aware RFID Tracking with Centimeter-level Accuracy. *IEEE Transactions on Mobile Computing* 20, 2 (2019), 634–646.

[21] Haojian Jin, Zhijian Yang, Swarun Kumar, and Jason I Hong. 2018. Towards Wearable Everyday Body-frame Tracking Using Passive RFIDs. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (UBiComp)* 1, 4, 1–23.

[22] Kiran Joshi, Steven Hong, and Sachin Katti. 2013. PinPoint: Localizing Interfering Radios. In *Proceedings of the USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 241–253.

[23] Wolfgang Kabsch. 1976. A Solution for the Best Rotation to Relate two Sets of Vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* 32, 5 (1976), 922–923.

[24] Rainer Kronberger, Thomas Knie, Roberto Leonardi, Uwe Dettmar, Markus Cremer, and Salah Azzouzi. 2011. UHF RFID Localization System Based on a Phased Array Antenna. In *Proceedings of the IEEE International Symposium on Antennas and Propagation (APSURSI)*. 525–528.

[25] Qiongzheng Lin, Lei Yang, Yuxin Sun, Tianci Liu, Xiang-Yang Li, and Yunhao Liu. 2015. Beyond One-dollar Mouse: A battery-free Device for 3D Human-computer Interaction via RFID tags. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*. 1661–1669.

[26] Feng Liu and Xiaoming Liu. 2021. Voxel-based 3D Detection and Reconstruction of Multiple Objects from a Single Image. *Advances in Neural Information Processing Systems(NeurIPS)* 34, 2413–2426.

[27] Jia Liu, Min Chen, Shigang Chen, Qingfeng Pan, and Lijun Chen. 2017. Tag-Compass: Determining the Spatial Direction of an Object with Small Dimensions. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*. 1–9.

[28] Tingjun Liu, Chuyu Wang, Lei Xie, Jingyi Ning, Tie Qiu, Fu Xiao, and Sanglu Lu. 2022. RF-Protractor: Non-Contacting Angle Tracking via COTS RFID in Industrial IoT Environment. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*. 390–399.

[29] Robert Miesen, Fabian Kirsch, and Martin Vossiek. 2013. UHF RFID Localization Based on Synthetic Apertures. *IEEE Transactions on Automation Science and Engineering* 10, 3 (2013), 807–815.

[30] Andrea Motroni, Paolo Nepa, Valerio Magnago, Alice Buffi, Bernardo Tellini, Daniele Fontanelli, and David Macii. 2018. SAR-based Indoor Localization of UHF-RFID Tags via Mobile Robot. In *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. 1–8.

[31] Arsalan Mousavian, Dragomir Anguelov, John Flynn, and Jana Kosecka. 2017. 3D Bounding Box Estimation Using Deep Learning and Geometry. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 7074–7082.

[32] Lionel M Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek P Patil. 2003. LANDMARC: Indoor Location Sensing Using Active RFID. In *In Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 407–415.

[33] Lionel M Ni, Dian Zhang, and Michael R Souryal. 2011. RFID-based Localization and Tracking Technologies. *IEEE Wireless Communications* 18, 2 (2011), 45–51.

[34] Yinyu Nie, Xiaoguang Han, Shihui Guo, Yujian Zheng, Jian Chang, and Jian Jun Zhang. 2020. Total3DUnderstanding: Joint Layout, Object Pose and Mesh Reconstruction for Indoor Scenes From a Single Image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 52–61.

[35] Stefan Georgiev Popov, Pablo Bauszat, and Vittorio Ferrari. 2020. CoReNet: Coherent 3D Scene Reconstruction from a Single RGB Image. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 366–383.

[36] Longfei Shangguan, Zhenjiang Li, Zheng Yang, Mo Li, Yunhao Liu, and Jinsong Han. 2013. OTrack: Towards Order Tracking for Tags in Mobile RFID Systems. *IEEE Transactions on Parallel and Distributed Systems* 25, 8 (2013), 2114–2125.

[37] Ken Shoemake. 1985. *Euler Angle Representation of 3D Rotations*.

[38] Jannis Strecker, Khakim Akhunov, Federico Carbone, Kimberly Garcia, Kenan Bektaş, Andres Gomez, Simon Mayer, and Kasim Sinan Yildirim. 2023. MR Object Identification and Interaction: Fusing Object Situation Information from Heterogeneous Sources. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (UBiComp)* 7, 3, 1–26.

[39] Maxim Tatarchenko, Stephan R Richter, René Ranftl, Zhuwen Li, Vladlen Koltun, and Thomas Brox. 2019. What Do Single-view 3D Reconstruction Networks Learn?. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 3405–3414.

[40] Chuyu Wang, Jian Liu, Yingying Chen, Lei Xie, Hong Bo Liu, and Sanclu Lu. 2018. RF-Kinect: A wearable RFID-based Approach Towards 3D Body Movement Tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (UBiComp)* 2, 1, 1–28.

[41] Chong Wang, Hongyi Wu, and N-F Tzeng. 2007. RFID-based 3-D Positioning Schemes. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*. 1235–1243.

[42] Jue Wang, Fadel Adib, Ross Knepper, Dina Katabi, and Daniela Rus. 2013. RF-Compass: Robot Object Manipulation Using RFIDs. In *Proceedings of the ACM International Conference on Mobile Computing and Networking (MobiCom)*. 3–14.

[43] Jue Wang and Dina Katabi. 2013. Dude, Where's My Card? RFID Positioning That Works with Multipath and Non-Line of Sight. In *Proceedings of the ACM SIGCOMM Conference (SIGCOMM)*. 51–62.

[44] Teng Wei and Xinyu Zhang. 2016. Gyro in the Air: Tracking 3D Orientation of Batteryless Internet-of-Things. In *Proceedings of the ACM International Conference on Mobile Computing and Networking (MobiCom)*. 55–68.

[45] Lei Yang, Yekui Chen, Xiang-Yang Li, Chaowei Xiao, Mo Li, and Yunhao Liu. 2014. Tagoram: Real-time Tracking of Mobile RFID Tags to High Precision Using COTS Devices. In *Proceedings of the ACM International Conference on Mobile Computing and Networking (MobiCom)*. 237–248.

[46] Songzhen Yang, Meng Jin, Yuan He, and Yunhao Liu. 2021. RF-Prism: Versatile RFID-based Sensing through Phase Disentangling. In *Proceedings of the IEEE International Conference on Distributed Computing Systems (ICDCS)*. 1053–1063.