



# Tamera: Contactless Commodity Tracking, Material and Shopping Behavior Recognition Using COTS RFIDs

FEI SHANG and PANLONG YANG, School of Computer Science and Technology, University of Science and Technology of China; CAS Key Laboratory of Wireless-Optical Communications, China  
JIE XIONG, University of Massachusetts Amherst, USA

YUANHAO FENG and XIANGYANG LI, School of Computer Science and Technology, University of Science and Technology of China; CAS Key Laboratory of Wireless-Optical Communications, China

RFID technology has recently been exploited for not only identification but also fine-grained trajectory tracking and gesture recognition. While contact-based (a tag is attached to the target of interest) sensing has achieved promising results, contactless sensing still faces severe challenges such as low accuracy and inability to sense multiple targets simultaneously in proximity, restricting its applicability in real-world deployment. In this work, we present **Tamera**, a contactless RFID-based sensing system, which significantly improves the tracking accuracy, enables multi-commodity tracking, and even material and shopping behavior recognition. We successfully address multiple technical challenges, and design and implement our prototype on commodity RFID devices. We test the positioning accuracy of **Tamera** in a 5 m × 6 m laboratory. **Tamera** achieves a median error of 1.3 cm and 2.7 cm for contactless single- and multi-commodity tracking, respectively. In our laboratory, two shelves commonly found in the supermarket are arranged and the goods are placed on them. **Tamera** successfully localizes and identifies the material type (metal, plastic, paper, and glass) of the commodities on the shelf with an accuracy higher than 95%. **Tamera** successfully recognizes four shopping behaviors (taking commodity, replacing commodity, buying commodity, and invoking commodity) with an accuracy higher than 93%.

CCS Concepts: • Human-centered computing → Ubiquitous and mobile computing design and evaluation methods;

Additional Key Words and Phrases: RFID, contactless, tracking, sensing, material, shopping behavior

43

## ACM Reference format:

Fei Shang, Panlong Yang, Jie Xiong, Yuanhao Feng, and Xiangyang Li. 2023. Tamera: Contactless Commodity Tracking, Material and Shopping Behavior Recognition Using COTS RFIDs. *ACM Trans. Sen. Netw.* 19, 2, Article 43 (February 2023), 24 pages.

<https://doi.org/10.1145/3563777>

The research is partially supported by NSFC with No. 62072424, U20A20181; the University Synergy Innovation Program of Anhui Province with No. GXXT-2019-024; National Key R&D Program of China 2018YFB0803400; National Key R&D Program of China under Grant No. 2021ZD0110400; China National Natural Science Foundation with No. 62132018; and Key Research Program of Frontier Sciences, CAS. No. QYZDY-SSW-JSC002.

Authors' addresses: F. Shang, P. Yang (corresponding author), Y. Feng, and X. Li, School of Computer Science and Technology, University of Science and Technology of China; CAS Key Laboratory of Wireless-Optical Communications, No. 96 Jinzhai Road, Hefei, Anhui, China, 230027; emails: shf\_1998@outlook.com, plyang@ustc.edu.cn; J. Xiong, University of Massachusetts Amherst, 374 Whitmore Building, 181 Presidents Drive, Amherst, MA 01003; email: jxiong@cs.umass.edu. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2023 Association for Computing Machinery.

1550-4859/2023/02-ART43 \$15.00

<https://doi.org/10.1145/3563777>

## 1 INTRODUCTION

Wireless sensing has attracted considerable attention in the last decade. Among the wireless technologies [10, 11, 19] that are employed for sensing, RFID is particularly promising due to the flexibility and cheap price of tags. The RFID reader is relatively expensive but one reader can communicate with many tags so the cost is well amortized. We have witnessed widespread RFID-based applications such as identification [3, 39], authentication [7, 12], tracking [8, 35, 40, 42], and counting [5, 13, 38]. Recent breakthrough research in this area has further enabled new and novel applications such as respiration sensing [28, 43], gesture recognition [25, 29], and even material sensing [33, 38]. These applications demonstrate a great potential of applying RFID devices for sensing besides the traditional identification usage.

In advanced tracking and localization application, RFID-based technologies have achieved centimeter-level accuracy recently by leveraging the fine-grained phase readings of the reflected signal. However, these high-accuracy approaches are contact based which means RFID tags need to be attached to the target commodity. Although an RFID tag is relatively cheap (~10 cents), attaching one tag to each commodity still incurs a significant amount of cost when there are millions of items such as in a huge supermarket or warehouse. Further, the labor cost of attaching and removing the tag to/from the item is also non-negligible.

Previous researches have shown that the RFID tags attached to the target can be utilized for target tracking. For example, Tagoram [40] utilizes moving tags to simulate the inverse **Synthetic Aperture Radar (SAR)** antennas and pinpoints a tag at an accuracy of a few centimeters. RF-IDraw [32] tracks a moving tag and deduces its trajectory in the air. RF-Echo [4] introduces a custom-designed active ASIC tag and achieves decimeter accuracy in the long-range indoor **non-line-of-sight (NLOS)** scenarios. But they all need to place an RFID tag on the target. In some applications, it is not always feasible to attach a tag to the target as the tag may affect the operation of a delicate mechanical/electronic device.

Inspired by these practical application requirements, in this work we focus on contactless sensing with commodity RFID hardware. The key idea behind contactless tracking is to retrieve useful information about the signal reflected on the target.

To enable fine-grained contactless sensing with RFID, multiple challenges need to be carefully addressed. First, the signal is weak after two reflections. Thus, the variation of reflected RFID signal induced by the commodity movement can be extremely small and easily buried in noise. Second, the disturbance of the environment is time-varying. Taking the supermarket environment as one example shown in Figure 1, customers walking around can interfere with the sensing. Furthermore, the environment in which the RFID tags and antennas are deployed is dynamically changing. Third, it is difficult to distinguish the target signal from the reflected signal of surrounding items. Without attaching tags to the targets of interest, the received signals are reflected by these commodities. This adds another layer of difficulty in differentiating these targets. In addition to performing contactless tracking of multiple commodities, we further propose to sense the material of the commodity.

To address the aforementioned challenges and extract subtle signal variation for sensing/tracking, we differentiate the signals received at different stages (before, during, after the target is picked up): employ the difference of these signals to remove the commonly shared strong direct-path signal, background reflections, and hardware noise in the three stages. Inspired by Tagoram [40], we further employ the hologram scheme to eliminate the effect of white noise. To address the interference caused by people moving around in the supermarket application, we combine information from multiple tags to distinguish the signal induced by target movement from the interference signal. The key observation is that the interference caused by the movement



Fig. 1. Supermarket scenario: Commodity tracking and understanding shopping behavior recognition.

of customers drastically depends on how far they are from the RFID tags. We observe that the commodities on the shelf are much closer (tens of centimeters) to the tags attached on the back of the shelf and thus the effect of the commodity movement on different tags can be dramatically different because a 10–20 cm distance difference among tags matters. Notice that the interfering customers are usually relatively far away (1–3 m) from the tags and thus the interference signal tends to be similar at all the tags. Based on this observation, we can filter out the interference caused by surrounding humans and track the targeted commodity. Since different commodity packaging materials have different dielectric properties, even if different commodities are placed in the same position, their effects on the RF signal are different. Therefore, when different commodities traverse the same track, the RF signals obtained are also different. Based on this difference, we have completed the identification of the product material. In order to avoid the influence of environmental noise, we choose to extract material features from the differential signal.

In this work, we carefully design **Tamera**, a contactless RFID-based multi-commodities tracking system. The major contributions of our work are summarized as follows.

- In terms of tracking, we propose a contactless, centimeter-level multi-commodity tracking solution. Taking advantage of the Gaussian nature of noise distribution, we propose an algorithm to assign weights to different labels (in Section 4.1), which achieves simultaneous tracking of multiple items with an average error of less than 3 cm. Furthermore, we combine geometric relations and gradient descent algorithms to improve the real-time performance of the system. Experimental results (in Section 7.2) show that the computation delays for single and multi-objectives are 0.01 s and 0.08 s, respectively.
- In terms of material identification, based on the model analysis, we construct features related to the permittivity of materials (in Section 5.3). Furthermore, we propose an algorithm to suppress human interference based on experimental observations (in Section 5.2). The experimental results (in Section 7.3) show that when there are three people in the environment, the accuracy of material recognition still exceeds 75%.

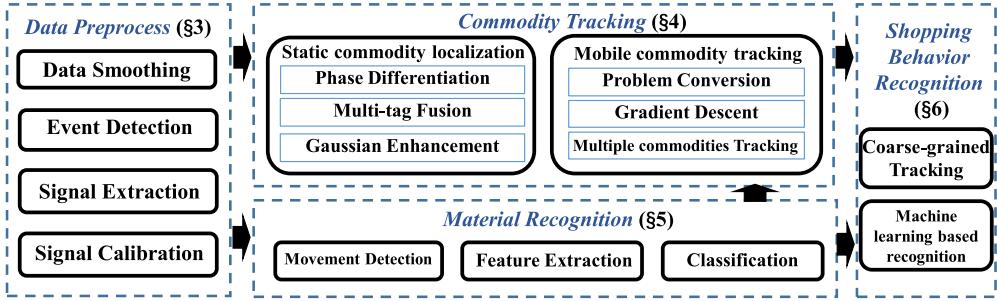


Fig. 2. System overview of **Tamera**.

—In terms of behavior identification, we build a lightweight classification model (in Section 6.2) to complete user behavior recognition. Our extensive studies show that **Tamera** can successfully infer the shopping behaviors of different customers at an accuracy higher than 93%.

The rest of this article is organized as follows. In Section 2, we present the overall structure of **Tamera**. In Section 3, we present our data preprocessing and calibration methods, and in Section 4 we present our contactless localization and tracking of multiple commodities simultaneously. We then introduce material recognition in Section 5 and shopping behavior recognition in Section 6. We present our extensive experiments to evaluate and validate our system in Section 7. We review the related work in Section 8 followed by a summary in Section 9.

## 2 SYSTEM OVERVIEW

Figure 2 shows the system architecture of **Tamera**. It consists of five main modules: (1) data preprocessing module, (2) localization module for static commodities, (3) tracking module for mobile commodities, (4) material recognition module, and (5) shopping behavior recognition module.

In the remaining sections, we will present these three key components of **Tamera** in detail.

