6-2017

Enabling gesture-based interactions with objects

Longfei SHANGGUAN

Zimu ZHOU
Singapore Management University, zimuzhou@smu.edu.sg

Kyle JAMIESON

Citation

SHANGGUAN, Longfei; ZHOU, Zimu; and JAMIESON, Kyle. Enabling gesture-based interactions with objects. (2017). *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services: MobiSys '17, Niagara Falls, USA, 2017 June 19-23*. 239-251.
Available at: https://ink.library.smu.edu.sg/sis_research/4741

Enabling Gesture-based Interactions with Objects

Longfei Shangguan
Dept. of Computer Science
Princeton University
longfeis@cs.princeton.edu

Zimu Zhou
Dept. of Electrical Engineering
ETH Zurich
zzhou@tik.ee.ethz.ch

Kyle Jamieson
Dept. of Computer Science
Princeton University
kylej@cs.princeton.edu

ABSTRACT

Increasing numbers of everyday objects in libraries, stores and warehouses are instrumented with passive RFID tags, resulting in a ripe opportunity for gesture-based interactions with people. By a simple act of picking up and gesturing with an RFID-tagged object, users can send their opinions and sentiments about that object to the cloud. Prior work in RFID-based gesture tracking relies on multiple bulky and expensive antennas and readers to function, which incurs unacceptable infrastructure costs for large-scale ubiquitous deployment (over an entire warehouse or mall, for example) thus hindering practical adoption. In this paper, we propose Pantomime, the first RFID-based gesture recognition system that uses just a single antenna per geographical area of coverage. Our key insight is to replace the conventional multiple antenna single tag tracking framework with an equivalent multiple tag single antenna system. Through a novel tag coordination protocol and a lightweight tracking algorithm, Pantomime enables accurate gesture tracking that works for objects tagged with just two RFID tags. We implement a real-time prototype of Pantomime with commercial off-the-shelf (COTS) RFID readers and antennas. Extensive evaluations and real-world case studies in a classroom and a retail store demonstrate that Pantomime achieves comparable gesture tracking accuracy (87%) to state-of-the-art multi-antenna schemes (88%) at a minimal deployment cost.

Keywords

RFID; Tracking; Human-object interaction

1. INTRODUCTION

A promising vision for the future Internet of Things (IoT) is to enable input [3, 33, 35, 40], control [1, 13, 28, 38], and interaction [19, 20, 34] via natural gestures: writing and drawing anywhere with any smart objects. Radio Frequency IDentification (RFID) technology holds promise for such capabilities due to the massive existing deployment of small, low-cost RFID tags attached to almost every product and commodity in stores and warehouses. Through tracking the detailed shape of motions of any tagged object held in a user's hand by an RFID reader, the user can perform gestures or write in the air to input his/her comments of the object as an augmented interaction mechanism in stores and warehouses.

Table 1— Comparison of infrastructure deployment in representative RFID-based localization and gesture tracking systems.

System	Antennas	Hardware
RF-IDraw [40]	8	COTS RFID
Tagoram [46]	4	COTS RFID
Backpos [22]	4	COTS RFID
Yunfei *et al.* [23]	4	dedicated tag & reader
PolarDraw [33]	2	linearly-polarized antenna
Pantomime	**1**	**COTS RFID**

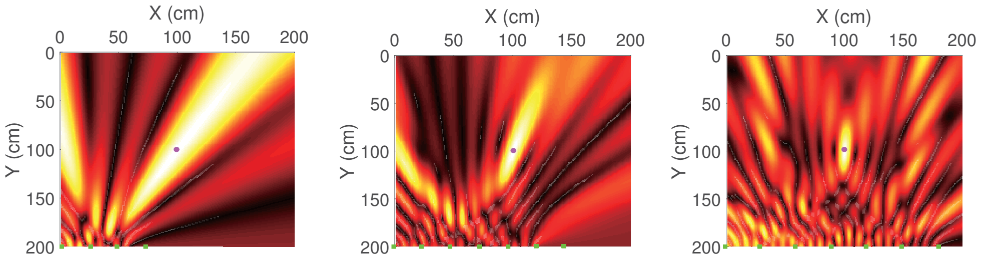
Despite active research on RFID-based localization and gesture tracking [22, 23, 33, 40, 46], we argue that existing schemes do not scale up to very large real world scenarios such as coverage over an entire mall or warehouse. State-of-the-art systems use techniques such as motion-based synthetic aperture radar [32, 46], multi-frequency continuous wave radar [23], and electromagnetic polarization discrimination [33], which achieve centimeter- or even millimeter-accurate localization or tracking accuracy. Yet all of them require multiple bulky and expensive antennas to function (Table 1). Due to the extremely asymmetric cost of RFID reader antennas (over $100) and passive RFID tags (lower than 15 cents) [37], real-world RFID systems are usually deployed to cover a large number of tagged objects with a minimal number of antennas. As a result, a tagged object is often covered by one antenna only, impeding the adoption of the above multi-antenna approaches. We quantify the infrastructure costs for covering a typical indoor space with multi-antenna based systems in §3.1.

In this paper, we present the design and implementation of Pantomime, the first RFID-based human-object interaction system that can track the object motion with a single antenna per geographical area of coverage. The basic idea is to attach multiple RFID tags to an object to compose a tag array. Owning to the reciprocity of the wireless channel, a multi-tag single antenna system is in a sense equivalent to a multi-antenna single tag system. Thus Pantomime can track the motion of a tag-array attached object using a single antenna by tracking the relative movements of the antenna in the frame of the reference of this tag array. The 650×[1] gap in cost between an RFID antenna and a passive tag guarantees that it is far more cost-effective to attach multiple tags to every object within a reading range than deploying additional antennas. With one antenna, Pantomime sets a new standard for RFID-based gesture tracking with a minimal supporting infrastructure, and opening new possibilities for pervasive in-air writing, non-intrusive shop-

[1] For a 15-cent tag and $100 antenna.

MobiSys'17, June 19-23, 2017, Niagara Falls, NY, USA

ISBN 978-1-4503-4928-4/17/06. . . $15.00
DOI: http://dx.doi.org/10.1145/3081333.3081364



(a)— $n = 4$, $s = 20$cm; **(b)**— $n = 7$, $s = 20$cm; **(c)**— $n = 7$, $s = 30$cm;

Figure 1— Hologram generated by an antenna array with different number of antennas and antenna spacing. The green square represents the antenna, the purple dot indicates the ground truth location of the tag. n and s is the number of antennas and antenna spacing, respectively.

Table 2— Impact of coupling effect on phase readings.

Tag spacing (cm)	**4.00**	**3.00**	**2.00**	**1.00**	**0.50**
Phase error (degree)	4.6	10.3	26.3	53.9	81.4

ping behavior identification, and instant commodity reviewing services.

A natural question arises over whether it is feasible to directly apply the algorithms designed for multi-antenna single tag systems such as Angle-of-Arrival (AoA) [40, 45] and hologram based methods [32, 46] to multi-tag single antenna systems by swapping the roles of antennas and tags. We argue that such a naive swapping will lead to severe degradation in gesture tracking performance. Multi-antenna based approaches often make certain requirements on antenna quantity and spacing. However, it is often difficult to attach tag arrays to small object to fulfill these requirements. Specifically, AoA based schemes [40, 45] achieve a high angular resolution by processing phase measurements collected from a large antenna array. Attaching large numbers of tags on small objects, *e.g.*, a book spine, will result in closely spaced tags. A small tag spacing then leads to significant coupling effects between adjacent tags [32, 44], which can cause sizable errors in phase readings (Table 2), and consequently, unacceptable errors for object motion tracking. On the other hand, hologram-based methods [32, 46] demand both sufficient antenna spacing and number of antennas to yield satisfactory beam resolution (Figure 1). Yet it is unrealistic to simultaneously achieve large tag spacing and tag quantity on the surfaces of small objects.

In addition to the difficulty to meet the requirements on tag spacing and tag population to yield a desirable resolution for gesture tracking, the adoption of tag arrays to each inventory object or retail item also brings a new challenge: *low per-tag reading rate*. By attaching two tags to each object within the reading range of a reader antenna, the tag population also doubles. Since standard UHF RFID readers exploit the slotted ALOHA protocol [29] to interrogate RFID tags, a boosted tag population will cause significant tag reply collisions. Hence the reading rate (the number of successful tag interrogations per second) of each tag will drop by almost half, leading to sparsely-collected measurements that may even fail to accurately track gestures performed at a normal speed. We quantify the impact of tag population on per-tag reading rate in §3.2.

Pantomime addresses the above challenges by incorporating (i) a MAC layer tag filtering algorithm to boost the reading rate of those tags attached to the target object, and (ii) a novel Extended Kalman Filter (EKF) based tracking algorithm that tracks fine-grained object motions in a 2D plane with a limited number of tags attached to an object. Specifically, Pantomime works as follows. It detects tag movement by examining the statistical distribution of their phase readings. These moving tags are target tags for tracking. Pantomime then stops those stationary tags from responding to the reader, thereby reducing the RF traffic load and response collisions to improve the reading rate of the target tags. As the object moves (so does the tag array attached to it), Pantomime tracks the relative location and the heading of the antenna in the frame of the reference of the tag array. Finally, the relative locations of the antenna, together with a kinetic model based tracking result are fed into an EKF fusion model to further enhance the tracking accuracy.

Contributions and Roadmap. We design and implement the first RFID-based human-object interaction system that enables users to input their opinions and sentiments of an object through instant gestures. We demonstrate the potential widespread applicability of Pantomime for augmented interactions with two real-world case studies: whiteboard handwriting tracking and retail-store item querying (§2). Although attaching multiple tags on an object to infer the object orientation has been studied [44], Pantomime is the first system we are aware of that uses a tag array for gesture-based human-object interaction. We present quantitative experimental measurements to demonstrate the antenna coverage problem and the low tag reading rate issue in real-world scenarios (§3). Novel techniques introduced in §4 (Design) address these challenges, and thus have the potential to be applied to existing RFID infrastructure deployments without deploying additional hardware. We implement Pantomime using COTS RFID devices (§5), evaluate both end-to-end performance and microbenchmarks in laboratory environments (§6), and compare it with two state-of-the-arts tracking systems: four-antenna Tagoram [46] and two-antenna PolarDraw [33]. The result shows that Pantomime achieves competitive tracking performance using a single antenna. We review related work in §7, discuss limitations in §8 and conclude in §9.

