RF-Dial: Rigid Motion Tracking and Touch Gesture Detection for Interaction via RFID Tags

Yanling Bu, Student Member, IEEE, Lei Xie, Member, IEEE, Yinyin Gong, Chuyu Wang, Member, IEEE, Lei Yang, Member, IEEE, Jia Liu, Member, IEEE, and Sanglu Lu, Member, IEEE

Abstract—With the rising demands for novel human-computer interaction approaches in the 2D plane, a number of intelligent devices come into being. For example, Microsoft Surface Dial supports simple clicks and rotations for the interaction with computer. However, these approaches are dedicated devices, and they might require batteries or have limited functions. In this paper, we propose RF-Dial to realize a light-weight, battery-free and functional 2D human-computer interaction solution via commercial off-the-shelf (COTS) passive RFID tags. What RF-Dial shines is that it can easily turn an ordinary object, e.g., a board eraser, into an intelligent interaction device. By deploying a tag array on the side face of the object together with a dipole tag on the top face, RF-Dial cannot only track the rigid motion of the object but also detect the touch gesture of a user on the surface of the object, including translation, rotation, click, press and hold, and swipe. To do the motion tracking, RF-Dial builds a phase-based model that captures the translation and the rotation of the tagged object simultaneously, by jointly exploiting the information of phase variations and the topology of the tag array. To detect the touch gesture, RF-Dial builds an RSSI-based model that uses the impact of the touching finger on the tag antenna’s impedance to estimate the touch position in real time, which is robust to environmental factors like position or orientation. We implemented a prototype of RF-Dial with commodity RFID devices. Extensive experiments show that RF-Dial achieves an accurate rigid motion tracking, with a small error of 0.6cm for the translation tracking, and a small error of 1.9 degrees for the rotation estimation. Besides, RF-Dial can also detect the touch gesture accurately, as the 90 percent of touch position errors are less than 2.09mm.

Index Terms—RFID, human-computer interaction, tag array, translation, rotation, coupling effect, touch gesture.

1 INTRODUCTION

In modern times, the widely used approaches for the human-computer interaction (HCI) are operated in the 2D plane, like the touch screen and the mouse. By moving or stroking these interaction devices, users can access the objects in the computer and manipulate them conveniently. With the rise of the computer aided art design and other novel applications, a number of intelligent devices have come into being as the response to the demand for brand-new 2D interaction solutions. For example, Microsoft Surface Dial [1] emerged in 2016, supporting simple clicks and rotations for the natural and friendly interaction.

The latest HCI approaches are mainly based on the computer vision or sensors. For computer vision-based approaches, they use cameras to monitor the movement of limbs or fingers. However, they are mainly limited by privacy concerns, the light condition and the viewing angle. For sensor-based approaches, they use commercial-off-the-shelf (COTS) sensors like inertial sensors to track the movement of devices. Their main constraints are the limited battery life and the high hardware cost. Thankfully, RFID provides the battery-free sensing technology to enable novel HCI designs [2-18]. RFID can even work in the non-line-of-sight situation due to its backscatter communication. Therefore, we hope to use RFID to answer such a question: “Is it possible to design a battery-free and light-weight solution to the 2D human-computer interaction, thereby even an ordinary object can be easily turned into an intelligent interaction device?”

In this paper, we propose RF-Dial to realize a novel 2D human-computer interaction solution via COTS passive RFID tags. We attach a tag array to the side face together with one tag on the top face of an object, denote them as movement tags and the touch tag, respectively. As shown in Fig. 1, we deploy two RFID antennas orthogonally to realize our vision. Specifically, we continuously track the rigid motion of the tagged object with movement tags, including the translation and the rotation simultaneously, and detect the touch gesture with the touch tag, including the click, the press and hold, and the swipe. In this way, an ordinary object such as a candy box can be turned into an intelligent interaction device. For example, we can realize an functional drawing application with RF-Dial. It tracks the translation of movement tags to draw lines, and adjusts the line color automatically by the rotation just during the drawing process; The touch tag functions as buttons and sliders, receiving the user’s commands, i.e., adjusting the line color and width, as shown in the case study (Section 9.4). Technically, based on RF-signals from movement tags, we build a rigid transformation model to reflect the relationship between the motion of the tagged object and the corresponding phase variations of each movement tag in the tag array. As the movement tags form a tag array with the fixed topology, we can derive the translation and the
rotation of the tagged object for each snapshot based on the rigid transformation model. Note that, the phase contours of RF-signals vary at different positions in the scanning area, the relationship between the tag movement and the phase variation is different, regarding to which we split the effective scanning area into linear region and non-linear region. Meanwhile, based on RF-signals from the touch tag, we build an RSSI-based model to depict the relationship between the touch position on the tag and the corresponding RSSI variation of received RF-signals. According to the RSSI-based model, the RSSI deviation is mainly related to the impedance change when touching the tag, so it is position-independent and orientation-insensitive. Consequently, we can rely on only one general RSSI deviation template to accurately and robustly determine the touch position, without the known start touch position or the fixed tag deployment, among the whole monitoring area.