Due to the environmental thermal noise and hardware imperfect synchronization, careful data preprocessing is required before we can achieve meaningful sensing. The processed data is passed to the signal extraction sub-module to extract the signal reflected from the target commodity. After that, we calibrate the signal by signal calibration sub-module to remove the random phase offset. The RSSI readings are also calibrated using a fitting method. Finally, the processed data will be used for accurate target tracking/localization, material recognition, and even shopping behavior recognition.

The localization and tracking modules are used to locate the absolute position of static commodities and the movement trajectory of mobile commodities. The material identification module is used to recognize the material type of the target when a commodity is moved by a customer. Localizing/tracking the items and identifying the packaging material with high precision are particularly important to infer which commodities are shopped by the customers. To enable high-precision tracking and recognition, we design an effective method for detecting whether a target item is moved by a customer. When it is moving, our system then triggers the tracking and material recognition methods to compute its initial position, the moving trajectory, and the material type of the item.

Finally, **Tamera** combines the material type, absolute position, and coarse-grained trajectory of the target movement to infer the customer’s shopping behavior. In this work, we predefine four different shopping behaviors for the demonstration. Our methods can be extended to recognize more shopping patterns if their trajectories differ from each other to some extent.

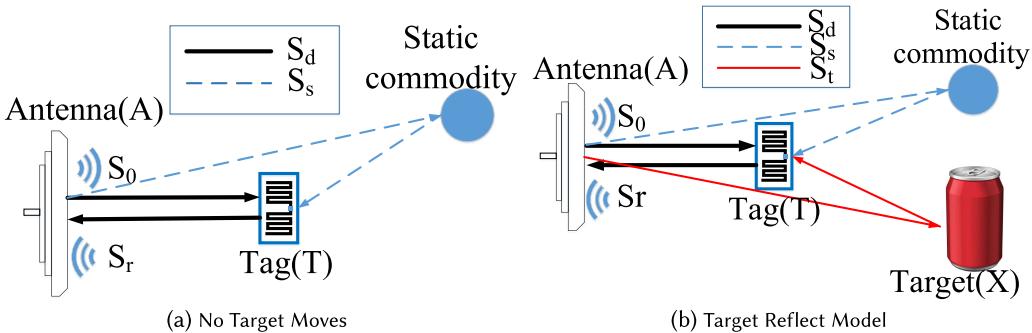


Fig. 3. Reflection model for RFID system.

### 3 DATA PREPROCESSING

### 3.1 Phase Data Preprocessing

**Signal acquisition.** Since we are using commercial RFID devices, we cannot obtain the original I/Q signals. Instead, we can only obtain the data (tag id, amplitude, phase, timestamp, etc.) returned by the reader according to the protocol. We convert RSSI into amplitude, and then combine the phase data to construct a complex number as the signal value.

**Smoothing.** We notice that even if the tags and the antennas are static, the phase and RSSI data will still fluctuate slightly in a short time. In addition to the small variation, the phase readings occasionally suffer from phase ambiguity, which introduces a sudden phase increase or decrease by  $\pi$ . For this significant phase change, we mitigate the  $\pi$  phase ambiguity from the readings. Then we employ a moving average filter to smooth the phase and RSSI variation for further processing.

**Interpolation.** RFID readers can only communicate with one tag at a time, which means sampling data for tracking is limited. On the other hand, we need to capture the data from multiple tags at the same time in order to track multiple moving targets. We indeed interpolate the available sampling data at the required time points. We choose the **Piecewise Cubic Hermite Interpolating Polynomial (PCHIP)** method which provides more accurate interpolation with fairly high efficiency.

### 3.2 Accurate Signal Extraction

**Raw Signal Extraction.** Figure 3 shows the single tag signal propagation model with RFID to illustrate how the target movement can be tracked with reflected signal variations. This concept can be easily generalized to multiple tags.

Before the tag-free target enters into the tracking area, the received signal contains the background signal components, including the direct path signal ( $s_d$ ) and other static reflection path signals ( $s_s$ ). We record the stationary signal at this time and the whole signal can be expressed as  $s_{env} = s_d + s_s$ . When the target enters the tracking area, the signals reflected from the target X ( $s_X$ ) will also be received. And the signal received at the RFID reader can be expressed as  $s_r$  [20]:

$$s_r = s_d + s_s + s_X. \quad (1)$$

To extract the target reflection signal, we subtract the signal when the target is not present from  $s_r$  in Equation (1):

$$s_X = s_r - (s_s + s_d) = s_{A \rightarrow X \rightarrow T \rightarrow A}, \quad (2)$$

where  $s_{A \rightarrow X \rightarrow T \rightarrow A}$  denotes the signal through a path from the antenna  $A$  to the commodity  $X$ , then to the tag  $T$ , and finally received at  $A$ .

**Propagation Signal Extraction.** The channel parameter is denoted by  $h = \alpha e^{-j\theta}$ , where  $\alpha$  and  $\theta$  represent the channel attenuation and phase shifting, respectively. In a commodity RFID system, we can obtain  $\alpha$  and  $\theta$  from the RSSI and phase readings.

We get the target-reflected signal  $S_X$  which can be further expressed as

$$S_X = s_A h_X h_{A \rightarrow X \rightarrow T \rightarrow A} * H_T, \quad (3)$$

where  $*$  indicates that the matrix is multiplied by element, and  $s_A$  represents the original signal from the antenna.  $H_{A \rightarrow X \rightarrow T \rightarrow A} = [h_{A \rightarrow X \rightarrow T_1 \rightarrow A}, \dots, h_{A \rightarrow X \rightarrow T_n \rightarrow A}]$  represents the signal change due to signal propagation in the air, and  $T_i$  is the  $i$ -th Tag. And the channel parameters  $h_X$  and  $h_A$  represent the signal change caused by the target surface and reader antenna, respectively.  $H_T = [h_{T1}, h_{T2}, \dots, h_{Tn}]$  is the influence of the tag array on the signal, and the subscript  $i$  corresponds to the  $i$ -th tag. And we name the phase shifting caused by these parameters as device-dependent phase shifting, which prevents us from obtaining location information of commodities.

**Eliminating Device-Dependent Phase Shifting.** To eliminate the device-dependent phase shifting, we apply a reference point model. Specifically, we place a target  $X_0$  at a randomly selected location which is called a reference point in the tracking space and measure the reflected signal following procedures presented in Section 3.2. Since we already know the location of the reference point, the propagation parameter  $h_{A \rightarrow X \rightarrow T_i \rightarrow A}$  can be calculated according to the following formula [24, 37]:

$$\begin{aligned} h_{A \rightarrow X_0 \rightarrow T_i \rightarrow A} &= c_{A \rightarrow X_0 \rightarrow T \rightarrow A} \cdot e^{-j\theta_{A \rightarrow X_0 \rightarrow T_i \rightarrow A}}, \\ \theta_{A \rightarrow X_0 \rightarrow T_i \rightarrow A} &= 2\pi \cdot \frac{r_{A \rightarrow X_0 \rightarrow T_i \rightarrow A}}{\lambda} \mod 2\pi, \end{aligned} \quad (4)$$

where  $c_{A \rightarrow X_0 \rightarrow T_i \rightarrow A}$  is the path loss factor along path  $A \rightarrow X_0 \rightarrow T_i \rightarrow A$ . The sum of the Euclidean distance of path  $A \rightarrow X_0 \rightarrow T_i \rightarrow A$  is  $r_{A \rightarrow X_0 \rightarrow T_i \rightarrow A}$ . The phase change due to signal propagation through path  $A \rightarrow X_0 \rightarrow T_i \rightarrow A$  is  $\theta_{A \rightarrow X_0 \rightarrow T_i \rightarrow A}$ . Let  $h_{X_0}$  denote the phase shift of target  $X_0$ , then the total device-dependent parameter  $s_A h_{X_0} h_A H_T$  can be calculated by  $S_X ./ H_{A \rightarrow X_0 \rightarrow T \rightarrow A}$ , where  $./$  is divided by the element. By eliminating the device-dependent phase shift calculated in the reference point, for any position in the measurement area, the channel parameters that are only related to the position of the commodity can be obtained by  $S_X ./ (s_A h_X \cdot h_A \cdot H_T)$ , and we name the phase of the channel parameter measurement as  $\Theta_m = [\theta_{m1}, \dots, \theta_{mn}]$ .

## 4 LOCALIZING AND TRACKING

We first present our scheme for contactless tracking commodities at high accuracy with COTS RFID devices. Due to the weak intensity of the RFID reflected signal, our measurement area is a square area of  $1.6 \text{ m} \times 1.6 \text{ m}$  surrounded by two tag arrays, each consisting of seven independent tags.<sup>1</sup>

### 4.1 Locate Static Commodities

**Basic Model.** Considering a commodity  $X$ , Equation (4) indicates the phase when the signal is reflected by the commodity. The sum of the distances from antenna  $A$  to commodity  $X$  and then to the  $i$ -th tag  $T_i$  is  $r_{A \rightarrow X \rightarrow T_i}$ . For one tag, the possible locations of the commodity  $X$  are on an ellipse [21] whose two focus points are the locations of antenna  $A$  and tag  $T$  as shown in Figure 4(a). For another tag, the possible locations of the commodity are on another ellipse. The intersection of ellipses produced based on info from different RFID tags is used to localize the target.

<sup>1</sup>We use seven tags to form a tag array because if the number is too small, such as three to four, it would be difficult to efficiently get sufficient data for processing; while if the number is far greater than 7, say 15–20, it would suffer from the heavy contention.

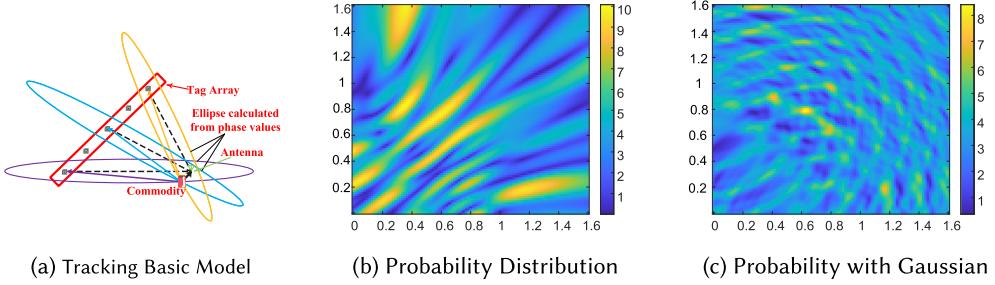


Fig. 4. Basic tracking model with Gaussian enhancement and prediction direction.