2. CASE STUDY

In this section, we demonstrate the potential applicability of Pantomime for gesture-based object interaction with two real-world case studies: whiteboard handwriting tracking and retail-store item querying.

2.1 Whiteboard handwriting tracking

Although previous work [33, 40] has shown the feasibility of handwriting tracking using multi-antenna RFID systems, this case study aims to demonstrate that handwriting tracking is viable with a single antenna. Enabling handwriting tracking with minimal RFID infrastructure is essential to promote RFID-based sensing technologies to smart homes, classrooms, and offices, where RFID systems haven't been pervasively deployed.

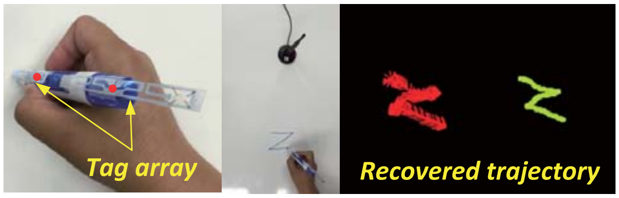


Figure 2— Handwriting tracking: the antenna is placed near the pen for a better illustration. Tag spacing = 2.7cm.

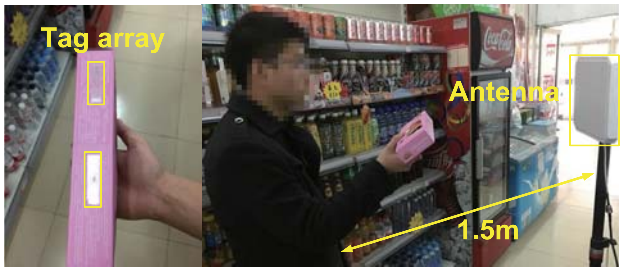


Figure 3— Retail store case study. Tag spacing = 5.4cm.

As an illustration, we deploy Pantomime in a classroom. We install an antenna on top of a whiteboard and attach two passive RFID tags to a marker pen as in Figure 2. A volunteer is then asked to write letters at random on the whiteboard using the tagged marker pen. We plot the ground-truth trajectory of the marker pen and the trajectory captured by Pantomime. In Figure 2, the red arrows show the motion headings inferred by Pantomime, which are input into our fusion algorithm to derive the final trajectory denoted by the green line. Compared with the ground-truth, the recovered letter is stretched and rotated due to tracking errors. However, the trajectory shape of this letter Z is well preserved and can be easily recognized. This result clearly demonstrates the viability of Pantomime to track the handwriting with high fidelity. In §6.3, we quantitatively evaluate the quality of the trajectories recovered by Pantomime and compare it with the state-of-the-art multi-antenna RFID tracking systems.

2.2 Retail-store item querying

Acquiring customers' opinions on groceries is desired for shopkeepers to optimize the trading strategy and profit. While large online retailers *e.g.* Amazon can directly collect opinions of items via online feedback systems, most small to medium sized offline retailers still lack effective methods to acquire customers' opinion on groceries. With proper incentive mechanisms such as coupons and discounts, customers might be motivated to provide instant and offline opinions on the goods if there is an interactive way to express their opinions. This case study aims to demonstrate the viability of enabling RFID-based customer interactions, including offering feedback on the price and quality of daily goods, and reporting out-of-date items by performing the following in-air gestures with the item held in hand:

♡: I like this item;	↓: Too expensive;
?: Price unclear;	O: Item (*e.g.*, milk, bread) Out-of-date;

This gesture set is designed to include gestures consisting of both *straight lines* and *curves*. Figure 3 shows the deployment of Pantomime in a retail store, where passive RFID tags are attached to 40 items, with two tags on each item. A volunteer is asked to randomly pick up the items and perform the four gestures above.

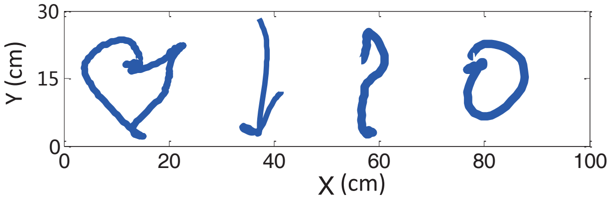


Figure 4— Recovered trajectory: ♡, ↓, ? and O.

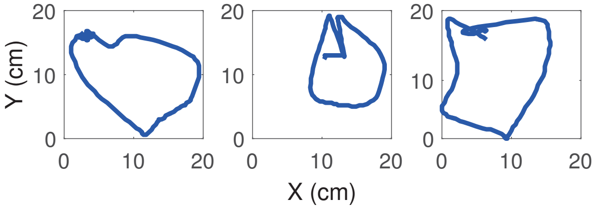


Figure 5— Recovered gesture when a user writes ♡ in the air using retail items with different weights: the weight of the retail item decreases from the left to the right.

Figure 4 shows the recovered trajectories of the gestures. Apparently, the strokes in each gesture are warped or stretched due to the fact that the volunteer may change the item orientation unintentionally during gesturing. An interesting finding is that the recovered trajectories are more easily recognized when performing gestures holding a heavier item, as shown in Figure 5. A partial explanation might be that the volunteer will hold the item more steadily if it is heavy, which helps to keep the orientation of the tag array unchanged. Note that we do not quantitatively evaluate the recognition accuracy of these four gestures. Our aim is to demonstrate that the trajectories of such free-hand drawing-like gestures tracked by Pantomime are recognizable by humans. Selecting a specific machine learning algorithm for gesture recognition is out of the scope of this work.

3. CHALLENGES

The previous section demonstrates potential applications of Pantomime for emerging augmented interactions. In this section, we conduct two quantitative measurements on (i) antenna coverage and (ii) per-tag reading rate with different antenna and tag population settings to show (i) the difficulty of adopting multi-antenna based tag tracking schemes with a cost-effective system deployment and (ii) the new challenge introduced by tag arrays.

3.1 Antenna crisis

To quantitatively measure the number of antennas needed to deploy existing RFID-based localization and tracking systems in medium-sized stores and warehouses, we conduct simulations in a 100m x 100m rectangular region. The region is divided into 1m x 1m *blocks*. We deploy different numbers of antennas in this region and calculate the *four-coverage ratio* (Backpos [22] and Tagoram [46]) and *eight-coverage ratio* (RF-IDraw [40]) of these blocks. Here an *n-coverage ratio* is defined as the percentage of blocks that can be covered by at least *n* reader antennas among all blocks. To make a conservative estimation of the number of antennas required for *n-coverage*, we set a wider beam-width (120°) and a higher antenna gain (20 dBm) of each antenna in our simulation. Figure 6 illustrates the coverage ratio under different numbers of antennas. We can see that over 85% of the region is covered by at least one antenna when deploying 220 antennas. In contrast, 900 antennas are

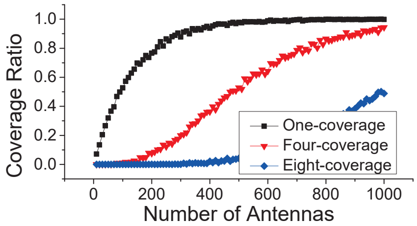


Figure 6— Coverage ratio vs. Number of antennas (simulation).

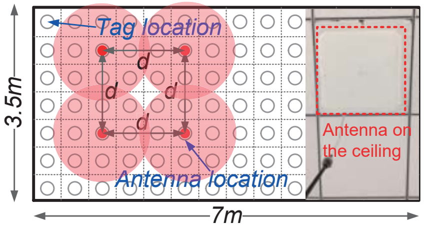


Figure 7— Coverage testing setup.

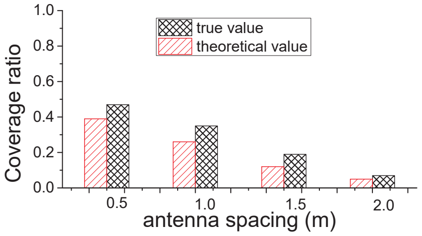


Figure 8— Four coverage ratio vs. antenna spacing *d* (experimental).

Table 3— The average readings (times/sec) of tag (**T**) and reader (**R**) under different tag population (*n*) settings.

	n = 1	2	5	20	50	100
T	51.23	46.32	21.92	15.34	7.78	4.14
R	51.23	92.64	109.6	306.79	388.84	413.52

required in order to provide a 85% four-coverage ratio. For eight-coverage, 1,000 antennas can only achieve around 54% coverage ratio. The result clearly shows that the multiple antenna coverage requirement will cause an "antenna crisis" – the number of antennas required grows dramatically with the expansion of the area to be covered.

To better understand the antenna coverage problem in real-world deployment, we deploy multiple antennas in a laboratory to examine the antenna coverage ratio. In this experiment, we mount four circularly polarized antennas [16] on the ceiling 3 m above the floor in a 3.5 m x 7 m laboratory, place an RFID tag in different locations in the room (white circles in Figure 7), and measure the number of antennas that can interrogate the tag. We estimate the coverage ratio as the number of locations that are covered by four antennas over different antenna spacing, and plot the results in Figure 8. As shown, four antennas with a spacing *d* of 0.5 m yields a 47% four-coverage ratio, which is moderately higher than the theoretical value due to multi-path reflections. Given a spacing *d* of 2 m, which resembles the antenna deployment in a real-world RFID system [41], the four-coverage ratio drops dramatically to only 7%. In summary, Deploying a multi-antenna RFID tracking system in malls, supermarkets, and warehouses will incur significant infrastructure costs.