There are three key challenges to realize RF-Dial. 1) How to estimate the rigid motion of the tagged object based on RF-signals of tags, including the translation and rotation simultaneously, is a key problem. To tackle this challenge, we build a rigid transformation model to reflect the relationship between the motion of the tagged object and the corresponding phase variations of each movement tag in the tag array. As the topology of the movement tag array is fixed, we are able to decompose the rigid motion of the tagged object referring to the phase variations of at least two movement tags, and then derive the translation and the rotation of the tagged object for each snapshot during the motion. 2) How to address the variation of phase contours at different positions in the effective scanning area is a key problem. Our empirical study shows that the phase contours are close to concentric circles with the antenna at the center. Hence, even for the same rigid motion of the tagged object, the antenna could collect different phase variations at different positions. To tackle this challenge, regarding to the relationship between the tag movement and the phase variation, we split the whole scanning area into linear region and non-linear region. Specifically, the tag movement in the linear region is linear to the phase variation, thus we can extract the tag movement based on the phase variations detected from the two orthogonal antennas. While in the non-linear region, we locate the tag first, then extract the tag movement based on the phase contours at the tag’s position. 3) How to obtain the absolute touch position of the tag when the tag moves to any position with different orientations within the monitoring area is a key problem. Existing work like [19] leverages the phase variation to detect the touch gesture, however, the phase is sensitive to the position and orientation of the tag, so it can only track the touch position with the known start touch point of a fixed tag. To tackle the challenge of determining the absolute touch position, we observe that the RSSI deviation during the swipe is position-independent and orientation-insensitive. We explore the signal variation when touching different positions on the tag, and build an RSSI-based model to verify the robustness of the RSSI deviation, which is mainly related to impedance change due to touch. Hence, based on the RSSI deviation, we can identify the absolute candidate touch positions. Note that, as the used linear tag has the dipole antenna, the RSSI variation during the swipe across the tag forms a symmetric Ω-wave pattern. To eliminate such ambiguity of touch positions, we use half of the tag as buttons and the other half as the slider. Referring to the RSSI variation, it is easy to determine which gesture is performed, thereby we can identify which part of the tag is touched and further derive the unique touch position. By tracking the consecutive touch positions of the swipe, we are able to estimate the swipe direction and distance.

Overall, we make the following three main contributions. First, we propose a novel interaction scheme via RFID technology, supporting the rigid motion tracking and the touch gesture detection. An ordinary object can be turned into an intelligent HCI device via attaching a tag array on the side face together with one linear tag on the top face, denoted as movement tags and the touch tag, respectively. Second, we build a phase-based model to reflect the relationship between the motion of tagged object and the corresponding phase variations of movement tags in the array. We also build an RSSI-based model to depict the relationship between the touch position and the corresponding RSSI deviation of the touch tag. Third, we implemented a prototype system of RF-Dial and evaluated its performance in the real environment. Extensive experiments show that RF-Dial achieves an accurate rigid motion tracking, with a small error of 0.6cm for translation tracking, and a small error of 1.9 degrees for rotation estimation. Besides, RF-Dial can also detect the touch gesture accurately, as the 90 percent of touch position errors are less than 2.09mm.

2 Related work

RFID-based Localization: A straightforward solution for RFID-based human-computer interaction is to utilize RFID localization schemes to directly locate tagged objects [2–6, 20–26]. State-of-the-art systems mainly use phase values for the accurate localization [2–4, 6, 22, 23]. PinIt [2] uses multi-path profiles of tags to accurately locate tags with the synthetic aperture radar created via the antenna motion. Rather than the absolute localization, STPP [6] is the first work to tackle 2D relative localization, which uses the spatial-temporal dynamics in the phase profiles to identify the relative positions of tags. More than only using the phase information, RFind [20] leverages the complete physical properties of RF-signals to realize the ultra-wideband localization. RFind is capable of emulating over 220MHz of bandwidth without changing the tag and remains compliant with current regulations. However, most approaches figure out the absolute positions of tags in a separate manner, whereas RF-Dial aims to track the movement of the tag array in a comprehensive manner. By referring to the fixed topology of tag array, RF-Dial can accurately track the rigid transformation of tagged object, including the translation and rotation simultaneously.
RFID-based Motion Tracking: Prior RFID-based motion tracking systems propose various approaches for the trajectory tracking [7–12, 27] and orientation tracking [13, 14, 28]. Representative work like RF-IDraw [7] and PolarDraw [8] use a single tag to reconstruct the handwriting, regarding the tag as a mass point for the motion tracking. TagCompass [14] uses a single tag to determine the orientation and position of the tagged object based on polarization properties of RF waves. Further, recent work uses the tag array to track the trajectory or orientation of the moving object. Specifically, Pantomime [9] enables the accurate trajectory tracking of the tagged object with a tag array, using a multiple tag single antenna system. Tagyro [13] realizes the 3D orientation tracking with an array of RFID tags, by converting the real-time phase offsets between tags into the orientation angle. However, these approaches track either the trajectory or the orientation of the moving object, without detecting the translation and rotation of the tagged object simultaneously. Tagball [16] is the closest work to RF-Dial, which studies the motion behavior, including the translation and rotation, of a ball attached with a tag array. However, Tagball solves the problem by the absolute localization on multiple tags. Specifically, it first estimates the absolute positions of multiple tags via the phase values, and then figures out the translation and rotation of the tagged object based on the estimated positions of tags. Hence, the localization errors are further introduced to the estimation of the translation and rotation. Therefore, it requires plenty of tags, i.e., 12 tags in total, to provide enough data to the Extended Kalman Filter-based tracking model to guarantee the tracking accuracy. In comparison, RF-Dial tracks the translation and rotation of tagged objects simultaneously by directly referring to phase variations from at least two tags, thus it achieves more accuracy in the motion tracking.