Unfortunately, in reality, it is non-trivial to obtain the position of the target by identifying the intersection of the two ellipses. First, for a phase  $\phi$  we calculate, there are multiple ellipses which satisfy the phase constraint  $\phi + k * 2\pi$  with  $k \in \{0, 1, 2, \dots\}$  being a non-negative integer. The second reason is that the intersection point of each tag pair may slightly deviate from the true position due to noise and errors, which causes the intersection points of multiple tag pairs not to coincide. If we solve multiple elliptic equations for the target's location, not only is the computational complexity high, but we also may not get the optimal solution.

**Phase Differentiation.** We thus divide the monitoring area into small grids with a side length of 0.1 cm (larger grids will bring more error and smaller grids will bring greater computing overhead). Every possible grid of the target is traversed to calculate the likelihood that the target appears.

When the target appears in the measurement area, we calculate the measurement phase  $\Theta_m$  of the propagation parameter by the method in Section 3.2 continuously. On the other hand, for each grid, if the target appears here, the reflected propagation distance is measurable, which can be expressed as  $R_{A \rightarrow X \rightarrow T \rightarrow A} = [r_{A \rightarrow X \rightarrow T_1 \rightarrow A}, \dots, r_{A \rightarrow X \rightarrow T_n \rightarrow A}]$ . The estimated phase can be given by

$$\Theta_e = \frac{2\pi}{\lambda}(R_{A \rightarrow X \rightarrow T \rightarrow A}), \quad (5)$$

where  $\lambda$  is the wavelength. For each tag, if the measured phase  $\theta_{mi}$  is closer to the estimated phase  $\theta_{ei}$ , we will regard that the target lies in this grid with high probability.

**Multi-Tag Fusion.** Since we have  $n$  deployed tags, after obtaining the measured phase  $\theta_{mi}$  and the estimated phase  $\theta_{ei}$  for the  $i$ -th tag, we can now calculate the probability for each grid by taking into consideration all the tags according to the following equation:

$$v = \left| \sum_{i=1}^n \|w_i\| e^{j(\theta_{mi} - \theta_{ei})} \right|. \quad (6)$$

For each tag, the difference between the measured phase value and the estimated phase value is treated as the phase of a constructed signal. And the amplitude  $\|w_i\|$  from each tag is equal to 1 since each grid has the same weight. If the  $K^{th}$  grid is the true position of the target, the measured phase  $\Theta_m$  will match the estimated phase at this grid well enough. Thus, a small phase difference causes the probability value  $v$  here to be combined constructively. Figure 4(b) shows the result of multi-tag fusion, and the color of the pixel at the estimated position is significantly darker than other areas.

**Gaussian Enhancement.** Due to the existence of additive white Gaussian noise,  $\theta_{mi}$  is a Gaussian distribution random variable while  $\theta_{ei}$  is constant, so  $\theta_{mi} - \theta_{ei} \sim \mathcal{T}(0, \delta^2)$ . Thus, the probability

value of a grid from a tag can be calculated by

$$\|w_i\| = f(|\theta_{mi} - \theta_{ei}|; 0, \delta^2), \quad (7)$$

$$f(x) = \frac{1}{\sqrt{(2\pi)\delta}} \exp\left(-\frac{(x - \mu)^2}{2\delta^2}\right), \quad (8)$$

where  $f(x; 0, \delta^2)$  is the probability density function of a Gaussian distribution  $\mathcal{N}(0, \delta^2)$ . The standard deviation  $\delta$  is determined by the thermal noise and independent of the distance. In our experimental settings, when we set the value of  $\delta$  as 0.4, it could be optimal comparing with other values, which is determined according to the experimental results in at least 500 rounds.

Figure 4(c) depicts the result when incorporating the aforementioned Gaussian enhancement method into our model. We can clearly see that our model greatly reduces the estimated region with higher probability, which could significantly reduce the searching overhead and improve the tracking accuracy as well.

**Event Detection.** When the location of the commodities is changed, the collected signal at the antenna will be changed. Compared to the signal amplitude, the signal phase is more sensitive to such environmental change. After smoothing the phase curve, we apply a sliding window to continuously detect whether the commodity location has been changed. The variance of the phase value in the sliding window reflects the degree of phase fluctuation. When the variance exceeds our preset threshold, we think that the commodity's location is being changed. When the variance drops below the threshold again, we regard the commodity has been changed and recalculate the position of the commodity. Specifically, we use a double sliding window for event detection, which is similar to WiFi's packet inspection mechanism [27]. First, we compute the signal variances within two adjacent time windows A and B of equal length, which are  $a_n$  and  $b_n$ . Then we set a threshold  $T_h$  to judge whether the ratio  $m_n$  of  $a_n$  and  $b_n$  has jumped. In the specific setting, the threshold is set to 10.

**Search Space Optimization.** In order to determine the location of the commodity  $X$ , for all grids in the area, we calculate the probability of the commodity at that grid one by one. However, such a global search method is likely to cause a lot of computational overhead, especially when the search area is large or the side length of the grid is small. As shown in Figure 4(a), the commodity  $X$  is on an ellipse with antenna  $A$  and tag  $T$  as the focus points in theory. This provides us with an opportunity to optimize the search space. Under ideal circumstances, the commodity  $X$  should satisfy the following equation:

$$\begin{aligned} & \sqrt{(x - X_A)^2 + (y - Y_A)^2} + \sqrt{(x - X_T)^2 + (y - Y_T)^2} \\ &= R_{A \rightarrow X \rightarrow T \rightarrow A} - \sqrt{(X_A - X_T)^2 + (Y_A - Y_T)^2}, \end{aligned} \quad (9)$$

where  $(x, y)$  is the coordinates of the commodity  $X$ ,  $\{(X_A, Y_A) | (x_{a1}, y_{a1}), \dots, (x_{an}, y_{an})\}$  is the set of antenna coordinates, and  $\{(X_T, Y_T) | (x_{t1}, y_{t1}), \dots, (x_{tm}, y_{tm})\}$  is the set of tag coordinates. On the left side of the equal sign is an equation about  $x$  and  $y$ , We let

$$\begin{aligned} g(x, y) &= \sqrt{(x - X_A)^2 + (y - Y_A)^2} \\ &+ \sqrt{(x - X_T)^2 + (y - Y_T)^2} + \sqrt{(X_A - X_T)^2 + (Y_A - Y_T)^2}. \end{aligned} \quad (10)$$

Combining Equations (5), (9) can be written as follows:

$$g(x, y) = \left(k + \frac{\theta_x}{2\pi}\right) \lambda, \quad (11)$$

where  $\theta_x$  is the measured phase,  $k$  is a non-negative integer, and  $\lambda$  is wavelength.

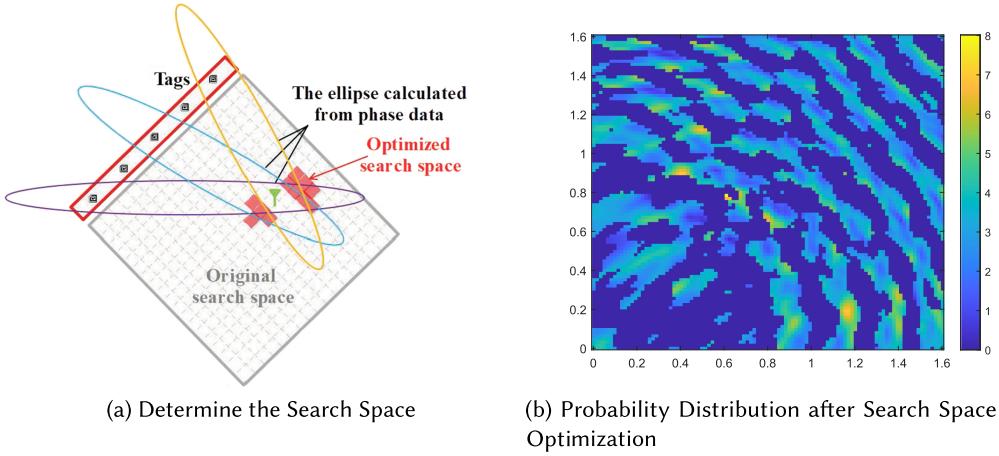


Fig. 5. Search space optimization.

However, due to multi-path interference and other reasons, there may not be a point where all ellipses pass. Therefore, when determining the search space, we do not strictly require that the points in the area appear on all ellipses, but require that the distance from them to each ellipse does not exceed a threshold. As shown in Figure 5(a), the area covered by gray is very close to each ellipse, and we select these areas as the search space. We modify Equation (11) to the following form:

$$\left| \frac{g(x, y)}{\lambda} - \left\lfloor \frac{g(x, y)}{\lambda} \right\rfloor - \frac{\theta_x}{2\pi} \right| < \theta_d, \quad (12)$$

where  $\lfloor x \rfloor$  represents the largest integer not greater than  $x$  and  $\theta_d$  is the threshold. According to Equation (12), we optimize the search space. If we simply use the traversal algorithm, the complexity of the algorithm is  $O(N \times M)$ , where  $N$  and  $M$  are the horizontal and vertical grid points in the area. By optimizing the search space, the complexity of the algorithm can be reduced to  $O(\frac{N \times M}{k})$ , where  $k$  is a constant related to  $\theta_d$ . The probability distribution after optimization of the search space is shown in Figure 5(b). Since most areas are not searched, the probability value of these places maintains the initial value of 0.

## 4.2 Tracking Mobile Commodities

In this section, we further explore the tracking of the mobile target. The difficulty in tracking the mobile target is that COTS RFID hardware has a low sampling rate. When the target moves, the data samples collected at a target location become insufficient for accurate localization. To address this issue, the gradient descent method is used to estimate the position of the target at the next moment which can make the most of every sample. We present the details below. We first discuss the movement of a single commodity in the area.

**4.2.1 Single Commodity Tracking. Problem Conversion.** According to Equation (6), our goal is to find the grid with the largest  $v$ . However, in addition to the insufficient sampling rate, the method of dividing the area into grids needs to traverse the entire tracking area, which is time-consuming and cannot meet the requirement of real-time performance. The size of the grid also affects the positioning accuracy. Thus, considering the continuity of space, we translate the problem into an optimization problem of identifying the value of  $(x, y)$  to make the  $v$  value the largest. However, Figure 4(b) shows the value of the function over the entire tracking area from which we can see that the function is non-convex in the definition domain.

The difficulty of the non-convex optimization problem is that the local optimum point is not the global optimum point. Fortunately, we can solve the problem by utilizing the continuity property of target movement. In a short period of time (<0.1 s), the target usually does not move more than a half wavelength (<16 cm) of distance. Therefore, if we know the current position of the target, as long as the sampling interval is small, the local optimum can be considered as the true position of the target [2].