3.2 Low tag reading rate

A high tag reading rate is essential to accurately track tag movement. Insufficient reading rates are especially harmful to phase-based rfid tracking schemes because phase readings repeat from 0 to 2π every wavelength, and a low reading rate will fail to capture the consecutive phase changes when the tag moves fast, resulting in an ambiguity in the tag's displacement. In gesture or handwriting tracking applications, a tagged object can move at a speed ranging from 0.5 m/s to 3.0 m/s [9]. Thus the tag should be read at a minimal rate of 8 Hz to 46 Hz to avoid this phase ambiguity, given a wavelength of 32 cm (UHF frequency 920.625 MHz).

Commercial RFID readers adopt Slotted ALOHA protocol [29] to interrogate tags within the reading range. Each tag randomly picks a time slot to reply. A collision occurs when multiple tags reply in the same slot and the reader will fail to decode any of them. Although the allocated slots can scale to the tag population [46], collisions occur much more frequently for a large tag population due to the limited frame length [31], which decreases the per-tag reading rate. In this experiment, we examine the average tag reading rate with different numbers of tags in an indoor laboratory environment. Specifically, we place an RFID antenna 2 m away from a bookshelf to interrogate tags attached to the spine of books. As shown in Table 3, the reader's reading rate (**R**) increases significantly as tag population grows. However, the average per-tag reading rate ((**T**)) drops significantly from 46 Hz to around 4 Hz with 100 tags. Such a low tag reading rate fails to meet the theoretical minimum requirement for gesture tracking. As we further attach multiple tags on each object, there will be more MAC contention and more collisions, resulting in a much lower per-tag reading rate.

4. DESIGN

This section first introduces our proposed technique for boosting the tag reading rate, followed by the virtual antenna tracking algorithm, and finally movement translation, the processing step that outputs the trajectory of the tagged object.

4.1 Boosting the tag reading rate

The basic idea to boost the reading rate of the tag array is to make the reader interrogate these target tags only, instead of all tags within the reading range of the reader. Here *target tags* refer to the tags on an object that a user holds in hand and performs gestures with. *Non-target tags* are the tags on other stationary objects nearby. Thus the reading rate boosting scheme contains two steps: (i) identifying the target tags and (ii) blocking non-target tags from responding to the reader.

Inferring the target tags. To differentiate target and non-target tags, we leverage the intuition that the pick-up action will make target tag moving, hence leading to continuous phase changes of these target tags. In contrast, the non-target tags keeps stationary and their phase reading remains. To validate the above intuition, we attach 40 tags to 20 books (two tags per book), and place an RFID antenna 4 m away to interrogate the tags. The per-tag reading rate in this experiment is around 10 Hz. We collect phase measurements under four settings: a clear Line-Of-Sight (LOS) path between the tags and the antenna, people stands by to create reflection paths, people stands in between to block the LOS path, and people pick up a book to review. Figure 9 illustrates these testing setups.

For each testing case, we randomly pick up a tag and plot their phase readings in Figure 10. As shown, the phase readings remain stable in clear LOS path condition. When the people walks to the bookshelf and generate a reflection path, we can see the phase readings of the stationary tag jumps from a stable level to another stable level. Similar trend happens when the human walks to block the LOS path between the tag and the antenna. However, when the tag is picked up by the volunteer, we can see the phase readings of this tag change constantly across the whole cycle of $[0, 2\pi]$. Hence it is possible to distinguish the target tags by observing its phase changing trend.

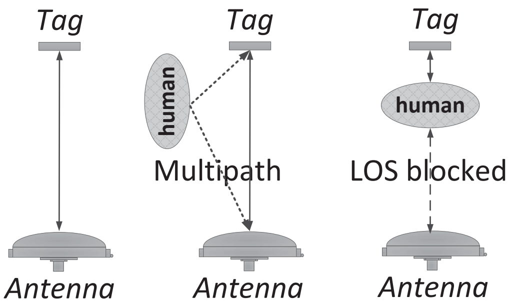


Figure 9— Examining the motion-induced phase changes.

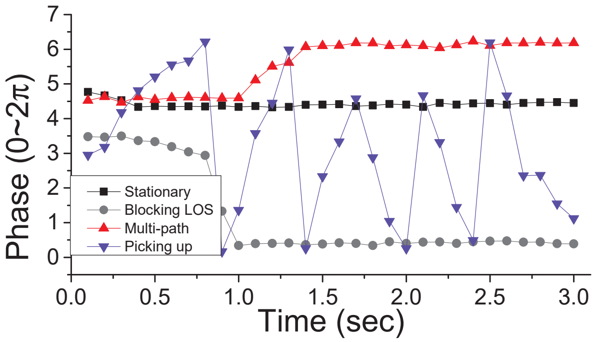


Figure 10— Phase pattern of tags under different motion states.

In Pantomime, we quantify the phase changing trend using an entropy metric. Low entropy correlates with narrow phase distribution likely from a stationary, non-target tag, while high entropy indicates a wider phase distribution possibly from a moving, target tag. Specifically, we split the range of phase readings [0, 2π] into N bins with equal bin size. The bin size is empirically set to 0.5 radians (see §6.2). Given a set of phase readings of tag i, let n_j be the number of phase readings within j^{th} bin. We write the entropy of tag i as:

$$e_i = -\frac{1}{\lg(N-1)} \sum_{j=1}^{N} p_i^j \lg p_i^j \qquad (1)$$

where $p_i^j = \frac{n_j}{\sum_{k=1}^{N} n_k}$[2]. Pantomime then formulates the following hypothesis test with H_0 representing the hypothesis of a moving, target tag and H_1 of a stationary, non-target tag:

$$\begin{cases} H_0 : e_i \geq \gamma \\ H_1 : e_i < \gamma, \end{cases} \qquad (2)$$

where γ denotes a predefined threshold. We conduct micro-benchmarks to test various γ settings and empirically set it to 0.7 (§6.2), which yields a satisfactory detection performance. As false positives may still exist, Pantomime further filters out those target tags whose siblings are detected as non-target tags. Here the siblings of a tag represent the remaining tags in the same tag array.

Blocking the reply of non-target tags. After finding out the target tags, our next step is to stop non-target tags from replying to the reader, thereby reducing the amount of RF traffic and response collisions to improve the reading rate of the target tags. To achieve this goal, we manipulate the EPC-standard C1G2 low-level tag interrogation process [7]. As Figure 11 shows, at the beginning of each inventory round (frame), the reader executes a *SELECT* command to choose tags for inventory and access. After receiving this command, the tags satisfying this selection criteria will enter the *READY* state, indicating that they are ready to reply. Other tags will keep silent (and not reply) until the end of this inventory round. By setting the *SELECT* criteria, we can thus stop the stationary tags from replying. The default setting of *SELECT* is to let all tags reply. In Pantomime, after detecting the target tags, we set the *SELECT* command to read the target tags alone by specifying the 96-bit EPC code of these tags. This is achieved by calling the function setC1G2TagInventoryMask of the C1G2Filter class provided by the Low-level Reader Protocol (LLRP) API.

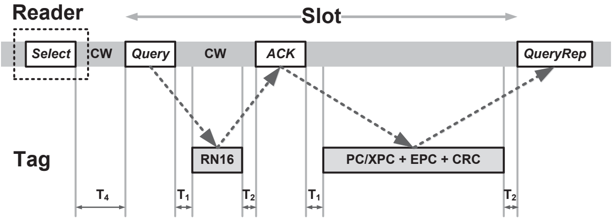


Figure 11— Low-level reader to tag communication.

After performing a gesture, the user may place the object back to the rack, resulting in a significant phase fluctuations due to the object motions, and then relatively stable phase readings when the object becomes stationary. Pantomime detects this state transition by examining the variance of phase readings within every five seconds. A small variance indicates the object has been placed back on the rack, hence Pantomime will reconfigure the tag mask to let all tags reply. We test different variance threshold and set it to 0.3, which yields an empirically minimum false positive rate.

4.2 Virtual antenna trajectory tracking

To track the object motion using a single antenna, we track the location changes of the antenna in the frame of the reference of the tag array and transform into a gesture trajectory in physical coordinates via motion translation. In Pantomime, we keep tracking the relative location of the antenna by estimating the change in tag-to-antenna distances. However, the accuracy of such ranging-based tracking algorithm is undesirable due to multi-path reflection and thermal noises [46]. In Pantomime, we further improve the accuracy of ranging-based tracking by incorporating a kinetic model and an Extended Kalman Filter (EKF) fusion framework. In the following parts, we first introduce our ranging-based tracking algorithm, and then describe how to improve its tracking accuracy with EKF model.

4.2.1 Pseudo-ranging tracking algorithm

Let $P_t = (x_t, y_t)$ be the virtual position of the antenna at time t; $R_j = (x_{R_j}, y_{R_j})$ be the location of tag j in the frame of reference of the tag array, which is known a prior. The distance from the antenna to tag j can be represented as:

$$d_t^j = \|P_t, R_j\| = \sqrt{(x_t - x_{R_j})^2 + (y_t - y_{R_j})^2} = \lambda \left(\frac{\theta_t^j}{4\pi} \right) + s_j \quad (3)$$

where θ_t^j is the phase reading of tag j at time t; s_j is an unknown distance offset. We assume the object moves within half a wavelength (≈16 cm) during consecutive tag readings.[3] Accordingly,

[2] we set $\lg p_i^j$ to 0 when $p_i^j = 0$.

[3] That is, no faster than 3 m/s given the reading rate of 46 Hz, which can be achieved after boosting the tag reading rate.