RFID-based Touch Sensing: Besides the traditional studies on localization or tracking, recently researchers also have studied the tag’s physical change when the conductor touches the tag, i.e., the liquid and human beings, and tried to utilize such characteristics for applications like liquid detection [29] and touch interaction [19, 30, 31]. PaperID [30] provides the capability to use COTS tags to sense the finger touch, swipe touch, and other gestures by the support vector machine (SVM). But this sensing capability is very coarse, i.e., it only can detect whether the touch happens but cannot determine where the exact touch position is on the tag. Meanwhile, the machine learning algorithm requires the high training overhead. RIO [19] observes that when a human finger touches the tag, the tag’s impedance changes, which causes the phase change among received RF-signals correspondingly. Based on the phase variation template, it tracks the finger position during a swipe using the segmental dynamic time warping (SDTW) method. However, RIO can only work with the known start finger position, otherwise it cannot use the SDTW to track the finger position. Also, the phase is sensitive to the distance from the tag and antenna, so the template is required to be updated when the tag is moved to different positions. Whereas, RF-Dial is designed to track the absolute finger position without the known start position. Our solution is based on the RSSI deviation resulted from the impedance change when the touch happens, which is position-independent and orientation-insensitive. With only one general RSSI deviation template, we can accomplish the light-weight and fine-grained touch detection accurately among the whole operating area.

3 System Overview

RF-Dial is designed to provide two functions: one is to track the rigid motion of the tagged object, including the translation and the rotation, the other is to detect the touch gesture on one linear tag, such as the click, the press and hold, and the swipe. The basic idea is to use a tag array attached on the side face of an object to track its movement, and use another single tag attached on the top face of the object to detect the touch gesture, as shown in Fig. 1. Suppose the tags in a array that track the motion are movement tags, the single tag that detects the touch gesture is the touch tag. To reduce the mutual coupling between tags, we separate the touch tag and the movement tags with enough distances, i.e., more than 6cm of height difference. Fig. 2 illustrates the system framework. After receiving the raw RF-signals from tags, we first determine which kind of the operation, i.e., touch or movement. The intuition is that for the touch gesture, the signal of the touch tag changes significantly but the signals of movement tags keep steady, while for the motion, the signal of either the touch tag or the movement tags changes along with the continuous motion, as shown in TABLE 1. Therefore, by extracting the features of the signal variation of tags, such as the phase or RSSI difference in time intervals, we can determine which operation the user performs using the random forest algorithm. Details are shown in Section 7.5.

As shown in Fig. 2, there are two main function modules to realize our goal of the rigid motion tracking and the touch gesture detection separately. In the following, we first build...
a phase-based model for the rigid motion tracking in Section 4, and propose a tracking solution based on the motion model in Section 5. Then, we build an RSSI-based model for the touch gesture detection in Section 6, and propose the corresponding solution to realize the absolute touch position detection robustly and accurately in Section 7.

4 RIGID MOTION TRACKING

In this section, we propose the definition of linear region and non-linear region based on phase contour variations. Then, we illustrate the relationship between the tag movement and phase variation in different regions. Further, we model the rigid motion to decompose the translation and rotation.

4.1 Linear Region and Non-linear Region

In RFID systems, the RF phase is a common attribute of the wireless signal, ranging from 0 to 2π. It is very sensitive to the tag-antenna distance. Suppose the distance between the tag and the antenna is d, so the signal traverses a distance of 2d in the backscatter communication. Then, the phase provided by the antenna can be expressed as:

\[ \theta = \left( \frac{2\pi}{\lambda} \times 2d + \mu \right) \mod 2\pi, \]

where \( \lambda \) is the wavelength, \( \mu \) represents the phase offset caused by the diversity of hardware characteristics. According to the phase expression in Eq. (1), besides the diversity term, the phase value mainly depends on the distance between the tag and the antenna. Therefore, the phase contours should form concentric circles with the antenna at the center in an ideal situation. We thus conduct an experiment to validate the above hypothesis. We build a 2D coordinate system according to the parallel direction (X-axis) and perpendicular direction (Y-axis) of the antenna, and set the origin \((0, 0)\) at the center of the antenna. Then, we collect phase values in a rectangle space in front of the antenna, ranging from \(-100cm\) to \(100cm\) along the X-axis and from \(100cm\) to \(180cm\) along the Y-axis, the step is 5cm. The collected phase values are plotted in Fig. 3. Based on the experiment results, we have the following observation:

Observation 1: The phase contours are very close to concentric circles with the antenna at the center. Besides, in the central beam region marked with blue lines in Fig. 3, the phase contours are almost parallel to each other and stretching along the X-axis. That is, in this region, the phase can be regarded as linearly related to the perpendicular distance from the tag to the antenna plane.