**Mini-Batch Gradient Descent.** **Tamera** employs the gradient descent method to find the optimal solution. However, not all the sample values can be used to calculate the gradient, because these samples are collected at different locations when the target is moving. We need to select samples collected at close-by target locations and make the algorithm converge faster than the target location change.

Because of the spatial continuity of target movement, adjacent samples in the time domain correspond to close-by target locations. On the other hand, the Mini-batch Gradient Descent method requires adjacent samples to be from random tags to ensure faster convergence. Fortunately, the RFID system adopts a random back-off mechanism and this inherent property guarantees the adjacent samples are exactly from random tags. Thus, we can directly use the data samples collected without the need of designing a sample selection algorithm.

**Function Simplification.** When determining the initial position of the commodity, the optimization function is the commodity of the probability distribution function and the complex exponential function. However, both of them are difficult to solve. We simplify the function in the moving target tracking module:

$$v' = \sum_{i=1}^n \cos(\theta_{mi} - \theta_{ei}), \quad (13)$$

where  $n$  is the number of sample points for each iteration.  $\theta_m$  is the measured phase of the reflection signal, and  $\theta_e$  is the estimated phase at the latest position  $(x, y)$ :

$$\theta_{ei} = \frac{2\pi}{\lambda} \cdot \left( \sqrt{(x - x_{Ti})^2 + (y - y_{Ti})^2} + \sqrt{(x - x_A)^2 + (y - y_A)^2} \right), \quad (14)$$

where  $(x_{Ti}, y_{Ti})$  and  $(x_A, y_A)$  denote the positions of the  $i$ -th tag and antenna, respectively. Then we calculate the partial derivatives of  $x$  and  $y$ , respectively:

$$\begin{aligned} \frac{\partial v'}{\partial x} &= \sum_i^n \frac{2\pi}{\lambda} \cdot \sin(\theta_{ei} - \theta_{mi}) \cdot \left( \frac{x - x_A}{dis_A} + \frac{x - x_{Ti}}{dis_{Ti}} \right), \\ \frac{\partial v'}{\partial y} &= \sum_i^n \frac{2\pi}{\lambda} \cdot \sin(\theta_{ei} - \theta_{mi}) \cdot \left( \frac{y - y_A}{dis_A} + \frac{y - y_{Ti}}{dis_{Ti}} \right), \end{aligned} \quad (15)$$

where  $dis_A$  and  $dis_{Ti}$  represent the distance from the target to the antenna and to the  $i$ -th tag, respectively. The sample values are fused to calculate the gradient, where the initial step size is set to 1 mm by default, which is the length of the coordinate changes after each update of the gradient. When the difference in the value of the function is less than a predefined threshold, we consider the localization algorithm has converged and then record the position of the target. If there is no convergence after a predefined period of time, we dynamically adjust the step size and the number of samples used in each iteration to increase the convergence speed.

**4.2.2 Multiple Commodities Tracking.** The difficulty in simultaneous positioning of multiple commodities is that the signals reflected from different targets are superimposed at the reader. RFID hardware usually has a small frequency bandwidth and the sampling rate is low. Thus, it is extremely difficult to separate the superimposed signals to track each individual target. We thus tackle this issue from another direction without requiring one to separate the mixed signals. The basic idea is that when there are  $l$  targets in the measurement area, the received superimposed signal can be calculated by the following formula:

$$\mathbf{S} = \sum_{i=1}^l \mathbf{C}_{\Pi_i} \cdot e^{-j \frac{2\pi}{\lambda} \mathbf{R}_{\Pi_i}}, \quad (16)$$

where matrix  $\mathbf{S}$  of size  $n \times m$  represents the superimposed reflected signal of  $l$  commodities at  $n$  antennas when receiving the signals of  $m$  tags. The signal propagation path for antennas and the  $i$ -th tag is  $\Pi_i$ . The path loss factor of paths  $\Pi_i$  when the signal propagates in the air is  $\mathbf{C}_{\Pi_i}$ , which is usually inversely proportional to distance. And,  $\mathbf{R}_{\Pi_i}$  is the distance of paths  $\Pi_i$ .

It's nontrivial to solve the aforementioned equations, since the inaccurate path-loss model as well as the measurement noise on RFID readers and tags. These errors make the results really far away from the ground truth, and sometimes there is even no feasible solution for this problem.

In dealing with the aforementioned challenges, we devise a virtual superimposition scheme, where  $l$  virtual nodes are selected to test the feasibility of the solution. Without loss of generality, we consider a two target scenario as one example. We traverse each pair of grids in the measurement area. For each pair of grids, we estimate the phase of their reflected signal. And then we can obtain the superimposed signal according to Equation (16).

Although the estimated phase for each grid can be calculated in advance, when the number of targets is more, there are more candidate grids to traverse and the computational complexity increases exponentially with the increased number of grids.

## 5 MATERIAL RECOGNITION

Many systems involve material analysis of commodities. However, these systems [33] are mostly limited to stationary commodities and require a specific container. In this section, we propose a method capable of material sensing even when the target is moving. More specifically, we employ the phase change caused by different materials when the commodity moves to identify the material type of the target.

### 5.1 Rationale of Material Recognition

**Reflection Model.** Figure 6(a) shows the basic principle of our material identification method. Different materials such as metal and plastic have different reflection intensities for wireless signals. The red line represents all stationary signals including the direct path signal ( $s_d$ ) of the RFID tag and the static signal ( $s_s$ ) reflected by the surrounding stationary commodities. The signal reflected by the target  $X$  ( $s_X$ ) is represented as a circle in the phase domain due to the movement.<sup>2</sup> The phase values are indicated by green and black lines for reflections on different materials. The three kinds of signals are mixed at the RFID receiver. When the target starts moving, the propagation distance of the reflected signal of the commodity has changed. The phase of the reflected signal  $s_X$  shifts with the distance.

The right part subfigures show the phase changes of the combined signals at the antenna. During this process, if the metal and plastic targets are moving in the same trajectory, due to the stronger

<sup>2</sup>The movement leads to the phase variation with time. Since the values are in the range of  $[0, 2\pi]$ , in phase domain, it could be drawn as a circle.

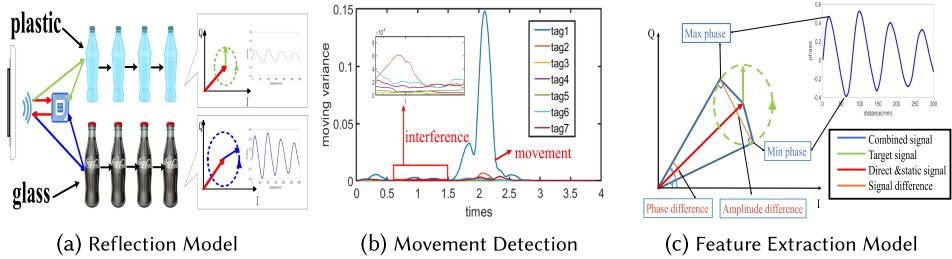


Fig. 6. Material recognition model.

reflectivity, the reflected signal from the metal surface, the phase change of the combined signal is significantly larger.

## 5.2 Movement Detection

Before we recognize the material of the moving targets, the change of signal caused by the target should be marked precisely. However, the frequency of moving targets, as well as walking, swinging, and other daily actions are all between 5 and 20 Hz and the sampling rate of a single tag of RFID is lower than 30 Hz, which makes it difficult to separate the target moving signal and surrounding noise in the frequency domain. In the time domain, different people take commodities at different speeds which results in different fluctuations and some gentle fluctuations are similar to ambient noise.

We detect the target movement based on the following observations. Relative to the difference in the position of different tags, the size of the person is so large that human interference has a similar effect on multiple tags. However, for the moving commodity, the relative position with different tags is significantly different so that the change of its position has a much larger effect on one or two tags. We calculate the moving variance of the RFID phase when the experimenter takes the target. The average moving variance of all tags is subtracted from the value of each tag. And then only one or two tags show significant peaks on the moving variance when the action of moving the item occurs as shown in Figure 6(b). We can thus employ this observation to differentiate interference and target-caused signal variations.

## 5.3 Feature Extraction

As mentioned earlier in Equation (1), the received signal contains the direct path signal ( $s_d$ ), static reflection path signal ( $s_s$ ), and the signals reflected from the target  $X$  ( $s_X$ ). Due to the dynamic changes of the environment, the static portion of the RFID signal  $s_s$  often changes unexpectedly, which will also affect the degree of fluctuation of the combined signal phase. As Figure 6(c) shows, the phase value fluctuation will change with the length of the red line, i.e., the combined amplitude of the direct signal and static signal.

We use a geometric model to calculate the amplitude of the target reflection signal from the commodity without knowing  $s_d$  and  $s_s$ . Assuming that a target moves slightly within half of the signal wavelength (i.e., about 16 cm), the phase shifting of the reflected signal would form a circle in the IQ coordinate system, as the green dotted line shows in Figure 6(c). When the phase of the received signal reaches the peak and valley value, we can think of the signal as the tangent to the circle as shown by the black lines. The magnitude of the difference ( $\Delta\alpha$ ) between the two signals can be regarded as a chord of the circle. And the phase of the difference ( $\Delta\phi$ ) between the two signals can be regarded as the complementary angle of the central angle corresponding to the

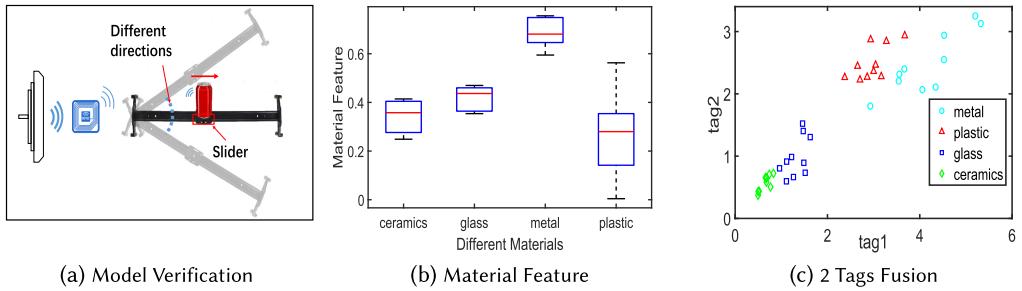


Fig. 7. Material recognition when commodities are moving in different directions, using different materials.

chord. These two values can now be used to determine the reflection amplitude feature (FAF) of the target according to  $f = \frac{\Delta\alpha}{\pi - \Delta\phi}$ . We use the reflection intensity as a feature of material identification.