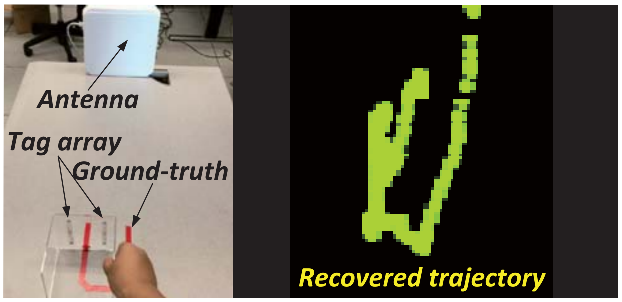


Figure 12— Recovering the trajectory of the moving cube based on pseudo-ranging.

the virtual moving distance of the antenna during time t and $t+1$ can be computed as follows:

$$\Delta d_{t+1}^{j}=\begin{cases}(\theta_{t+1}^{j}-\theta_{t}^{j})\cdot\lambda/(4\pi), & \text{if}\quad |\theta_{t+1}^{j}-\theta_{t}^{j}|<\pi\\ (\theta_{t+1}^{j}-\theta_{t}^{j}-2\pi)\cdot\lambda/(4\pi), & \text{if}\quad \theta_{t+1}^{j}-\theta_{t}^{j}\geq\pi\\ (\theta_{t+1}^{j}-\theta_{t}^{j}+2\pi)\cdot\lambda/(4\pi), & \text{if}\quad \theta_{t+1}^{j}-\theta_{t}^{j}\leq-\pi\end{cases} \tag{4}$$

Now the distance between the reader antenna and each tag can be updated as follows:

$$d_{t+1}^{j}=\|P_{t+1},R_{j}\|=d_{t}^{j}+\Delta d_{t+1}^{j}=d_{1}^{j}+s_{j}+\sum_{k=1}^{t+1}\Delta d_{k}^{j} \tag{5}$$

In the above equation, we cannot compute d_{t+1}^{j} due to the unknown distance offset s_j. However, s_j is a constant value and will not change throughout the object's movement. Hence we omit s_j and define the *pseudo-distance* between the antenna and tag j as follows:

$$\hat{d}_{t+1}^{j}=d_{1}^{j}+\sum_{k=1}^{t+1}\Delta d_{k}^{j} \tag{6}$$

In the above equation, d_1^j and $\sum_{k=1}^{t+1}\Delta d_k^j$ can be computed using Equation 3 and Equation 4, respectively. Thus we can estimate the antenna's position (P_t) at each time point and determine its virtual moving trajectory in the frame of the reference of the tag array with at least two tags.

The pseudo-ranging tracking algorithm works in two steps: computing the pseudo initial distance d_1^j between the antenna and each tag j, and then consecutively estimating the new location of the antenna as the object (antenna array) moves. Specifically, in each time frame, the algorithm computes the virtual moving distance Δd_{t+1}^{j} of the antenna and updates the pseudo-distance $\hat{d}_{t+1}^{j}$ between each tag and the antenna. With $\hat{d}_{t+1}^{j}$, the algorithm estimates the antenna's virtual location via trilateration.

Figure 12 shows the trajectory of a letter U recovered by the pseudo-ranging tracking algorithm. As shown, the recovered trajectory preserves the rough shape of the letter. However, there are sharp and abrupt discontinuities in the trajectory, especially at the corners of the shape. This is because the location of the antenna is estimated independently each time. Therefore, large location deviation occurs during consecutive location estimates due to phase noises and multi-path effect. To avoid unexpected discontinuities in the recovered trajectory, we further leverage heading information to bound the deviation of successive location estimates.

Antenna heading inference. In Figure 13(b), suppose the antenna moves from position A to B within two tag readings. Since the sampling rate of the tag is high, hence we know $|AB|\ll|AT_2|$. Thus it is reasonable to approximate the antenna's displacement along the radial direction $|AB_2|$ with Δd. On the other hand, since the antenna's position A is estimated through the position-velocity model (described later), it is thus feasible to determine the heading of $\angle\overrightarrow{AB_2}$ as well. Taking one step further, we can infer the displacement and the heading of $\angle\overrightarrow{AB_1}$ and $\angle\overrightarrow{AB_3}$ in a similar way (as shown in Figure 13(c)), and finally determine the antenna heading by summing up these vectors.

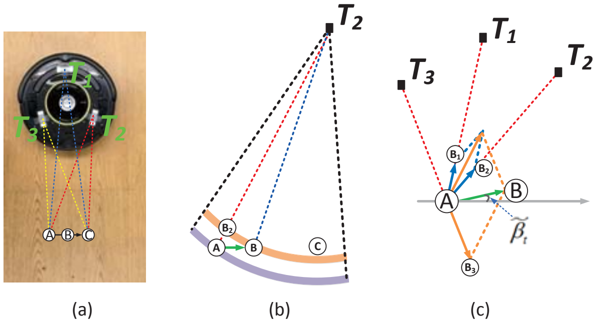


Figure 13— An illustration of the antenna heading inference. The antenna virtually moves from point A to point B and our algorithm continuously tracks the antenna heading during its virtual movement.

4.2.2 *Kinetic model based tracking algorithm*

Position-velocity model. We employ constant acceleration model [21] to characterize the hand motion. So the antenna movement suits this model as well. In the frame of reference of the tag array, consider the discrete time model: let T be the window size in *ms*. Within each window, the state of the antenna can be characterized by its location $P=[x_t,y_t]$, velocity $V=[\dot{x}_t,\dot{y}_t]$ and acceleration $A=[\ddot{x}_t,\ddot{y}_t]$:

$$X_t=[P,V,A]=[x_t,\dot{x}_t,\ddot{x}_t,y_t,\dot{y}_t,\ddot{y}_t] \tag{7}$$

The state transition model can be written as:

$$X_{t+1}=f(X_t)+W_t=AX_t{}^{T}+W_t$$
$$=\begin{bmatrix}1&T&T^2/2&0&0&0\\0&1&T&0&0&0\\0&0&1&0&0&0\\0&0&0&1&T&T^2/2\\0&0&0&0&1&T\\0&0&0&0&0&1\end{bmatrix}\cdot\begin{bmatrix}x_t\\\dot{x}_t\\\ddot{x}_t\\y_t\\\dot{y}_t\\\dot{y}_t\end{bmatrix}+\begin{bmatrix}T^2/2&0\\T&0\\1&0\\0&T^2/2\\0&t\\0&1\end{bmatrix}\cdot\begin{bmatrix}n_t^x\\n_t^y\end{bmatrix}$$

where A models the state transition of the antenna from window t to $t+1$. Vector W_t takes thermal noise into account, which follows normal distribution with zero mean and covariance matrix Q_t: $W_t\sim\mathcal{N}(0,Q_t)$. With the antenna heading and the position-velocity model, it is viable to track the virtual movement of the antenna in the frame of reference of the tag array, and accordingly the movements of the tagged object. However, since the acceleration of the hand gesture may change abruptly, the constant acceleration model fails to accurately characterize the shape of the virtual antenna motion.

4.3 Fusion Algorithm

As pointed out previously, the ranging-based method can recover the rough shape of the virtual antenna movement yet fail to characterize the trajectory smoothly. Conversely, the kinetic model based method could track the trajectory smoothly, yet fail to accurately characterize its shape with natural hand movements. We thus leverage the Extended Karman Filter (EKF) model to fuse the tracking results of these two tracking algorithms for a better result.

Observation. Let N be the number of tags attached to the object. The observation function is defined as:

$$Y_{t+1} = h(X_{t+1}) + V_{t+1} = \begin{bmatrix} \|P_{t+1}, R_1\| \\ \|P_{t+1}, R_2\| \\ \vdots \\ \|P_{t+1}, R_N\| \end{bmatrix} + V_{t+1} = \begin{bmatrix} d_t^1 + \Delta d_{t+1}^1 \\ d_t^2 + \Delta d_{t+1}^2 \\ \vdots \\ d_t^N + \Delta d_{t+1}^N \end{bmatrix} + V_{t+1} \qquad (8)$$

where V_{t+1} is the ranging errors due to phase noises. We empirically set V_{t+1} to $\frac{1}{10} \cdot mean(\Delta d_t^j)$, where $mean(\Delta d_t^j)$ is the average value of the change of the distance between the antenna and each tag. $h(.)$ is a non-linear function that relates observations and states.

EKF-based Motion Tracking. The EKF-based fusion algorithm consists of three steps: linearization, prediction and updating. We detail them in turn in the following parts.
Linearization. In Equation 8, the nonlinear function $h(.)$ can be linearized about X_0 as follows:

$$Y_t - Y_{t-1} = H_t \Delta X_t + V_t \qquad (9)$$

where H_t is the Jacobian matrix of $h(.)$ with respect to X_n.
Predicting phase. Given the estimate $\hat{X}_{t|t}$ of X_t, the predicted state $\hat{X}_{t+1|t}$ is calculated as:

$$\hat{X}_{t+1|t} = F_t \hat{X}_{t|t} \qquad (10)$$

where $F_t = \frac{\partial f}{\partial X}|_{\hat{X}_{t|t}}$ is the Jacobian matrix of $f(.)$. The prediction error covariance matrix $\hat{P}_{t+1|t}$ associated to the predicted state estimation $\hat{X}_{t+1|t}$ is evaluated from the previous estimate $\hat{P}_{t|t}$:

$$\hat{P}_{t+1|t} = F_t \hat{P}_{t|t} F_t^T + Q_t \qquad (11)$$

The predicted measurement is then computed as follows:

$$\hat{Y}_{t+1|t} = h(\hat{X}_{t+1|t}) + V_{t+1} \qquad (12)$$

Updating phase. Once the system acquires new phase readings at time $t+1$, it computes the difference between the measurement and the predicted measurement as follows:

$$\gamma_{t+1} = Y_{t+1} - \hat{Y}_{t+1|t} \qquad (13)$$

with the covariance:

$$S_{t+1} = H_{t+1} \hat{P}_{t+1|t} H_{t+1}^T + R_{t+1} \qquad (14)$$

where $H_{t+1} = \frac{\partial h}{\partial X}|_{\hat{X}_{t+1|t}}$ is the Jacobian matrix of the state transition function $f(.)$ with respect to the predicted state $\hat{X}_{t+1|t}$. After that, it computes the EKF gain as follows:

$$K_{t+1} = P_{t+1|t} H_{t+1} (\hat{S}_{t+1})^{-1} \qquad (15)$$

Finally, the algorithm computes a posterior state estimate $X_{t+1|t+1}$ and the corresponding covariance matrix $\hat{P}_{t+1|t+1}$ by correcting the priori state estimate $\hat{X}_{t+1|t}$ and $\hat{P}_{t+1|t}$:

$$\hat{X}_{t+1|t+1} = \hat{X}_{t+1|t} + K_{t+1} \gamma_{t+1} \qquad (16)$$

$$\hat{P}_{t+1|t+1} = (I_N - K_{t+1} H_{t+1}) \hat{P}_{t+1|t} \qquad (17)$$

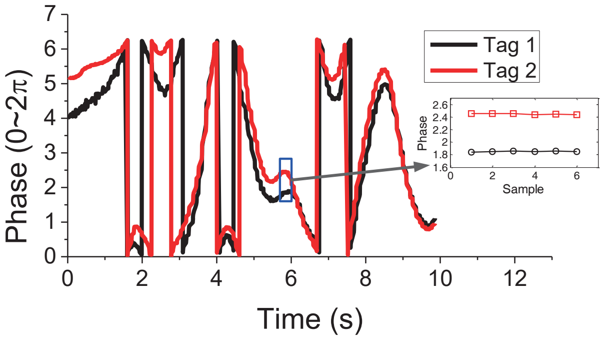


Figure 14— Phase readings of two tags attached to an object that has been picked up for gesturing.

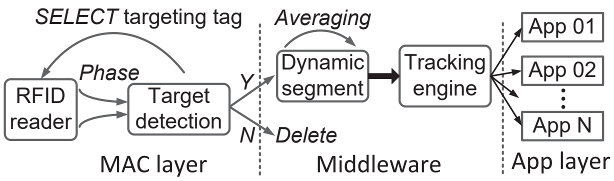


Figure 15— System workflow of Pantomime.

4.4 Motion Translation

Let $S = \{P_1, P_2, ..., P_t\}$ be the recovered trajectory of the antenna in the frame of reference of the tag array and S' be the real trajectory of the object in the frame of the reference of the physical world. We can recover S' by rotating S by π:

$$S' = S \cdot \begin{bmatrix} \cos\pi & -\sin\pi \\ \sin\pi & \cos\pi \end{bmatrix} \qquad (18)$$

5. IMPLEMENTATION

So far we have provided the algorithmic basis of how Pantomime works. This section describes its implementation.

RF-sensing module. The hardware of the RF-sensing module consists of three parts: an ImpinJ Speedway R420 RFID reader [11], an RFMAX indoor RFID antenna [16], and multiple Avery Dennison AD-227m5 UHF passive RFID tags [5]. The reader interrogates RFID tags and sends phase readings to a Lenovo ThinkCentre PC for processing.

Frontend software. Figure 15 shows the workflow of Pantomime. The reader interrogates tags using the Low Level RFID Protocol (LLRP) API. We implement the reading rate control function in the MAC layer. Once the target tags are detected, the reader stops other stationary tags from replying in the following frames. The phase readings of target tags are then delivered to the middleware layer for segmentation and tracking. The output of the middleware layer is the trajectory of the antenna in the frame of the reference of the tag array, which can be integrated to various applications (after trajectory translation) such as virtual touch screen and in-air book reviewing. The software is implemented in C# and it tracks the object in real-time.

Dynamic segmentation. To minimize the noises and interference, Pantomime segments phase streams into windows and uses the average phase in each window as the input of the tracking engine. The window size is critical to our system. If it is too small, Pantomime needs to frequently feed the phase readings to the tracking engine, leading to significant computational overhead. If the window size

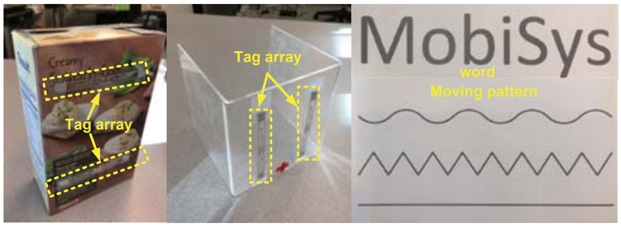


Figure 16— Experimental setup: We construct tag arrays by attaching multiple tags to a food package and a plastic cube, and move them along trajectories printed on paper.

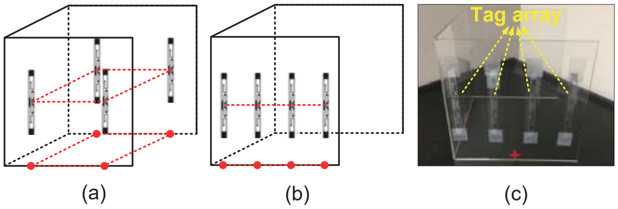


Figure 17— Different layout of tag arrays.

is too large, the phase changes induced by motions are likely to average out, resulting in tracking errors. Either case will degrade the system performance.

In Pantomime, we design a dynamic segmentation mechanism. It is based on the following finding: if the object moves slowly, the phase will change slowly and smoothly, and we can safely expand the window size. If the object moves fast, the phase will change rapidly, we need to reduce the window size to preserve the detail of the phase profile. Specifically, Pantomime updates the window size as follows:

```
if $v_t$ - $v_{t-1}$ < -$v_{TH}$ & $W_t$ + $W_a$ < $W_{max}$ then
    $W_{t+1} \leftarrow W_t + W_a$;
else if $v_t$ - $v_{t-1}$ > $v_{TH}$ & $W_t$ - $W_a$ > $W_{min}$ then
    $W_{t+1} \leftarrow W_t - W_a$;
else
    $W_{t+1} \leftarrow W_t$;
end if
```

where W_α is a constant to control the window length; v_t is the moving speed of the tag in the time window t; v_{TH} is the moving speed threshold. Within each window, the window size updates until it surpasses or falls below the maximum and minimum window size W_{max} and W_{min}, respectively. We test a wide range of parameter settings, and empirically set v_{TH}, initial window size W, W_α, W_{max} and W_{min} to 0.1 m/s, 40 ms, 20 ms, 200 ms and 20 ms, respectively.

6. EVALUATION

In this section, we present the experimental evaluation results of Pantomime, starting with the experimental methodology, microbenchmarks and ending with field studies.

6.1 Experimental methodology

We attach RFID tags to both a food package and a plastic cube to form a tag array and move the array along printed trajectories. The trajectories cover different English letters, words (with different lengths) and paths (see Figure 16).

We use *recognition accuracy* to quantify the performance of Pantomime. The *recognition accuracy* is defined as the fraction of correctly recognized hand-written letters/words over total number of hand-written letters/words. For letter recognition, we use LipiTk [10], an open-source handwriting recognition toolkit.

Table 4— FP/FN rate under different γ and bin size settings.

		Bin size			
		0.5 (rad)	**1**	**1.5**	**2**
γ	**0.1**	0.72/0.01	0.72/0.01	0.72/0.01	0.81/0.03
	0.3	0.51/0.01	0.51/0.01	0.53/0.02	0.74/0.05
	0.5	0.21/0.03	0.21/0.03	0.21/0.04	0.43/0.06
	0.7	**0.03/0.03**	0.06/0.03	0.06/0.07	0.14/0.09
	0.9	0.02/0.06	0.02/0.06	0.02/0.10	0.06/0.12

6.2 Microbenchmarks

In these microbenchmark experiments, we answer the following questions:

1. *How do bin size and γ affect tag detection?*
2. *How does tag reading rate affect tracking performance?*
3. *How does array size affect tracking performance?*
4. *How does array layout affect tracking performance?*
5. *Does the EKF fusion algorithm improve tracking accuracy?*

Configuring bin size and γ. We first perform a sensitivity analysis, examining how the bin size and the threshold γ affect the targeting tag detection accuracy. In this experiment, two volunteers randomly pick up a book from 20 tagged books. Each experiment is repeated 50 times, and we get 100 phase series in total. We also collect 100 phase series of those stationary tags, whose wireless links are blocked or reflected by nearby humans. We then vary the bin size and γ to find an appropriate setting. Table 4 summarizes the False Positive (FP) rate and False Negative (FN) rate with different the bin size and γ. The desired parameter setting should minimize both the FP and FN rates. Suggested by the experimental result, we set the bin size and γ to 0.5 rad and 0.7, which optimize the overall detection performance.

Impact of tag reading rate. We then examine the impact of the tag reading rate. In this experiment, we randomly pick 10 letters from 26 English letters and let a volunteer write these letters using the two-tag array cube (see Figure 16). We change the tag population in the antenna's reading range to control the tag reading rate. Figure 18 shows the recognition accuracy. As expected, the recognition accuracy is sensitive to the average tag reading rate. It achieves an accuracy below 20% when the reading rate is around 8 Hz, and then jumps to around 80% when the average per-tag reading rate is 31 Hz. In contrast, Pantomime achieves constantly high recognition accuracy after boosting the tag reading rate, with the accuracies all above 87%, which verifies the efficacy of our tag reading rate boosting protocol.

Impact of array layout. We then examine the impact of the array layout. As shown in Figure 17, we attach four tags to a plastic cube, forming a rectangle array and a linear array. We randomly choose 10 letters from the English alphabet, and invite a volunteer to write each letter 10 times using these two different tag-array. The letter recognition accuracy is shown in Table 5. As the result shows, Pantomime achieves 92% letter recognition accuracy with the rectangle array. The recognition accuracy drops slightly to 86% with the linear array. The performance gap here indicates that the nature of sparse tag distribution in the rectangle array helps to improve the tracking accuracy.