As shown in Fig. 4, assume that two antennas \( A_x \) and \( A_y \) are deployed in a mutually orthogonal manner and separated with a fairly large distance. Then, according to Observation 1, in the intersection area of the central beams of the two antennas, the displacement of a tag along the X-axis and Y-axis should be linear to the phase variations received by the antenna \( A_x \) and \( A_y \), respectively. We thus denote this intersection region as the linear region. The size of the linear region depends on two factors: the central beam region of each antenna, and the perpendicular distance to each antenna. Specifically, the central beam region relies on the tolerance of the small displacement along the horizontal direction for the antenna. E.g., if the small displacement along the X-axis at one position incurs the phase difference of antenna \( A_y \) smaller than a certain value, we think this position belongs to the central beam of \( A_y \). Also, as shown in Fig. 3, the width of the central beam changes with the perpendicular distance. The larger the perpendicular distance is, the wider the central beam is at that distance, the central beam region is trapezoidal. Therefore, taking Fig. 4, if setting the small displacement as 5cm, the phase different threshold as 0.4 radians, for the position with the distance of 1.2m from two antennas, the size of the linear region is about \(0.6 \times 0.6m^2\), centered at that position. Apart from the linear region, the phase variations in the other scanning area are not linear to the displacements along either the X-axis or Y-axis, they depend on the exact tag position instead. Hence, we denote the other area as the non-linear region.

4.2 Rigid Transformation

During the continuous movement of an object, its position and orientation are changing all the time. For a rigid body, such change of the position and orientation in the 2D space can be defined by the rigid transformation \([R, S]\), where \( R \) is a \(2 \times 2\) rotation matrix and \( S \) is a \(2 \times 1\) translation matrix. Here, the rotation means a circular movement that the device rotates around a rotation center, and the translation means a linear movement that every point of the device moves with the same displacement. As the continuous movement of an object consists of a series of instant movements at different time, we denote the instant movement as the micro-movement, each micro-movement can be expressed with the rotation and translation. Thus, we can use the rigid transformation to depict the micro-movement.

By attaching a tag array on an object, it is possible to track the rigid transformation based on the movement of each tag in the array. Note that, different from the rigid body, i.e., the tagged object, the tag attached on the object actually represents a single point of the object, so its movement can be regarded as the particle movement, which only has the translation rather than the rotation. E.g., assume an object is attached with a tag array with the layout of rectangle, as shown in Fig. 5, the tags are denoted as solid points on the rectangle. For any micro-movement in the continuous
movement, it can be intuitively observed that, the rigid transformation of the tagged object, including the translation and rotation, can be derived from the movement of different tags.

### 4.3 Model of Tag Movement and Phase Variation

According to Observation 1, the phase contours can be depicted as concentric circles with the antenna at the center. Thus, we can build a polar coordinate system by setting the center of the antenna as the origin. Then, given a tag movement $s$, we can further depict the relationship between the phase variation and the movement $s$ in this polar coordinate system. As shown in Fig. 6, the antenna is deployed at position $A$, we use the vector $s$ to denote the tag movement, the starting point of $s$ is $P$. Besides, we use the vector $I$ to denote the polar axis $AP$, and use $\gamma$ to denote the angle between $s$ and $I$. Thus, if we use $\Delta d$ to denote the projection of $s$ on the polar axis, then $\Delta d = ||s|| \cos \gamma$.

Note that, for any tag movement in the micro-movement, its moving distance should be smaller than half-wavelength, i.e., $||s|| \leq \frac{\lambda}{2} \approx 16.4\text{cm}$. According to Eq. (1), by offsetting the constant diversity term, the phase variation $\Delta \theta$ caused by $s$ is as follows:

$$\Delta \theta = \frac{2\pi}{\lambda} \times 2 \Delta d = \frac{2\pi}{\lambda} \times 2 ||s|| \cos \gamma.$$  

Meanwhile, as $I \cdot s = ||I|| \cdot ||s|| \cos \gamma$, according to Eq. (2),

$$\frac{1}{||I||} \cdot s = \frac{\lambda}{4\pi} \Delta \theta.$$  

Note that, $\frac{1}{||I||}$ is a normalized vector of $I$, it depends on the position of $P$ relative to $A$. Assume $s = (\Delta x, \Delta y)$, $\frac{1}{||I||} = \langle x_l, y_l \rangle$, then, according to Eq. (3),

$$\begin{cases} x_l \Delta x + y_l \Delta y = \frac{\lambda}{4\pi} \Delta \theta, \\ x_l^2 + y_l^2 = 1. \end{cases}$$

Then, to compute the tag movement $s = (\Delta x, \Delta y)$ according to the phase variations, we investigate their relationships in the linear region and non-linear region, respectively.

#### 4.3.1 Tag Movement in the Linear Region

In the linear region, the phase variations detected from the two orthogonally deployed antennas are linear to the tag’s moving distances along the two orthogonal axes, respectively. E.g., as shown in Fig. 4, antenna $A_x$ detects the phase variation of the tag movement along the $X$-axis, whereas antenna $A_y$ detects the phase variations of the tag movement along the $Y$-axis. Let $\Delta \theta_x$ and $\Delta \theta_y$ be the phase variations from antenna $A_x$ and $A_y$, respectively, so the tag movement $s$ is computed as follows:

$$\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} \frac{\lambda}{4\pi} \Delta \theta_x \\ \frac{\lambda}{4\pi} \Delta \theta_y \end{bmatrix}.$$  