#### 5.4 Classification

**Basic Observation.** We first validate our model with benchmark experiments. As shown in Figure 7(a), we use the track slider to move the commodity, and the track slider is placed at different orientations to demonstrate the effectiveness of the model. Figure 7(b) shows the extraction results of the four materials including metal, plastic, glass, and paper. There are clear differences in the reflected signals of different materials.

**Multi-Tag Fusion.** One may wonder, is this feature reasonable when commodities in different materials are moving at different distances? If a commodity with stronger reflectivity is moving at a farther distance in comparison with the one in material with weaker reflectivity, how does one discriminate these two commodities? In fact, for the feature of a single tag, as shown in Figure 7(b), there are indeed some overlaps between different materials because of the offset of the relative position of the target with respect to the antenna and tag. Therefore, we employ multiple tags and fuse the information from multiple tags to address the issues mentioned above. Figure 7(c) shows the result of the feature fusion of two tags. And the different materials can be artificially separated. In fact, in order to prevent the impact of signal errors or large position offset, we used more than seven tags and each tag is separated by 15 cm.

Based on the features extracted from tags, we try to classify different materials using the following three classifiers: Decision Tree, KNN (K-Nearest Neighbor), and SVM (Support Vector Machine with different kernels). For our dataset, the performance of these three classifiers is presented in Table 1. Based on accuracy and time cost, the decision tree is selected as the classifier for material identification. We set the split criterion to Gini's diversity index. And the number of leaves is limited to 20 to prevent overfitting.

## 6 BEHAVIOR RECOGNITION

In this section, we will introduce how **Tamera** recognizes customers' shopping behavior in a typical supermarket environment. Since we already obtained the tracking and material information as presented in the above sections, a straightforward method is to leverage the obtained trajectory information directly for behavior recognition. However, with this trajectory-based method it is often hard to achieve the expected performance due to the following reasons.

- First, due to the complicated placement of commodities on the shelves, estimate the true location of each commodity is a time-consuming task.

Table 1. Performance of Different Classifiers in Material Recognition

Classifier	Accuracy	Training Cost	Prediction Cost
KNN	84.6%	N/A	1.66 ms
SVM (Gaussian Kernel)	88.3%	2.07 s	5.4 ms
SVM (Linear Kernel)	78.3%	2.07 s	5.2 ms
SVM (Polynomial Kernel)	84.7%	2.02 s	5.5 ms
Decision Tree	96%	0.59 s	1.46 ms

- Second, the limited shelf space limits the number of deployed tags, which results in limited location performance.
- Third, when a customer takes the commodity, the trajectory is a complex 3D movement rather than a simple 2D movement, which makes it very difficult to judge customers' shopping behavior through trajectories, moreover, customers' different shopping habits, such as speed, distance, and other factors, also make this problem worse.

Next, we will describe how **Tamera** overcomes these challenges and achieves a high identification accuracy for customers' behavior.

## 6.1 Target Commodity Positioning

To recognize a customer's shopping behavior, **Tamera** first needs to know which commodity the customer has picked up. Therefore, the initial position of the picked commodity should be known in advance. An intuitive way is to use the method mentioned in Section 4.1 to directly locate the picked commodity. However, due to the rich multi-path effect caused by narrow shelf space and complicated commodity placement, it is a very difficult and time-consuming task to obtain the real position of the picked commodity. To this end, **Tamera** applies position recognition instead of estimate real location to reduce the time overhead.

The key observation behind the position recognition scheme is that the closer the commodity is to the tag, the greater the impact on the tag signal. Therefore, the movement of the commodity will obviously change the phase feature of the tag signal. Specifically, **Tamera** divides each shelf layer into several even zones. As shown in Figure 8(d), seven tags are linearly arranged behind the commodity to form a tag array. When a customer takes a commodity, **Tamera** collects the phase change caused by the commodity movement and calculates the difference between the peak value and valley value as the location feature. The position recognition result can be obtained by feeding such features into a KNN classifier. Note that the length of the shelf is 115 cm and the average size of the commodity (e.g., paper cup in this scenario) is usually 9 cm. For that reason, **Tamera** divides the shelf layer into  $\lfloor 115/9 \rfloor = 12$  zones, which is dense enough for commodity position recognition.

## 6.2 Customer Behavior Recognition

After identifying the picked commodity, **Tamera** can begin to recognize the customer's shopping behavior. An intuitive idea is to track the commodity's movement directly as we mentioned in Section 4, and then recognize customer's behavior. However, such method suffers high computational overhead and processing delay, which makes it infeasible. Fortunately, **Tamera** does not need to accurately track a customer's entire trajectory to identify their shopping behavior; in practice, it only needs to identify several basic shopping actions; **Tamera** can recognize most of the customer's shopping behaviors. In this article, **Tamera** predefines four basic shopping actions: (*PickUp*) pick the commodity up from the shelf; (*PutDown*) put the commodity down to the shelf; (*PutIn*) put the commodity in the cart; and (*PutOut*) remove the commodity from the cart. By combining these basic actions, **Tamera** can recognize the customer's shopping behavior. For example,

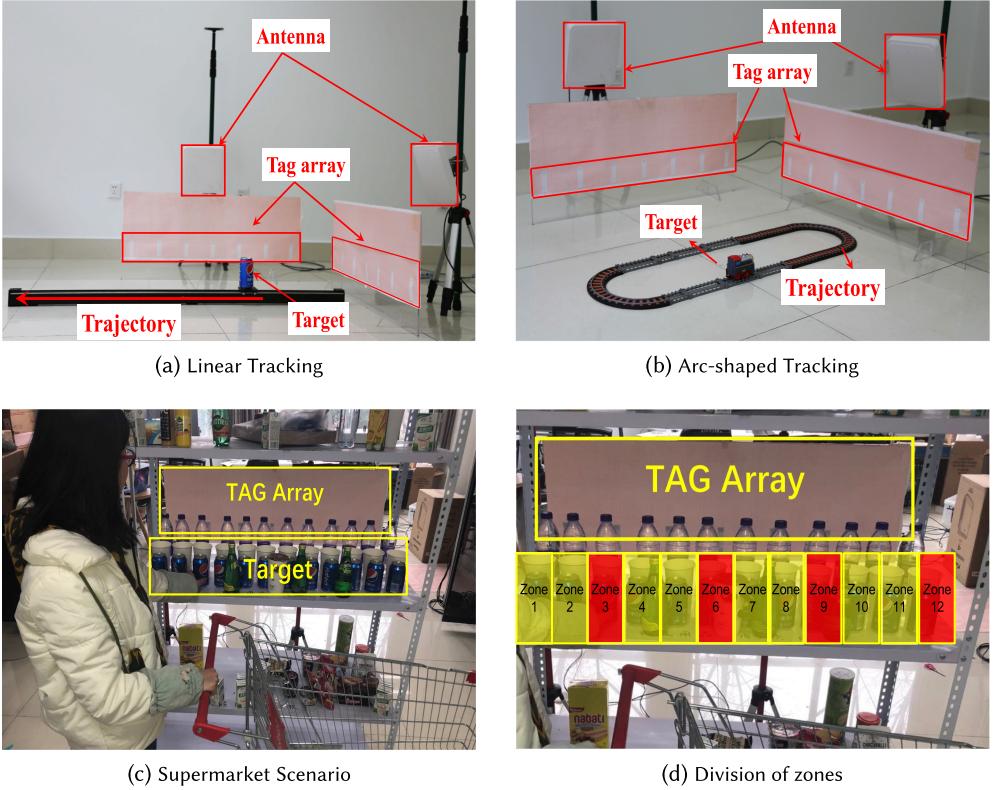


Fig. 8. Experiment environment and typical setup for evaluations.

if **Tamera** detects action *PickUp* and action *PutDown* that occur continuously, then it can be judged that this customer just picked up and checked the commodity, but has no intention to buy.

For each action, **Tamera** recognizes it by identifying its phase feature. Specifically, **Tamera** continuously records tag signals. When a commodity movement is detected, based on the recognition result of the initial position of this commodity, **Tamera** will subtract the corresponding relatively stationary signal from the received signal to eliminate multi-path in the environment and use the phase value of the resulted signal as the feature for action recognition. The whole phase values are re-divided into 80 discrete points by the interpolation method to remove the Doppler effect caused by taking the commodity.

However, we noticed that different users have different habits, so when they do the same action, the signal phase will be very different. Specifically, the time and duration of the peaks and troughs of each phase signal will be different due to the speed at which the user completes the action. This difference makes it difficult for classifiers such as KNN to achieve the required performance.

Therefore, before classifying the signal, we first use the **dynamic time warping (DTW)** [16] algorithm to standardize the data, so that the phase characteristic curves of the same action have a higher similarity. We test the effectiveness of the DTW algorithm on three classifiers, including KNN, SVM, and Decision tree. The results of the test are shown in Table 2. We find that after using the DTW algorithm to standardize the data, the accuracy of using KNN for behavior recognition reached 93.2%, and it only took 0.19 ms to classify each action. Therefore, we choose KNN as the classifier for behavior recognition.

Table 2. Performance of Different Classifiers in Behavior Recognition

Classifier	Accuracy without DTW	Accuracy with DTW	Prediction Cost
KNN	65.4%	93.2%	0.19 ms
SVM	59.3%	82.6%	1.14 ms
Decision Tree	51.1%	73.2%	0.26 ms

## 7 EVALUATION

In this section, we evaluate **Tamera** from aspects on tracking accuracy, material recognition, and behavior recognition. We first introduce the basic experiment settings. Then, different materials combining with the human interference are fully evaluated across different settings. Moreover, supermarket environments are fully respected, where user behaviors are evaluated with fairly complicated settings for multiple racks and multiple walking persons nearby.

### 7.1 Experiment Setup

We build a prototype of our system with commodity ImpinJ Speedway R420 reader and Alien EPC Gen2 UHF tags. The reader supports up to four antennas and is compatible with the EPC Gen2 standard. The RFID system operates in the 920–926 MHz band. The size of the antenna is 260 mm × 260 mm × 40 mm. The reader is connected to the server through an Ethernet cable. We adopt the timestamp function provided by the reader instead of the absolute timing information to eliminate the influence of network latency. We attach the tags on the acrylic board for the convenience of deployment as shown in Figure 8. The tags are evenly attached to the board. The dimension of the board is 100 cm × 40 cm. Adjacent tags are separated by 15 cm.