Impact of array size. We next examine the impact of the array size. Similar to previous experiments, we invite a volunteer to write

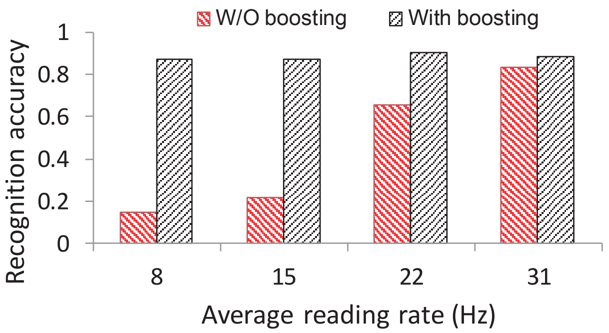


Figure 18— Impact of reading rate.

Table 5— Impact of Array layout.

	Rectangle array	**Linear array**
Recognition accuracy	92%	86%

10 letters using a linear tag array of size 2 to 5. Figure 19 shows the recognition accuracy. As we see, the recognition accuracy increases as we attach more tags to the cube. This is because compared with a small tag array, the large tag array can help average out ranging noise. However, the performance gap between the small array (*e.g.*, two-tag array) and the larger one (*e.g.*, five-tag array) is confined to a small range (below 5% in the experiment), indicating that the large array marginally improves tracking accuracy.

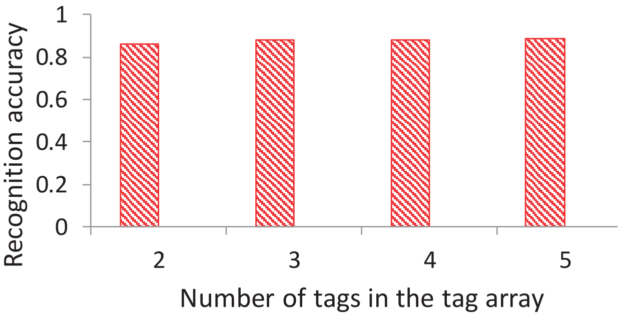


Figure 19— Impact of array size.

Gain of fusion algorithm. This experiment validates the gain of the EKF fusion algorithm for gesture tracking. Four volunteers are asked to use the two-tag array to write 10 letters randomly chosen from the 26 English letters. These written letters are then put into the LipiTk for letter recognition. The result is shown in Figure 20. Pantomime achieves a relatively stable performance over all four volunteers, with a maximum and minimum recognition accuracy of 87% and 83%, respectively. In contrast, when solely running the pseudo ranging-based tracking algorithm, Pantomime achieves undesirable performance over the four volunteers, with a maximum recognition accuracy of around 56%. This result demonstrates that the EKF fusion algorithm significantly improves the tracking accuracy.

6.3 Field study

We conduct field studies in a laboratory environment. In the field study, we attach two tags to a food package to form a stylus (see Figure 16). A volunteer is asked to use this light-weighted stylus

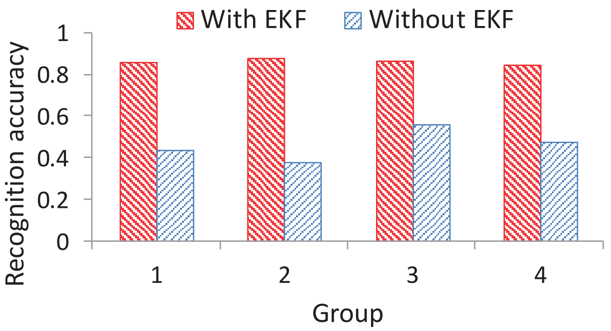


Figure 20— Examine the EKF gain.

to write given characters. We are of particular interest in answering the following questions:

1. *Is* Pantomime *robust to different moving patterns?*
2. *Is* Pantomime *resilient to multi-path effect?*
3. Does tag-to-reader distance affect the tracking accuracy?
4. *Does backing material affect the tracking accuracy?*
5. *How does* Pantomime *perform compared with other state-of-the-art systems?*
6. Can we adapt Pantomime to multi-antenna settings?

Robustness to moving patterns. We first examine whether Pantomime can successfully track different kinds of trajectories for future drawing applications. These trajectories include three kinds of fundamental line drawings [36]: straight lines, wave lines, and jagged lines, as shown in Figure 16. Figure 21 shows the recovered trajectories. As shown, the recovered wave line and jagged line are not self-symmetric due to tracking errors. However, each of them still preserves its unique line pattern, making them distinguishable to human beings. Also, we can see the recovered straight line is still smooth despite deviating from its ground truth at small scale.

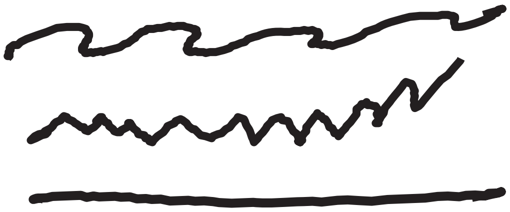


Figure 21— Wave line, jagged line, and straight line, as outputs of Pantomime.

We further show a sample of the recovered trajectory of letters M o b i S y s in Figure 22. The recovered trajectories of letters o and s are quite smooth, manifesting that Pantomime is capable of capturing the continuously moving heading changes. The detailed shapes of these seven letters are well preserved.

Resilience to multi-path reflections. We then examine the impact of both static and dynamic multi-path on the tracking performance.

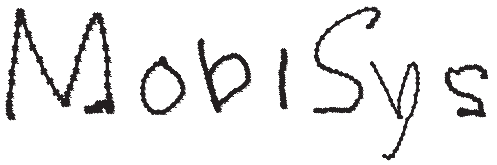


Figure 22— Recovered trajectories for M, o, b, i, S, y, s.

For static multi-path, we put two metal lockers around the experiment field to reflect the backscatter signals. For dynamic multi-path, a volunteer walks around near the experiment field at normal walking speed to generate dynamic reflections. Another volunteer then writes 10 letters and 10 words using the stylus, with each letter written for 10 times. Figure 23 shows the recognition result of these letters and words under different multi-path reflections. The recognition accuracy for both the letter and the word slightly decreases as we introduce the multi-path effect. However, with static multi-path, Pantomime still achieves around 85% and 80% accuracy for letter and word recognition, respectively. With dynamic multi-path, Pantomime achieves around 80% and 71% accuracy for letter and word recognition, respectively. The small performance gap between the no multi-path case and the multi-path cases shows that Pantomime is not very sensitive to both the static and dynamic multi-path.

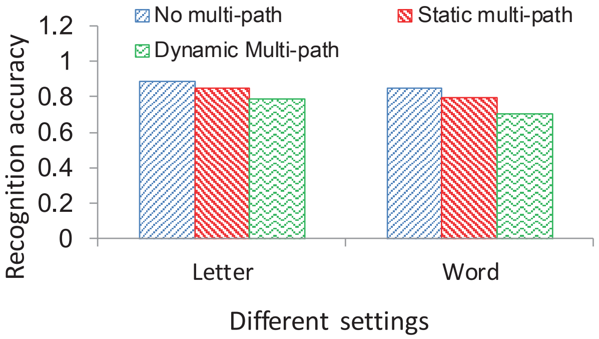


Figure 23— Impact the multi-path effect.

Impact of tag-to-reader distance. We further examine the system performance under different tag-to-reader distance settings. In this trail of experiments, we place an RFID antenna 3m, 2m, 1m, and 0.5m away from a stage. In each tag-to-reader distance setting, we invite a volunteer to stand on this stage and write ten letters and ten words using the plastic stylus. Each letter and word is repeated ten times. Figure 24 shows the result. As shown, when the stage is 0.5m away from the reader antenna, Pantomime achieves 92% letter recognition accuracy and 89% word recognition accuracy, respectively. As we expand the distance between the stage and the reader antenna, the signal will experience relatively intense multi-path propagation. Hence both the letter and the word recognition accuracy drops slightly. Nevertheless, both the letter and the word recognition accuracy still maintain above 80% when the reader is 3m away from the stage.

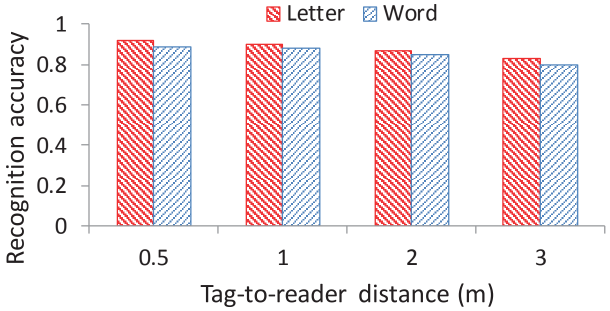


Figure 24— Impact of tag-to-reader distance.

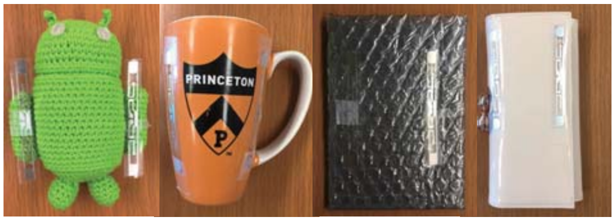


Figure 25— Making a stylus using everyday object.

Impact of backing material. We next examine the impact of backing material on the system performance. In this trail of experiments, we attach a two-tag array on four kinds of items (as shown in Figure 25): a woolen puppet (W), an empty ceramic mug (C), a plastic bubble cushion (P), and a leather wallet (L). We then ask a volunteer to use these four kinds of items as a stylus to write 10 letters in the air, with each letter been written for 10 times. The letter is randomly chosen from 26 English alphabet letters. The letter recognition accuracy of these four kinds of stylus is shown in Table 6. As shown, we can category the recognition accuracy of these five kinds of stylus into two groups. The first groups include the woolen puppet, the bubble cushion and the leather bag, they all achieve a letter recognition accuracy over 84%, and there is no significant difference in their recognition accuracy. Another group only contains the empty ceramic mug, which achieves 2% recognition accuracy. To understand the reason behind this poor performance, we further check the phase readings of this stylus and find that the tag reading rate maintains in a very low level (below 5 Hz), probably due to the strong signal reflection property of this kind of material. Such a low tag reading rate renders the phase readings discontinuous during the stylus movement. Accordingly, Pantomime is unable to recover the profile of the stylus's movement, resulting in a low letter recognition accuracy.