#### 4.3.2 Tag Movement in the Non-linear Region

In the non-linear region, since the corresponding phase variations are not linear to the tag movement, we need to figure out their relationship according to the geometric property. Given the phase variations $\Delta \theta_x$ and $\Delta \theta_y$ respectively collected from the two orthogonally deployed antennas $A_x$ and $A_y$, according to Eq. (4), we have:

$$\begin{cases} x_l \Delta x + y_l \Delta y = \frac{\lambda}{4\pi} \Delta \theta_x, \\ x_l \Delta x + y_l \Delta y = \frac{\lambda}{4\pi} \Delta \theta_y, \end{cases}$$

where $\langle x_{l,s}, y_{l,s} \rangle$ and $\langle x_{l,e}, y_{l,e} \rangle$ denote the normalized vector for the polar axis $AP$ from the antenna $A_x$ and $A_y$, respectively. Therefore, as long as the starting position of movement $s$, i.e., $P$, is known, the values of $\langle x_{l,s}, y_{l,s} \rangle$ and $\langle x_{l,e}, y_{l,e} \rangle$ can be figured out. Then, by solving the linear equations in Eq. (6), we can directly compute $[\Delta x, \Delta y]^T$.

### 4.4 Model of Rigid Motion Decomposition

As aforementioned, during the continuous moving process of the rigid body, the micro-movement can be defined by the rigid transformation including the rotation and translation. Meanwhile, the tag movement can be regarded as the particle movement only with the translation. Therefore, we investigate the relationship between the tag movement and the rigid transformation of the tagged object, i.e., translation, rotation, and translation with rotation, respectively.

#### 4.4.1 Translation

The translation means a linear movement that every point of the device moves with the same displacement. Suppose a rigid body is attached with a tag array $T$, when the center of the rigid body translates from position $P_s$ to position $P_e$ at each tag $T_i$ in the tag array has the same translation $S = [s_x, s_y]^T$. Let $[x_{i,s}, y_{i,s}]^T$ and $[x_{i,e}, y_{i,e}]^T$ be the coordinates of tag $T_i$ when the rigid body is at position $P_s$ and $P_e$, respectively, then:

$$\begin{bmatrix} x_{i,e} \\ y_{i,e} \end{bmatrix} = \begin{bmatrix} x_{i,s} \\ y_{i,s} \end{bmatrix} + S.$$  

Fig. 7(a) shows an example of the translation when the rigid body is attached with a rectangle tag array.

#### 4.4.2 Rotation

The rotation means a circular movement that the device rotates around a rotation center. Suppose a rigid body is attached with a tag array $T$, when the rigid body rotates around a rotation center $P_a$ by the angle of $\alpha$, all the tags should have the same rotation angle. Specifically, let $[x_{i,s}, y_{i,s}]^T$ and $[x_{i,e}, y_{i,e}]^T$ be the coordinates of tag $T_i$ when the rigid body starts rotation and ends rotation, respectively, let $(x_a, y_a)$ be the coordinates of rotation center $P_a$, then:

$$\begin{bmatrix} x_{i,e} - x_a \\ y_{i,e} - y_a \end{bmatrix} = R \begin{bmatrix} x_{i,s} - x_a \\ y_{i,s} - y_a \end{bmatrix},$$

where $R$ is a rotation matrix $\begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix}$, representing the counter-clockwise rotation of angle $\alpha$. Fig. 7(b) shows an example of the rotation when the rigid body is attached with a rectangle tag array.
4.4.3 Translation with Rotation

According to the definition of the rigid transformation, any arbitrary rigid body motion can be decomposed into the combination of rotation and translation. Suppose a rigid body is attached with a tag array \( T_r \) when the center of the rigid body translates from the position \( P_s \) to the position \( P_e \), the rigid body also rotates around a local rotation center \( P_a \) by the angle of \( \alpha \), the local rotation center has the same translation as the rigid body as well. Without loss of generality, we can model the process of the rigid body motion into two successive steps one after the other, i.e., performing the rotation first and then the translation. Specifically, the rigid body first rotates around \( P_a \) by the angle of \( \alpha \), then it translates from position \( P_s \) to position \( P_e \). According to Eq. (7) and Eq. (8), let \( [x_{i,s}, y_{i,s}]^T \) and \( [x_{i,e}, y_{i,e}]^T \) be the coordinates of tag \( T_i \) when the rigid body starts moving and ends moving, respectively, let \( (x_a, y_a) \) be the coordinates of local rotation center \( P_a \) when the rigid body starts moving, then:

\[
\begin{bmatrix}
    x_{i,e} - x_a \\
    y_{i,e} - y_a
\end{bmatrix} = \mathbf{R} \begin{bmatrix}
    x_{i,s} - x_a \\
    y_{i,s} - y_a
\end{bmatrix} + \mathbf{S}.
\]  

(9)

According to Eq. (9), the movement of tag \( T_i \), i.e. \( [\Delta x_i, \Delta y_i]^T \), can be decomposed into the following components:

\[
\begin{bmatrix}
    \Delta x_i \\
    \Delta y_i
\end{bmatrix} = \begin{bmatrix}
    x_{i,e} - x_{i,s} \\
    y_{i,e} - y_{i,s}
\end{bmatrix} = (\mathbf{R} - \mathbf{I}) \begin{bmatrix}
    x_{i,s} - x_a \\
    y_{i,s} - y_a
\end{bmatrix} + \mathbf{S},
\]  

(10)

where \( \mathbf{I} \) is an identity matrix. Fig. 7(c) shows an example of the translation with rotation, when the rigid body is attached with a rectangle tag array. Such rigid body motion is equivalent to first rotating around \( P_a \) with angle \( \alpha \), i.e., from the green array to the blue one, then translating from \( P_s \) to \( P_e \), i.e., from the blue array to the yellow one.