Due to the inherent anti-collision mechanism adopted by the commodity RFID reader, when the reader queries multiple tags, a counter mechanism is adopted to prevent the tag from replying very quickly [13]. This mechanism greatly affects the reader sampling (reading) rate. But if the counter parameter is set too low, collision will happen frequently so that the continuity of the experimental data will be destroyed and the sampling rate will be reduced. We thus modify the counter parameter to optimize the time the tag waits and successfully increase the sampling rate to about 40 Hz for one tag array with 14 tags. We use Ethernet cables to transfer RFID reader data to a computer with Intel(R) Core i5-8500 at 2.8 GHz and 8 G memory. Using the TCP protocol, the data obtained by the RFID reader is transmitted to the computer, and the data processing is completed in MATLAB 2018a to determine the target’s position and movement trajectory.

**Tracking Environment.** The tracking experiments are conducted in the lab environment with a size of 5 m × 6 m. Each of the two antennas is placed 70 cm away from the acrylic board. Since any trajectory can be decomposed into linear and arc-shaped trajectories, we mainly focus on these two basic trajectories. As shown in Figure 8(a), we place the target on an electric track with a length of 135.2 cm. For arc-shaped trajectory, we place a toy train running in an arc-shaped orbit as shown in Figure 8(b). The track consists of two linear parts with a length of 30 cm and two half circles with a diameter of 22 cm. We employ a camera to record the movements as ground-truths. The above toy track is also reassembled into various irregular shapes in the experiment.

**Recognition Environment.** To evaluate the shopping behavior recognition, we conduct comprehensive experiments with the supermarket shopping cart and shelves as shown in Figure 8(c). One tag array with less than 14 tags is deployed at the back of the shelf. The antennas of the reader are located 50 cm away from the tag array. The antennas and tag arrays are required to remain stationary, and commodities are placed densely on the shelves. A total of six people with two females and four males are involved in the experiment. We collect data when the volunteers perform shopping behaviors in their own manner. For material identification, we use

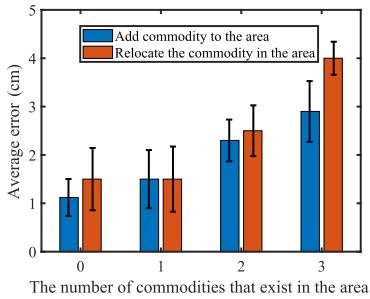


Fig. 9. Localization accuracy in different module.

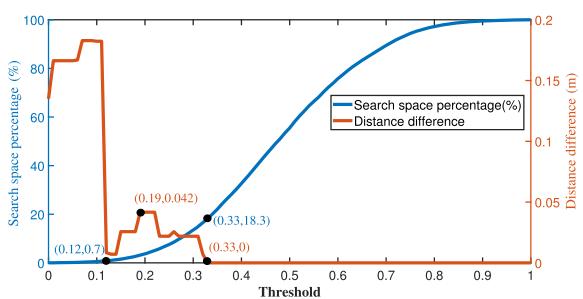


Fig. 10. The effect of search space optimization.

paper, metal, glass, and plastic materials to train our model. The shopping behaviors we evaluated include (1) taking a commodity from the shelf, (2) putting back the target to the shelf, (3) putting a commodity in the cart, and (4) taking the target out of the cart and putting it back to the shelf.

## 7.2 Tracking Accuracy

**Static Commodities.** We test the positioning accuracy in two scenarios: (1) add a new commodity to the positioning area, and then determine the location of the new commodity; and (2) move a commodity in the area to determine the position of the commodity after it has been moved. However, during daily use, there are often multiple commodities on the desktop. Therefore, we measure the position of multiple commodities one by one. The positioning accuracy is as shown in Figure 9. For the first scenario, the average of positioning error is within 3 cm. As the number of commodities increases, a small number of data are positioned incorrectly, resulting in some abnormal points, so the average error will increase from 1 cm to 3 cm. However, in the second scenario, environmental signals need to be synthesized artificially from the reflected signals of other commodities in the measuring area which will cause some deviation.

**Search Space Optimization.** When locating the target, if we traverse each grid in the area without restriction, it will bring greater computational overhead. Therefore, we constrain the search space based on the geometric relationship. Figure 10 shows the impact of the selection of the threshold on the search space percentage and distance difference when optimizing the search space. The threshold varies from zero to one, which corresponds to the restriction from strict to loose. When the threshold is zero, the search space contains only the intersection of several ellipses; when the threshold is one, the search space contains all grids in the area. Search space percentage indicates the ratio of the number of grids in the search space to the number of all grids in the area. The distance difference represents the distance between the points obtained by positioning before and after the search space optimization. It can be seen from the figure that when the restriction conditions are strict, the distance difference is relatively large. This is because the measurement error makes the intersection of the ellipses not accurately correspond to the position of the commodity. When the threshold is greater than 0.33, the distance difference is already 0 cm, and the search space at this time only accounts for 18.3% of the total number of grids. Even under the loose requirement that the maximum distance difference does not exceed 5 cm, the size of the search space can be optimized to 0.7% of the number of all grids.

**Mobile Commodities.** Figure 11 shows the comparison between the trajectory generated by Tamera and the actual trajectory when the target moves in a linear trajectory, a circular trajectory, and an irregular shaped trajectory. We also show the tracking results when two commodities

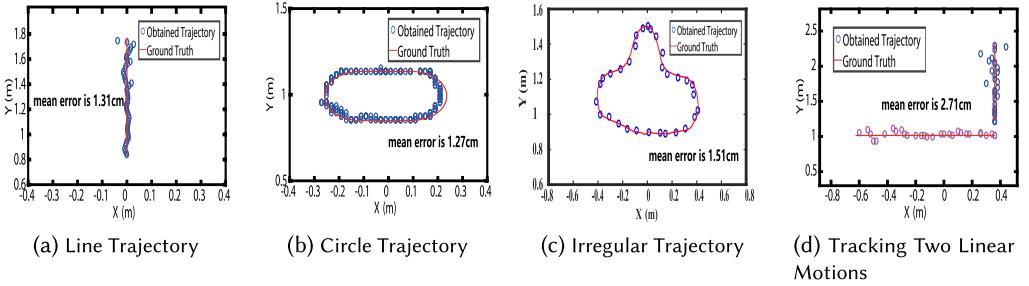


Fig. 11. Tracking results when the tag-free commodities moving in different types of trajectories. The red line is the true trajectory of the target's movement, and the black point is the trajectory obtained by **Tamera** during the movement of the target.

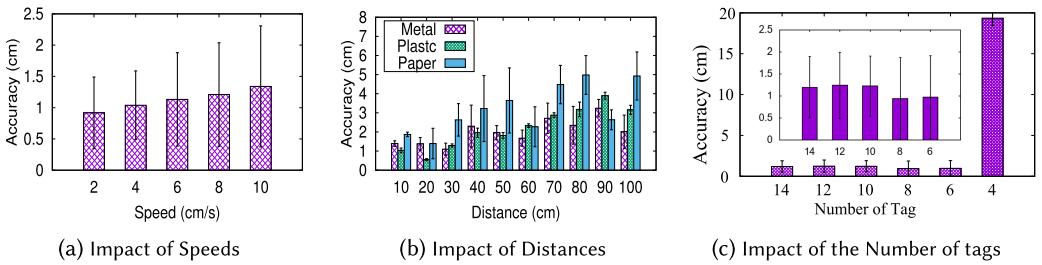


Fig. 12. Impact of different moving speeds, distances to tags, and different commodity materials.

move simultaneously. As shown in Figure 11, the ground-truth is represented in red, and the black dots represent the location of the real-time tracking. According to Figure 11(a), (b), and (c), the estimated trajectories are consistent with the ground-truth. Moreover, as shown in Figure 11(d), **Tamera** achieves good tracking performance even for two concurrent moving commodities.

For the linear track, **Tamera** achieves a mean error of 0.74 cm, 0.98 cm, and 1.31 cm in the  $x$ -axis,  $y$ -axis, and  $x-y$  two-dimensional space, outperforming Tadar [42] and D-Watch [34] by around 10 $\times$  and 3 $\times$ , respectively. For the circular track, **Tamera** achieves a mean error of 0.6 cm, 0.92 cm, and 1.27 cm in the  $x$ -axis,  $y$ -axis, and  $x-y$  two-dimensional space. We arrange two sliders in the monitoring area and use blacktooth to control the motion of the two targets. We calculate the error of the two targets separately and average them as the mean error. Compared to single target tracking, multi-target tracking is much more challenging. **Tamera** can still achieve a mean error of 2.71 cm when two targets move simultaneously.

**Impact of Target Moving Speed.** The target's moving speed affects the tracking performance. We conduct five sets of experiments by setting the target's movement speed<sup>3</sup> as 2 cm/s, 4 cm/s, 6 cm/s, 8 cm/s, and 10 cm/s with the help of a sliding track. As shown in Figure 12(a), as the speed increases, due to the low sampling rate of RFID, the data used for smooth calibration is reduced so that the tracking accuracy is slightly reduced. Moreover, the variance of the data also increases with speed. Even if the speed reaches 10 cm/s, the average error of DTW distance is within 1.4 cm. However, we do believe that when the speed is increased to meter level per second, the induced Doppler shift will greatly affect the tracking accuracy. In a high-speed scenario, to achieve high tracking accuracy, the speed effect does need to be carefully addressed and this is one important direction of our future work.

<sup>3</sup>Note that the maximum achievable speed with this sliding track is 10 cm/s.

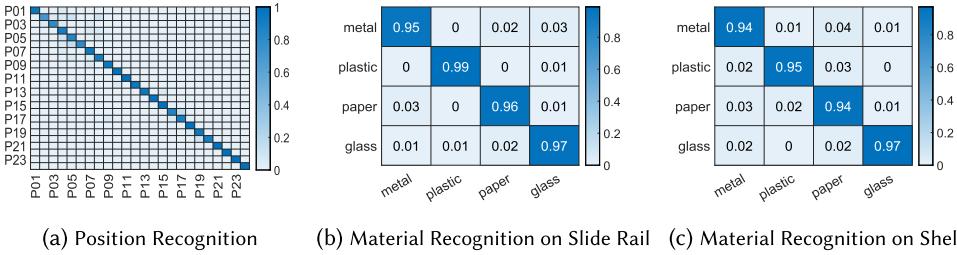


Fig. 13. Material recognition accuracy: Its position accuracy, and under different moving trajectories.