Table 6— Letter recognition accuracy using different stylus.

	W	**C**	**P**	**L**
Letter recognition accuracy	85%	2%	84%	88%

Comparison with state-of-the-art. We further compare Pantomime against two state-of-the-art RFID-based tracking systems[4]: Tagoram [46] and PolarDraw [33]. Tagoram adopts four circularly polarized antennas to track one moving tag using a hologram method. PolarDraw adopts two linearly polarized antennas and simultaneously estimates the moving direction (based on RSS change) and distance (based on phase change) of a moving tag.

For a fair comparison, we conduct two sets of experiments. In the first experiment, we deploy multiple antennas (four for Tago-

[4] We do not compare with Wi-Fi based solutions because most of Wi-Fi based works cannot track the gesture trajectory and associate the gesture with the object that the user is interacting with.

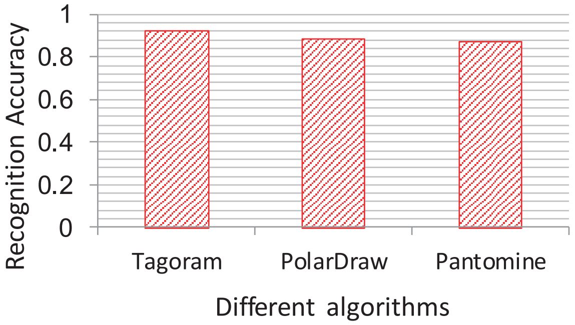


Figure 26— Recognition accuracy achieved by different system in Multi-antenna settings.

ram and two for PolarDraw) to track handwriting and compare the recognition accuray with Pantomime. Figure 26 shows the result. With four antennas, Tagoram achieves an average recognition accuracy of 92%. While for the two-antenna PolarDraw, the average recognition accuracy is 88%. Our single antenna based Pantomime achieves an average recognition accuracy 87%, which is slightly lower than Tagoram but similar to PolarDraw. With natural language processing techniques like [25], we believe recognition accuracy can be improved even further.

In the second experiment, we use one antenna and attach multiple tags (four for Tagoram and Two for PolarDraw) to the food package to compose a multi-tag single antenna tracking system. We then implement Tagoram and Polardraw to track the virtual movement of the antenna in the frame of the reference of the tag array. The recovered trajectories are rotated and fed into the letter recognition software. Figure 27 shows the recognition result. As expected, Tagoram achieves inferior performance due to the low beam resolution. PolarDraw's performance also degrades because of the limited array spacing and the stiff writing style (*i.e.*, writing without rotation). In contrast, Pantomime achieves consistently high recognition accuracy (88%), which is superior to both Tagoram and PolarDraw.

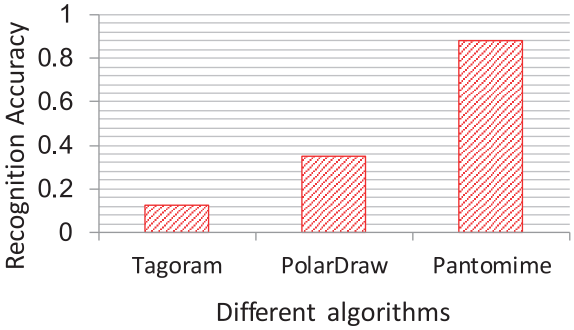


Figure 27— Recognition accuracy achieved by different system in Multi-tag settings.

Multi-antenna Pantomime **vs. multi-tag** Pantomime**.**We further build a multi-antenna version of Pantomime and compare it with the multi-tag Pantomime. The multi--antenna version of Pantomime is composed of three RFID antennas. The antenna spacing is 0.3m. A volunteer stands 1.5m away from the antenna array, writing 10 letters and 10 words using a mark pen (attached with one RFID tag), as shown in Figure 28. For comparison, we let the volunteer write the same letters and words using a plastic cube, where we attach

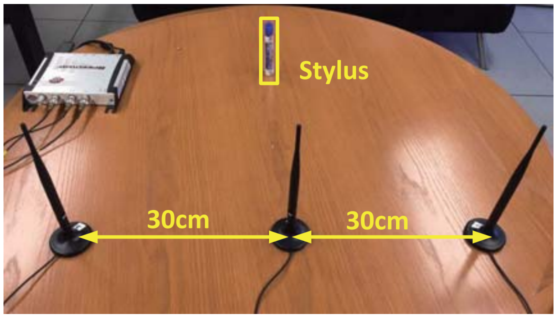


Figure 28— Setups of the multi-antenna version of Pantomime.

three tags to compose a linear tag array. The tag spacing is 4cm. Both the letters and the words are chosen randomly. We summarize the letter and word recognition accuracy in Table 7. As shown, the multi-antenna version of Pantomime (termed as A-Pantomime) achieves a better performance than the tag-array based Pantomime in both letter and word recognition. Specifically, A-Pantomime achieves an overall accuracy of 92% and 89% in letter and word recognition, respectively. In contrast, Pantomime achieves 85% and 82% word and letter recognition accuracy, respectively. The performance gap between A-Pantomime and Pantomime may be due to the asymmetric layout: the spacing between antennas is 0.3m while the spacing between tags is 4cm due to the limited surface space of the item held in hand. However, the performance gap is marginal, which verifies our assumption that a multi-antenna-single-tag system is equivalent to a multi-tag-single-antenna system in performance, yet at much lower cost due to the huge price gap between an antenna and a tag.

Table 7— Recognition accuracy of A-Pantomime and Pantomime.

	Antenna-array	**Tag-array**
Letter recognition accuracy	92%	85%
Word recognition accuracy	89%	82%

7. RELATED WORK

Despite commercial gesture recognition systems such as Kinect-[14] and Leap Motion [18], these vision or infrared based solutions are infeasible for deployment in stores and warehouses, where they are expected to cover a wide range and operate even in non-line-of-sight conditions. Alternatively, researchers have explored context sensing via wireless signals, especially Wi-Fi, with the vision to enable pervasive context-awareness with minimal costs [1, 2, 4, 28, 35, 42, 43, 47]. Although these systems enable device-free sensing in smart homes and offices, they do not associate a gesture with a specific object, which motivates our work. Pantomime is inspired by this trend of wireless sensing, and is particularly related to the following categories of efforts.

Wireless gesture and activity sensing. There has been active research on leveraging wireless signals to track gestures and activities [1, 2, 4, 28, 35, 42, 43], especially using Wi-Fi signals from commodity infrastructure. E-eyes [43] exploits CSI from COTS Wi-Fi network interface cards (NICs) for household activity recognitions. CARM [42] models the relationship between motion speeds and the CSI amplitude, and builds an SVM classifier to recognize location-independent human activities. WiKey [4] decodes keystrokes based on CSI patterns while WiGest [1] leverages RSSI patterns for gesture based device control. WiSee [28] extracts

Doppler frequency shifts from Wi-Fi signals for arm gesture recognition. WiDraw [35] tracks hand motions by measuring angle-of-arrival (AoA) extracted from CSI. Although these systems enable device-free sensing for context-awareness in smart homes and offices, they do not associate a gesture with a specific RFID-tagged object, as Pantomime does.

RFID-based gesture and activity sensing. The ubiquitous deployment of RFID tags makes them promising for interactions and activity recognition with everyday objects. TagFall [30] leverages RFID for fall detection. FEMO [8] extracts Doppler signatures to classify and assess the quality of free-weight activities by attaching RFID tags to dumbbells. ShopMiner [34] harnesses the phase patterns of RFID tags attached to garments to infer shopping behaviors in physical clothing stores. IDSense [20] combines RSSI and phase patterns of RFID tags to identify touching and moving events of the tagged objects. PaperID [19] extract features from RSSI, phase and reading rate for gesture recognition on traces printed or drawn on paper with conductive ink. AllSee [13] powers an RFID-like gesture recognition interface by backscattering ambient wireless signals. While some of them [13, 19] have enabled gesture recognition with a single antenna, they rely on machine learning to recognize a predefined set of gestures. Pantomime is complementary in that Pantomime outputs the trajectory shapes of object motions, which can be further fed into different machine learning algorithms for scalable gesture recognition and other applications such as handwriting tracking and drawing. Pantomime is most close to works on RFID-based gesture tracking [33, 40]. RF-IDraw [40] exploits carefully designed antenna arrays to track the AoA of finger motions. PolarDraw [33] leverages the polarization mismatch of two linearly polarized antennas to track handwriting. Pantomime advances the area by achieving comparable gesture tracking performance using one single antenna, thus yielding the minimal costs for RFID-based gesture tracking.

RFID-based localization and tracking. LANDMARC [26] is the first active RFID based localization scheme by deploying anchor tags. PinIt [39] exploits the multi-path profile of RFID tags attached to books to locate these tagged books in NLOS environments. The adoption of anchor tags for profile matching incurs high cost and the decimeter localization accuracy is unfit for fine-grained gesture tracking. Other works leverage Synthetic Aperatures Radar (SAR) or Inverse SAR (ISAR) to track RFID tags at centimeter-scale accuracies. In [24], the authors introduce a hologram-based scheme to track a moving tag on a transponder. In [27], the authors leverage the tag mobility and adopt ISAR to localize an RFID tag. Tagoram [46] introduces a multi-antenna based hologram method to localize and track tag motion trajectories with error centimeter accuracies. However, all these SAR-based schemes require multiple bulky and expensive RFID antennas to function, impeding their deployment in stores and warehouses. In addition, both SAR and ISAR based schemes track the RFID tag by calculating its position for each time point. Conversely, Pantomime tracks the tag array via the changes of the tag-to-reader distances without pinpointing its absolute position. Tagyro [44] attaches multiple tags on an object to track its 3D *orientation* using differential phase-based hologram. However, it does not consider the limited tag reading and insufficient antenna coverage issues. Pantomime is designed for the fine-grained gesture tracking and we show that hologram-based methods fail to achieve this goal due to the space limitation of the object and the antenna crisis.