5 Design of Rigid Motion Tracking

We use a set of tags to track the rigid transformation of the tagged object, including the translation and the rotation. To support the tracking of small objects, we choose the small tag AZ9629 as the movement tag. Note that, during the movement, the polarization angle of the tag relative to the antenna changes as well, which can bring in additional phase offsets apart from the change of position. Whereas, for the movement tracking, we take each movement tag as a particle, and only focus on the phase offset caused by the position. Thus, we should reduce the phase offset due to the change of the polarization angle during the movement of the tagged object. Empirically, we fix the tag orientation as shown in Fig. 20, which is insensitive to the rotation due to the movement on the table.

RF-Dial extracts phase variations of the tag array received by the orthogonally deployed RFID antenna pair to track the rigid motion of the tagged object, the design of the rigid motion is shown in Fig. 2. Detailed steps are as follows.

1) Data Preprocessing: With RF-signals from the tag array, we extract the series of phase values for each tag from the two antennas and segment them into the phase value for each snapshot. Then, we estimate the initial state of the tagged objects, including rotation state and rough position.

2) Movement Tracking: We derive the tag movement according to the phase variations, respectively, according to the situations of the linear region and vast region (including both the linear region and non-linear region). Further, we estimate the rigid transformation of the tagged object, including the rotation and translation.

3) Movement Calibration: We detect the outliers of the phase values from the tag array, by comparing the estimated movement of each single tag with the estimated movement of the tag array. Then, we eliminate the outlier(s) and re-estimate the rigid transformation of the tagged object with the remaining tags for calibration.

5.1 Data Preprocessing for Rigid Motion Tracking

5.1.1 Data Segmentation

Due to the issues such as the multi-path effect and ambient noises, the measured phase values may contain some fluctuations in the waveforms. Hence, after receiving the RF-signals from the tag array, we extract the series of phase values for each tag from the two antennas, and calibrate the phase values first. Specifically, due to the operation of \( \text{mod} \) in Eq. (1), the measured phase values are discontinuous. Thus, we stitch the phase values and remove the periodicity among the phase values. Besides, due to the diversity term in Eq. (1), each tag has its own phase offset, so we measure the diversity term among tags in advance and eliminate the tag diversity by offsetting the diversity term. Further, we utilize the Kalman Filter to filter the corresponding noises in the phase values. After that, suppose there are \( n \) tags in the tag array, we segment the phases of \( n \) tags from the two antennas into \( m \) snapshots, denoted as \( \Theta = [\Theta_{1,1}, \Theta_{1,2}, \ldots, \Theta_{1,m}, \Theta_{n,1}, \Theta_{n,2}, \ldots, \Theta_{n,m}] \), where \( \Theta_{i,j} = (\theta_{x,i,j}, \theta_{y,i,j}) \) means the phase values of tag \( T_i \) in the \( j \)th snapshot, \( \theta_{x,i,j} \) and \( \theta_{y,i,j} \) represent the phase values from antenna \( A_x \) and \( A_y \), respectively. The time interval \( \Delta t \) for each snapshot is usually set to a small value, e.g. \( \Delta t = 200 \text{ms} \) in our implementation.

5.1.2 Initial State Estimation

According to Eq. (10), to compute the rigid transformation \( [\mathbf{R}, \mathbf{S}] \), it is essential to determine the initial state of the
rigid body first. E.g., the matrix $\mathbf{T} = [x_{i,s} - x_a \ y_{i,s} - y_a]$ in Eq. (10) depends on the relative positions of tags in the tag array, i.e., the topology and rotation state of the tag array. Hence, since the rotation state of the tag array depends on both initial and subsequent rotation angles of the tag array, we need to estimate the initial rotation angle first.

Without loss of generality, we use the rectangle tag array as an example to illustrate our method to compute the initial rotation angle for the tag array. As shown in Fig. 8, we set the rotation center $P_a$ at the center of the rectangle, and set two orthogonal polar axes $I_x$ and $I_y$ according to the $X$ and $Y$-axis in the global coordinate system. Each tag can be regarded as a particle point in the coordinate system, e.g., $T_1 \sim T_4$. The initial rotation angle of the tag array, i.e., the angle between $I_x$ and $P_aT_i$, is $\beta$.

**Topology Matrix $\mathbf{T}$**. According to the rotation angle $\beta$, we can depict the matrix $\mathbf{T} = [x_{i,s} - x_a \ y_{i,s} - y_a]$ for each tag $T_i$. E.g., for the rectangle tag array in Fig. 8, let $\|T_1T_4\| = h$, $\|T_1T_3\| = \|T_2T_3\| = w$, $\|P_aT_i\| = l$, $\angle T_1P_aT_2 = \eta$, these parameters can be regarded as constants, as long as the topology of the tag array is fixed. Then, according to the relative positions of the tags in the rectangle tag array,

$$
[x_1 - x_a, y_1 - y_a]^T = [l \cos \beta, l \sin \beta]^T, \\
[x_2 - x_a, y_2 - y_a]^T = [l \cos (\beta + \eta), l \sin (\beta + \eta)]^T, \\
[x_3 - x_a, y_3 - y_a]^T = [-l \cos (\beta), -l \sin (\beta)]^T, \\
[x_4 - x_a, y_4 - y_a]^T = [-l \cos (\beta + \eta), -l \sin (\beta + \eta)]^T.
$$

As aforementioned in Section 4.2, the distances between tag pairs are linear/non-linear to the phase differences between tag pairs in the linear region/non-linear region, respectively. In the following, we provide solutions to figure out the rotation angle $\beta$ for the linear region and the vast region (including both the linear region and non-linear region), respectively.