**Impact of Distance.** Due to the contactless nature of our system which relies on the weak reflected signal for tracking, when the target is far away from the tags, the signal may become too weak to be utilized for tracking. We conduct experiments to evaluate the effect of distance between the tag array and the target. It is shown in Figure 12(b) that overall the error increases with a larger distance. However, even when the target is 1 m away from the tag array, the median error is still within 5 cm. To achieve high tracking accuracy, we suggest to keep the distance between the tag array and the target smaller than 60 cm.

**Impact of Material.** Target material directly affects the reflected signal, resulting in performance variations. We evaluate the system performance with three different materials, i.e., metal, plastic, and paper. Due to the stronger reflection from metal commodities, as shown in Figure 12(b), we observe that both the tracking accuracy and tracking distance of the metal targets are higher than that of plastic and paper targets.

**Real-Time Performance.** The end-to-end latency is composed of time spent on the following parts: data preprocessing, signal calibration, and augmented hologram processing. The RFID reader spends 0.2 s to collect the data samples. We employ a Lenovo desktop computer, which is equipped with Intel(R) Core i5-8500 at 2.8 GHz and 8 G memory. We conduct 500 rounds of experiments and the average computer processing time is 0.01 s and 0.08 s for single-target and two-target tracking, respectively. The end-to-end latency is well below 1 s even for a two-target tracking scenario.

### 7.3 Material Recognition

We evaluate the performance of the material recognition module in both the slide rail scene and supermarket scenario, and fully explore the impact of the number of tags, the number of positions, and the human interference.

**Slide Rail Scenario.** In the slide rail scene, metal, plastic, paper, and glass boxes are placed on the slider of the rails as shown in Figure 8(a) in turn. Only one tag array with seven tags and one antenna are used to record RFID signals when the slider moves. The box for each material is moved 50 times, 20 of which are used as test data. The results of the test are shown in Figure 13(b), with an average accuracy of more than 96%.

**Supermarket Scenario.** In the store scene, commodities with different materials are placed on the shelf as shown in Figure 8(c). One tag array with less than 14 tags is placed behind the commodities, and one RFID directional antenna behind the tag array reads the RFID signal. The data is collected once when the experimenter takes an item from the shelf. When the same shape of the glass bottle and the plastic bottled mineral water are used as the recognition target, the system can still accurately recognize the material of the package.

**Number of Tags.** In our recognition system, each antenna tag pair can extract a feature, and a small number of tags cannot obtain enough features to help identify. If the number of tags is too large, the sampling rate will be reduced, which will affect the accuracy of the feature. For a single

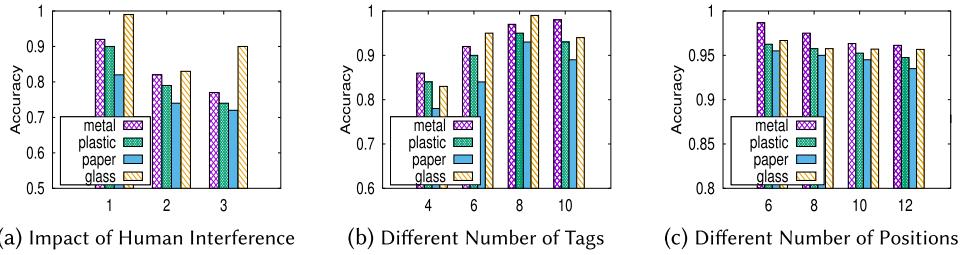


Fig. 14. Impact of various factors on material recognition, mainly human interference and number of tags.

shelf, we need to find an appropriate number of tags, in keeping low interference and maintaining a relatively high sampling rate for accuracy concern. And we evaluate the performance of the material identification module with 4, 6, 8 and 10 tags, respectively. As shown in Figure 14(b), when the number of tags reaches 6, the accuracy of material recognition can be stabilized at over 93%.

**Number of Positions.** We further study whether different number of positions of the commodities affects the accuracy of material recognition. The feature of different positions of the target will form different clusters. If the number of locations increases, a large number of clusters may affect the accuracy of the classification. The commodities are placed in 6 to 12 given positions and we need to evaluate the accuracy when users are taking goods from these candidate positions.

As depicted in Figure 14(c), with the increased number of positions, the accuracy of position recognition drops. Considering the size of the packed goods, it could not be deliberately small (generally larger than 5 cm in diameter). Thus, the number of candidate positions could not exceed 15. To this concern, we can claim that the overall material recognition accuracy is above 94%.

**Human Interference.** We explore the impact of interference from people moving around on the performance of the system. The number of interfering people is increased from 1 to 3. When an experimenter collects data in front of the shelf, different numbers of other experimenters are free to move within 3 meters of the shelf as a disturber. As shown in Figure 14(a), as the number of disturbers increases, the accuracy of the system fluctuates slightly, but still maintains at a relatively high level.

#### 7.4 Behavior Recognition

We evaluate the performance of the behavior recognition module in this section. Four behaviors including (1) taking a targeted item from the shelf, (2) putting back the target to the shelf, (3) putting a commodity in the cart, and (4) taking the target out from the cart are recognized in this module. Figure 15(a) shows the recognition results of four shopping behaviors with multiple experiments, with an average accuracy of 85.5%. And the accuracy of position recognition and the factors that influence behavioral identification are discussed in this section.

**Initial Position Recognition.** As depicted in Figure 8(c), we mimic the supermarket environment with items on the shelf. We collect the phase change at each tag when the item at different initial locations starts moving in advance into the KNN algorithm for training. In this experiment, we put 24 bottles of beverage on two layers of shelves, each with 12 bottles. As shown in Figure 13(a), we can achieve 96.6% accuracy in detecting which bottle is picked.

**Intensive Deployment Scenario.** The scenes in which the shelves are densely arranged in the mall must be considered. We place two shelves vertically together. Shelf A is attached to a tag array with 14 tags and shelf B is attached with 7 tags to compare. Two experimenters are required to make actions simultaneously to evaluate the impact of human interference within close range.

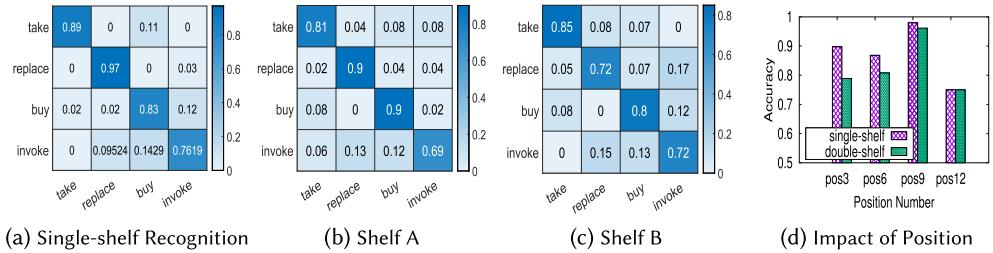


Fig. 15. Shopping behavior recognition accuracy in different scenarios.

As Figure 15 shows, the accuracy of behavior recognition is reduced due to the reduction of the sampling rate and the interference of the action in an excessively dense scenario, but for the case of 14 tags, the average accuracy is still over 83%.

**Impact of Position.** Each experimenter is asked to repeat the above four actions to collect data at different positions in front of the shelf as shown in Figure 8(c). Since we deploy the tags horizontally due to the shape of the shelves, items moving at different locations will have different effects on each tag. One layer in the shelf is divided into 12 zones from left to right. The location of each zone is shown in Figure 8(d). Data from different locations are trained in the same model. We test the accuracy of the model in Zone 3, Zone 6, Zone 9, and Zone 12 as shown in Figure 15(d). Since Zone 12 is too far edged, the reflected signal of the target has less influence on the leftmost tag resulting in low behavior recognition accuracy in this zone. And our models maintain high precision in other zones. For the double-shelf scenario, when the experimenter takes a commodity at position 3, the distance of another experimenter who is simultaneously performing the action is less than 0.5 m, which reduces the recognition accuracy.

## 8 RELATED WORK

In this section, we briefly review the related literature in RFID tracking and recognition, which can be broadly divided into the following categories.

**Device-Based Tracking.** Previous researches have shown that the RFID tags attached to the target can be utilized for target tracking. For example, Tagoram [40] utilizes moving tags to simulate the inverse SAR antennas and pinpoints a tag at an accuracy of a few centimeters. RF-IDraw [32] tracks a moving tag and deduces its trajectory in the air. RF-Echo [4] introduces a custom-designed active ASIC tag and achieves decimeter accuracy in long-range indoor NLOS scenarios. Trio [6] provides an RF interference-based tag localization solution, which locates the target at centimeter level. These systems require the commodity to carry an RF device such as an RFID tag in order to be identified and localized.

**Contactless Tracking.** As an emerging solution for positioning, device-free approaches gain lots of attention in recent years. Tadar [42] establish a Tag-free commodity reflection model and proposed a signal difference method, which shows the potential of device-free sensing in an RFID system. Its accuracy is decimeter level. RF-finger [29] uses RSSI and tag arrays to track the finger movement at centimeter level accuracy. It requires the fingers to be very close to the tag and it is not able to track multiple fingers. D-Watch [34] uses the AOA method to locate multiple commodities at an accuracy of 6 cm. However, it requires a dense deployment of readers and tags which significantly limits its real-life application. **Tamera**, on the other hand, only needs a sparse deployment and can track multiple targets at high accuracies.

**RFID-Based Sensing.** Many exciting new applications have been enabled by RFID sensing. Liquid level [1], vibration frequency [9, 41], orientation [36], breathing [14], and shape [15, 33]

can be perceived with an RFID system. TagScan [33] can accurately identify the liquid type, but a fixed container and stable environment are required. Echoscope [30] extracts unique features from the backscatter signals to reveal the internal status of packages. The idea of replacing RFID chips with sensors [31] makes RFID more effective in the field of sensing. **Tamera** is able to recognize the material type of the target as long as the target is moved.

**RFID-Based Behavior Analysis.** RF-finger [29] and Rio [22] implement several gesture recognitions in close range. Li et al. [17] exploit a deep convolutional neural network for activity recognition with passive RFID readings. ShopMiner [26], TagBooth [18], and Konark [23] deeply analyze the user's shopping behavior based on the information feedback from the tag attached to the commodity. **Tamera**, on the other hand, aims to realize contactless shopping behavior recognition at high accuracies without attaching any tag to the target.