8. DISCUSSION

We discuss limitations and opportunities for improvement.

Identifying the beginning of the gesture. Pantomime needs to detect the beginning of the gesture in the phase stream for gesture tracking. The technique we propose is based on instructing users to hold the tagged object for a short period before performing a gesture, which results in relatively stable phase readings that naturally separate the pick-up action and the gestures (see Figure 14). Similar to the activity detection algorithm used in [8], Pantomime uses K-L divergence [15] to identify these short resting periods, and tracks the gesture hereafter.

Support for multiple users. Our current prototype supports single user gesture tracking per antenna, because when a user performs gestures in front of an antenna, Pantomime blocks all the remaining tags within its reading range from replying. Pantomime can support multiple users by letting the reader interrogate both target and non-target tags at different sampling rates, *e.g.*, sampling all target tags at a high frequency, yet sampling only one tag attached to each of the remaining objects at a lower rate for multiple user detection. However, this requires modifications of the MAC protocol and we leave it for future work.

Trade-off between multi-tag and multi-antenna systems. Recent research on MIMO RFID readers [17] has lowered down the cost of a multi-antenna system, yet an RFID antenna can still cost 100 USD, more than 600 times the cost of an RFID tag (15 cent). Thus the additional cost to deploy 3 more antennas to support four-coverage ratio can support attaching a 2-tag array to around 2000 objects. Since the typical effective reading area of a ceiling-mounted RFID antenna is about 10m^2, it is reasonable that there will be fewer than 1500 objects placed on shelves in stores and warehouses within this effective reading area. Therefore, a multi-tag system like Pantomime is a more cost-effective alternative to current multi-antenna solutions. Multi-tag solutions also have the advantages of improving the item detection probability [6], which is desirable in store and warehouse management to reduce potential financial losses.

Practicality of hardware solutions. Note that Pantomime serves as a proof-of-concept to show the benefits of attaching multiple tags on each object. Our aim is to provide a new object interaction mechanism for stores and warehouses, where the adoption of low-cost, widely accessible infrastructure is essential. In principle, recent research on multi-antenna RFID tags [12] is also applicable in our scenarios with customized MAC protocols, and will avoid the challenge of reading rate. We envision the future commercialization of multi-antenna RFID tags will further decrease the cost of multi-tag systems like Pantomime.

9. CONCLUSION

In this paper, we present the design, implementation and evaluation of Pantomime, a gesture tracking system with a single RFID antenna. Through both comprehensive experiments and real-world case studies, we demonstrate that Pantomime achieves comparable gesture tracking accuracy to the state-of-the-art multi-antenna methods, showing the potential to support various gesture-based object interactions with a minimal deployment cost. In the future we plan to extend Pantomime to multi-user cases and deploy it in an entire warehouse/mall for large-scale and long-term studies.

Acknowledgements

We thank our shepherd Shyam Gollakota and the anonymous reviewers for their helpful feedback. This material is based upon work supported by the National Science Foundation under Grant No. 1617161.

References

[1] H. Abdelnasser, M. Youssef, K. A. Harras. Wigest: A ubiquitous wifi-based gesture recognition system. *INFOCOM*, 2015.

[2] F. Adib, Z. Kabelac, D. Katabi, R. Miller. 3D Tracking via body radio reflections. *NSDI*, 2014.

[3] S. Agrawal, I. Constandache, S. Gaonkar, R. Roy Choudhury, K. Caves, F. DeRuyter. Using mobile phones to write in air. *MobiSys*, 2011.

[4] K. Ali, A. X. Liu, W. Wang, M. Shahzad. Keystroke recognition using WiFi signals. *MobiCom*, 2015.

[5] Avery-Dennison UHF RFID inlay *AD-227m5*. Web page.

[6] L. Bolotnyy, S. Krize, G. Robins. The practicality of multi-tag rfid systems. *IWRT*, 2007.

[7] EPC radio-frequency identity protocols class-1 generation-2 uhf rfid protocols for communications at 860 MHz - 960 MHz. .

[8] H. Ding, L. Shangguan, Z. Yang, J. Han, Z. Zhou, P. Yang, W. Xi, J. Zhao. Femo: A platform for free-weight exercise monitoring with RFIDs. *SenSys*, 2015.

[9] S. Gupta, D. Morris, S. Patel, D. Tan. Soundwave: using the doppler effect to sense gestures. *CHI*, 2012.

[10] LipiTk: online hand writing recognition toolkit. Web page.

[11] ImpinJ Speedway R420 reader. Web page.

[12] N. C. Karmakar, M. Zomorrodi, C. Divarathne. *Advanced Chipless RFID: MIMO-Based Imaging at 60 GHz-ML Detection*. John Wiley & Sons, 2016.

[13] B. Kellogg, V. Talla, S. Gollakota. Bringing gesture recognition to all devices. *NSDI*, 2015.

[14] Microsoft Kinect. Web page.

[15] Kullback-Leibler Divergence. Web page.

[16] RFMAX S9028PCR (LP) outdoor RFID antenna. Web page.

[17] R. Langwieser, C. Angerer, A. L. Scholtz. A uhf frontend for mimo applications in rfid. *RWS*, 124–127, 2010.

[18] Leap Motion. Web page.

[19] H. Li, E. Brockmeyer, E. J. Carter, J. Fromm, S. E. Hudson, S. N. Patel, A. Sample. Paperid: A technique for drawing functional battery-free wireless interfaces on paper. *CHI*, 2016.

[20] H. Li, C. Ye, A. P. Sample. Idsense: A human object interaction detection system based on passive uhf rfid. *CHI*, 2015.

[21] X. R. Li, V. P. Jilkov. Survey of maneuvering target tracking: dynamic models. *AeroSense 2000*, 2000.

[22] T. Liu, Y. Liu, L. Yang, Y. Guo, C. Wang. Backpos: High accuracy backscatter positioning system. *IEEE TMC*, **15**(3), 586–598, 2016.

[23] Y. Ma, X. Hui, E. C. Kan. 3d real-time indoor localization via broadband nonlinear backscatter in passive devices with centimeter precision. *MobiCom*, 2016.

[24] R. Miesen, F. Kirsch, M. Vossiek. Holographic localization of passive uhf rfid transponders. *RFID*, 2011.

[25] A. Nasr, F. Béchet, J.-F. Rey, B. Favre, J. Le Roux. Macaon: An nlp tool suite for processing word lattices. *ACL*, 2011.

[26] L. M. Ni, Y. Liu, Y. C. Lau, A. P. Patil. LANDMARC: indoor location sensing using active RFID. *Wireless networks*, 2004.

[27] A. Parr, R. Miesen, M. Vossiek. Inverse sar approach for localization of moving rfid tags. *RFID*, 2013.

[28] Q. Pu, S. Gupta, S. Gollakota, S. Patel. Whole-home gesture recognition using wireless signals. *MobiCom*, 2013.

[29] L. G. Roberts. Aloha packet system with and without slots and capture. *ACM SIGCOMM Computer Communication Review*, 1975.

[30] W. Ruan, L. Yao, Q. Z. Sheng, N. Falkner, X. Li, T. Gu. Tagfall: Towards unobstructive fine-grained fall detection based on uhf passive rfid tags. *MOBIQUITOUS*, 2015.

[31] M. Shahzad, A. X. Liu. Every bit counts: fast and scalable rfid estimation. *MobiCom*, 2012.

[32] L. Shangguan, K. Jamieson. The design and implementation of a mobile rfid tag sorting robot. *MobiSys*, 2016.

[33] L. Shangguan, K. Jamieson. Leveraging electromagnetic polarization in a two-antenna whiteboard in the air. *CoNEXT*, 2016.

[34] L. Shangguan, Z. Zhou, X. Zheng, L. Yang, Y. Liu, J. Han. Shopminer: Mining customer shopping behavior in physical clothing stores with cots RFID devices. *SenSys*, 2015.

[35] L. Sun, S. Sen, D. Koutsonikolas, K.-H. Kim. WiDraw: Enabling hands-free drawing in the air on commodity WiFi devices. *MobiCom*, 2015.

[36] Line Drawing: A Guide for Art Students. Web page.

[37] Tag price. Web page.

[38] J. Wang, F. Adib, R. Knepper, D. Katabi, D. Rus. RF-compass: Robot object manipulation using RFIDs. *MobiCom*, 2013.

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

[40] J. Wang, D. Vasisht, D. Katabi. RF-IDraw: Virtual touch screen in the air using RF signals. *SIGCOMM*, 2014.

[41] L. Wang, B. A. Norman, J. Rajgopal. Placement of multiple rfid reader antennas to maximise portal read accuracy. *International Journal of Radio Frequency Identification Technology and Applications*, 2007.

[42] W. Wang, A. Liu, M. Shahzad, K. Ling, S. Lu. Understanding and modeling of WiFi signal based human activity recognition. *MobiCom*, 2015.

[43] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, H. Liu. E-eyes: device-free location-oriented activity identification using fine-grained WiFi signatures. *MobiCom*, 2014.

[44] T. Wei, X. Zhang. Gyro in the air: tracking 3d orientation of batteryless internet-of-things. *MobiCom*, 2016.

[45] J. Xiong, K. Jamieson. ArrayTrack: A fine-grained indoor location system. *NSDI*, 2013.

[46] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, Y. Liu. Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices. *MobiCom*, 2014.

[47] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, Y. Liu. Smokey: Ubiquitous smoking detection with commercial wifi infrastructures. *INFOCOM*, 2016.