**Linear Region**: The linear region usually has an area of $0.6 \times 0.6m^2$ in the intersection of the central beam regions for two orthogonally deployed antennas. If we set the operation area in this linear region, e.g., a tabletop with the size smaller than $0.6 \times 0.6m^2$, then all the movement can be controlled in this region. Suppose the tag array is located in the linear region, we first consider the relationship between the distance and the phase difference among tags for antenna $A_x$ along the $X$-axis. For any two arbitrary tags $T_i$ and $T_j$ in the tag array, if the distance between $T_i$ and $T_j$ along the $X$-axis, i.e., $\Delta x_{i,j} = x_i - x_j$, is smaller than half-wavelength $\lambda/2$, according to the phase values of $T_i$ and $T_j$ collected from antenna $A_x$, the estimation of $\Delta x_{i,j}$, i.e., $\Delta \hat{x}_{i,j}$, can be computed as follows:

$$
\Delta \hat{x}_{i,j} = \begin{cases}
(\theta_{x,i} - \theta_{x,j}) \times \lambda/(4\pi), & |\theta_{x,i} - \theta_{x,j}| < \pi \\
(\theta_{x,i} - \theta_{x,j} - 2\pi) \times \lambda/(4\pi), & \theta_{x,i} - \theta_{x,j} > \pi \\
(\theta_{x,i} - \theta_{x,j} + 2\pi) \times \lambda/(4\pi), & \theta_{x,i} - \theta_{x,j} < -\pi
\end{cases}
$$

Similarly, we can compute the estimated distance between $T_i$ and $T_j$ along the $Y$-axis, i.e., $\Delta \hat{y}_{i,j}$. Meanwhile, according to Eq. (11), given the rotation angle $\beta$, we can also compute the theoretical value for $\Delta x_{i,j} = x_i - x_j$ and $\Delta y_{i,j} = y_i - y_j$. Therefore, we leverage the Minimum Mean Square Error (MMSE) estimator to estimate the rotation angle $\beta$, we define the mean squared error $e$ by enumerating the squared errors for all pairs of tags, that is,

$$
e(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} (\Delta x_{i,j}(\beta) - \Delta \hat{x}_{i,j})^2 + (\Delta y_{i,j}(\beta) - \Delta \hat{y}_{i,j})^2.
$$

We then compute the optimal value of $\beta$ that achieves the minimal mean squared error for $e(\beta)$:

$$
\beta^* = \arg \min_{\beta} e(\beta).
$$

**Vast Region**: The vast region usually covers a large area of $4 \times 4m^2$ in the intersection of the scanning areas for two orthogonally deployed antennas. As the vast region includes both the linear region and non-linear region, we thus consider the solution for the non-linear region, as it is a more generalized solution for both regions. Suppose the tag array is located at the non-linear region, the distance between the tags are not linear to the phase differences between the tags, which depends on the exact positions of the corresponding tags. Therefore, it is essential to estimate both the position and rotation state of the tag array in the non-linear region. For the tag array, we set the position of the rotation center as $(x, y)$, and set the rotation angle as $\beta$. Then, different from the solution in the linear region, in regard to the non-linear region, for any tag $T_i$ in the tag array, we estimate the Euclidean distance $d_i$ between tag $T_i$ and the corresponding antenna $(A_x$ or $A_y)$, and compute the difference of the Euclidean distance $\Delta \hat{d}_{i,j} = d_i - \hat{d}_j$ for any two arbitrary tags $T_i$ and $T_j$. Specifically, for antenna $A_x$, the difference of the Euclidean distance $\Delta \hat{d}_{x,i,j}$ can be computed as follows:

$$
\Delta \hat{d}_{x,i,j} = \begin{cases}
(\theta_{x,i} - \theta_{x,j}) \times \lambda/(4\pi), & |\theta_{x,i} - \theta_{x,j}| < \pi \\
(\theta_{x,i} - \theta_{x,j} - 2\pi) \times \lambda/(4\pi), & \theta_{x,i} - \theta_{x,j} > \pi \\
(\theta_{x,i} - \theta_{x,j} + 2\pi) \times \lambda/(4\pi), & \theta_{x,i} - \theta_{x,j} < -\pi
\end{cases}
$$

Similarly, we can also compute the difference of the Euclidean distance $\Delta \hat{d}_{y,i,j}$ for antenna $A_y$. Meanwhile, according to the position $(x, y)$ and rotation angle of the tag array $\beta$, we can also compute the theoretical value for the difference of the Euclidean distance $\Delta d_{x,i,j}$ and $\Delta d_{y,i,j}$. Hence, by leveraging the MMSE estimator, we can estimate the $(x, y, \beta)$ by finding the optimal set of parameters that minimizes the squared errors for all pairs of tags:

$$(x, y, \beta)^* = \arg \min_{x, y, \beta} e'(x, y, \beta),$$

$$
e'(x, y, \beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} (\Delta d_{x,i,j} - \Delta \hat{d}_{x,i,j})^2 + (\Delta d_{y,i,j} - \Delta \hat{d}_{y,i,j})^2.$$
5.2 Movement Tracking

5.2.1 Derive the tag movement from the phase variation

To derive the tag movement from the phase variation collected by the antenna pair, we also provide solutions for the linear region and the vast region, respectively.