## 9 CONCLUSION

In this work, we propose **Tamera**, a contactless RFID-based multi-commodities deep sensing system. It combines high-precision trajectory tracking of tag-free commodities, with material recognition and shopping behavior recognition for an in-depth customer's shopping behavior study. We propose a novel method of using signal superposition instead of signal separation to address the well-known challenging tag-free multi-commodities tracking. Our design seamlessly integrates the physical model and the popular deep learning schemes to achieve high-precision tracking and recognition. We implement the prototype of **Tamera** using COTS RFID devices. Our comprehensive experiments demonstrate that **Tamera** can achieve an average accuracy of 1.3 cm and 2.7 cm for single- and multi-commodities tracking, 95% for material recognition, and 93% for shopping behavior recognition. Compared with video-based approaches, this work has the benefit of better accuracy and protecting users' privacy naturally.

This work is a stepping stone for the vision of recognizing users' behavior accurately with contactless approaches. A number of challenging issues are left for future study. First, we need to further improve the accuracy (say, sub-centimeter) of tag-free tracking of multiple commodities (say tens concurrently) for a range of innovative applications. Second, system robustness is a critical requirement for the possible adoption of this technique in daily usage. We need to improve its robustness even in a hostile environment, possibly integrating non-wireless techniques.

## REFERENCES

- [1] Rahul Bhattacharyya, Christian Floerkemeier, and Sanjay Sarma. 2010. RFID tag antenna based sensing: Does your beverage glass need a refill? In *Proceedings of the 2010 IEEE International Conference on RFID*.
- [2] Léon Bottou. 2010. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*. Springer, 177–186.
- [3] Jae-Ryong Cha and Jae-Hyun Kim. 2005. Novel anti-collision algorithms for fast object identification in RFID system. In *Proceedings of the 11th International Conference on Parallel and Distributed Systems, 2005*.
- [4] Li Xuan Chuo, Zhihong Luo, Dennis Sylvester, David Blaauw, and Hun Seok Kim. 2017. RF-Echo: A non-line-of-sight indoor localization system using a low-power active RF reflector ASIC tag. In *Proceedings of the International Conference on Mobile Computing and NETWORKING*.
- [5] Han Ding, Jinsong Han, Alex X. Liu, Wei Xi, JiZhong Zhao, Panglong Yang, and Zhiping Jiang. 2018. Counting human objects using backscattered radio frequency signals. *IEEE Transactions on Mobile Computing* (2018).
- [6] Han Ding, Jinsong Han, Chen Qian, Fu Xiao, Ge Wang, Nan Yang, Wei Xi, and Jian Xiao. 2018. Trio: Utilizing tag interference for refined localization of passive RFID. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM'18)*. IEEE.
- [7] Han Ding, Jinsong Han, Yanyong Zhang, Fu Xiao, Wei Xi, Ge Wang, and Zhiping Jiang. 2018. Preventing unauthorized access on passive tags. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM'18)*. IEEE.
- [8] 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 Conference on Computer Communications (INFOCOM'17)*. IEEE.

- [9] Chunhui Duan, Lei Yang, Qiongzheng Lin, Yunhao Liu, and Lei Xie. 2018. Robust spinning sensing with dual-RFID tags in noisy settings. *IEEE Transactions on Mobile Computing* (2018).
- [10] Chuhuan Gao, Yilong Li, and Xinyu Zhang. 2018. LiveTag: Sensing human-object interaction through passive chipless WiFi tags. In *Proceedings of the 15th USENIX Symposium on Networked Systems Design and Implementation (NSDI'18)*. 533–546.
- [11] Xiaonan Guo, Bo Liu, Cong Shi, Hongbo Liu, Yingying Chen, and Mooi Choo Chuah. 2017. Wifi-enabled smart human dynamics monitoring. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*. ACM, 16.
- [12] Haitham Hassanieh, Jue Wang, Dina Katabi, and Tadayoshi Kohno. 2015. Securing RFIDs by randomizing the modulation and channel. In *Proceedings of NSDI*.
- [13] Yuxiao Hou, Jiajue Ou, Yuanqing Zheng, and Mo Li. 2016. PLACE: Physical layer cardinality estimation for large-scale RFID systems. *IEEE/ACM Transactions on Networking* (2016).
- [14] Yuxiao Hou, Yanwen Wang, and Yuanqing Zheng. 2017. TagBreathe: Monitor breathing with commodity RFID systems. In *Proceedings of the 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS'17)*. IEEE, 404–413.
- [15] Haojian Jin, Jingxian Wang, Zhijian Yang, Swaran Kumar, and Jason Hong. 2018. WiSh: Towards a wireless shape-aware world using passive RFIDs. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 428–441.
- [16] Eamonn J. Keogh and Michael J. Pazzani. 2001. Derivative dynamic time warping. In *Proceedings of the 2001 SIAM International Conference on Data Mining*. SIAM, 1–11.
- [17] Xinyu Li, Yanyi Zhang, Ivan Marsic, Aleksandra Sarcevic, and Randall S. Burd. 2016. Deep learning for RFID-based activity recognition. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*. ACM, 164–175.
- [18] Tianci Liu, Lei Yang, Xiang-Yang Li, Huaiyi Huang, and Yunhao Liu. 2015. Tagbooth: Deep shopping data acquisition powered by RFID tags. In *2015 IEEE Conference on Computer Communications (INFOCOM'15)*. IEEE, 1670–1678.
- [19] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. 2015. Contactless sleep apnea detection on smartphones. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 45–57.
- [20] Sophocles J. Orfanidis. 1995. *Introduction to Signal Processing*. Prentice-Hall, Inc.
- [21] Ales Povalac and Jiri Sebesta. 2011. Phase difference of arrival distance estimation for RFID tags in frequency domain. In *Proceedings of the 2011 IEEE International Conference on RFID-Technologies and Applications*. IEEE, 188–193.
- [22] Swadhin Pradhan, Eugene Chai, Karthikeyan Sundaresan, Lili Qiu, Mohammad A. Khojastepour, and Sampath Rangarajan. 2017. Rio: A pervasive RFID-based touch gesture interface. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. ACM, 261–274.
- [23] Swadhin Pradhan, Eugene Chai, Karthik Sundaresan, Sampath Rangarajan, and Lili Qiu. 2017. Konark: A RFID based system for enhancing in-store shopping experience. In *Proceedings of the 4th International on Workshop on Physical Analytics*. ACM, 19–24.
- [24] Longfei Shangguan, Zheng Yang, Alex X. Liu, Zimu Zhou, and Yunhao Liu. 2016. STPP: Spatial-temporal phase profiling-based method for relative RFID tag localization. *IEEE/ACM Transactions on Networking* 25, 1 (2016), 596–609.
- [25] Longfei Shangguan, Zimu Zhou, and Kyle Jamieson. 2017. Enabling gesture-based interactions with objects. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*.
- [26] Longfei Shangguan, Zimu Zhou, Xiaolong Zheng, Lei Yang, Yunhao Liu, and Jinsong Han. 2015. ShopMiner: Mining customer shopping behavior in physical clothing stores with COTS RFID devices. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*. ACM, 113–125.
- [27] John Terry and Juha Heiskala. 2002. *OFDM Wireless LANs: A Theoretical and Practical Guide*. Sams Publishing.
- [28] Shrenik Vora, William Mongan, Kapil Dandekar, Adam Fontecchio, and Timothy Kurzweg. 2016. Wireless heart and respiration monitoring for infants using passive RFID tags. In *Proceedings of the International Conference on Biomedical and Health Informatics (BHI'16)*.
- [29] Chuyu Wang, Jian Liu, Yingying Chen, Hongbo Liu, Lei Xie, Wei Wang, Bingbing He, and Sanglu Lu. 2018. Multi-touch in the air: Device-free finger tracking and gesture recognition via COTS RFID. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM'18)*. IEEE.
- [30] Ge Wang, Jinsong Han, Chen Qian, Wei Xi, Han Ding, Zhiping Jiang, and Jizhong Zhao. 2018. Verifiable smart packaging with passive RFID. *IEEE Transactions on Mobile Computing* (2018).
- [31] Ju Wang, Omid Abari, and Srinivasan Keshav. 2018. Challenge: RFID hacking for fun and profit. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. ACM, 461–470.
- [32] Jue Wang, Deepak Vasishth, and Dina Katabi. 2014. RF-IDraw: Virtual touch screen in the air using RF signals. In *ACM SIGCOMM Computer Communication Review*.

- [33] Ju Wang, Jie Xiong, Xiaojiang Chen, Hongbo Jiang, Rajesh Krishna Balan, and Dingyi Fang. 2017. TagScan: Simultaneous target imaging and material identification with commodity RFID devices. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*.
- [34] Ju Wang, Jie Xiong, Hongbo Jiang, Xiaojiang Chen, and Dingyi Fang. 2017. D-Watch: Embracing “Bad” multipaths for device-free localization with COTS RFID devices. *IEEE/ACM Transactions on Networking (TON’17)*.
- [35] Ju Wang, Jie Xiong, Hongbo Jiang, Kyle Jamieson, Xiaojiang Chen, Dingyi Fang, and Chen Wang. 2018. Low human-effort, device-free localization with fine-grained subcarrier information. *IEEE Transactions on Mobile Computing* (2018).
- [36] Teng Wei and Xinyu Zhang. 2016. Gyro in the air: Tracking 3D orientation of batteryless internet-of-things. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. ACM, 55–68.
- [37] Haibing Wu, Bo Tao, Zeyu Gong, Zhouping Yin, and Han Ding. 2019. A fast UHF RFID localization method using unwrapped phase-position model. *IEEE Transactions on Automation Science and Engineering* 16, 4 (2019), 1698–1707.
- [38] Qingjun Xiao, Bin Xiao, and Shigang Chen. 2013. Differential estimation in dynamic RFID systems. In *Proceedings of the International Conference on Computer Communications (INFOCOM’13)*. IEEE.
- [39] Lei Xie, Jianqiang Sun, Qingliang Cai, Chuyu Wang, Jie Wu, and Sanglu Lu. 2016. Tell me what I see: Recognize RFID tagged objects in augmented reality systems. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*.
- [40] 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 20th Annual International Conference on Mobile Computing and Networking*.
- [41] Lei Yang, Yao Li, Qiongzhen Lin, Xiang-Yang Li, and Yunhao Liu. 2016. Making sense of mechanical vibration period with sub-millisecond accuracy using backscatter signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. ACM, 16–28.
- [42] Lei Yang, Qiongzhen Lin, Xiangyang Li, Tianci Liu, and Yunhao Liu. 2015. See through walls with COTS RFID system!. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*.
- [43] Mingmin Zhao, Fadel Adib, and Dina Katabi. 2016. Emotion recognition using wireless signals. In *International Conference on Mobile Computing and NETWORKING*.

Received 29 March 2022; revised 15 July 2022; accepted 8 September 2022