**Linear Region:** If we set the operation area in the linear region, e.g., a tabletop with the size smaller than $0.6 \times 0.6 \text{m}^2$, we can simply derive the tag movement $[\Delta x_i, \Delta y_i]^T$ from the phase variation for each tag $T_i$ according to Eq. (5).

**Vast Region:** As aforementioned in Section 5.1.2, the position of the tag array can be effectively estimated, then, given the starting position of the tag movement, we can derive the tag movement $[\Delta x_i, \Delta y_i]^T$ from the phase variation for each tag $T_i$, by solving Eq. (6).

5.2.2 Track the continuous movement via the tag array

The ultimate goal is to track the continuous movement of the tagged object via the tag array. Hence, it is essential to track the rigid transformation $[R, S]$ of the tagged objects for each micro-movement. As shown in Fig. 9, we build a global 2D coordinate system according to the deployment of the two mutually orthogonal antennas in Fig. 4. Since we only focus on the rigid transformation rather than the absolute position of the tagged objects, the origin can be set at any position. For simplicity, we can set the initial position of the local rotation center of the tagged object as the origin.

Hence, according to Eq. (10), during the moving process of the tagged object, the rigid transformation of the tagged object can be continuously estimated as follows: For each snapshot $t$, suppose the rotation state in the previous snapshot $t-1$ is $\beta_{t-1}$, then we can estimate the matrix $T_{t-1} = \begin{bmatrix} x_{i,t-1} - x_{a,t-1} \\ y_{i,t-1} - y_{a,t-1} \end{bmatrix}$ from the previous snapshot $t-1$ according to the tag array’s topology and the rotation state.

Then, according to the derived movement of tag $T_i$ at snapshot $t$, i.e. $[\Delta x_{i,t}, \Delta y_{i,t}]^T$, and the matrix $T_{t-1}$ from the previous snapshot $t-1$, we can further figure out the rotation matrix $R_t$ and translation matrix $S_t$ at snapshot $t$ from the following equation:

$$
\begin{bmatrix} \Delta x_{i,t} \\ \Delta y_{i,t} \end{bmatrix} = (R_t - I) \begin{bmatrix} x_{i,t-1} - x_{a,t-1} \\ y_{i,t-1} - y_{a,t-1} \end{bmatrix} + S_t. \tag{12}
$$

After we figure out the rotation angle $\alpha_t$ from the rotation matrix $R_t$, we then update the rotation state in snapshot $t$ as $\beta_t = \beta_{t-1} + \alpha_t$. This process iterates until the tagged object stops moving. For the initial rotation state of the tag array, i.e., the angle $\beta_0$, it can be figured out based on the phase differences among the tags in the tag array, as aforementioned in Section 5.1.2.

Specifically, to compute the parameters from the rotation matrix $R_t$ and translation matrix $S_t$, i.e., $\alpha_t$, $s_{x,t}$, and $s_{y,t}$, we can expand Eq. (12) and obtain the following equations:

$$
\begin{align*}
\{(\cos \alpha_t - 1) \times \delta_{x,t} - \sin \alpha_t \times \delta_{y,t} + s_{x,t} &= \Delta x_{i,t}, \\
\sin \alpha_t \times \delta_{x,t} + (\cos \alpha_t - 1) \times \delta_{y,t} + s_{y,t} &= \Delta y_{i,t},
\end{align*}
$$

(13)

where $\delta_{x,t} = x_{i,t-1} - x_{a,t-1}, \delta_{y,t} = y_{i,t-1} - y_{a,t-1}$. As we can obtain two such equations according to the phase values from a single tag received by two antennas, to solve the three unknown parameters $\alpha_t$, $s_{x,t}$ and $s_{y,t}$ from Eq. (13). At least two tags are essential to provide four equations to solve the unknown parameters. However, due to the issues such as the multi-path effect and ambient noises, it is possible that the estimated tag movement might differ from the actual tag movement in the rigid transformation $[R_t, S_t]$ to a certain extent. Hence, we need to deploy more tags in tag array to track the rigid transformation in a more accurate manner.

Suppose that we are able to collect the RF-signals from $n$ tags, given the parameters $\alpha_t$, $s_{x,t}$ and $s_{y,t}$, according to Eq. (12), we can compute the theoretical tag movement $[\Delta x_{i,t}, \Delta y_{i,t}]$ derived from the phase variations, by leveraging the MMSE estimator, we aim to find the optimal solution $\hat{\alpha}_t$, $\hat{s}_{x,t}$, and $\hat{s}_{y,t}$ to minimize the difference between the theoretical movement $[\Delta x_{i,t}, \Delta y_{i,t}]$ and the derived movement $[\Delta \hat{x}_{i,t}, \Delta \hat{y}_{i,t}]$:

$$
\arg\min_{\alpha_t, s_{x,t}, s_{y,t}} \sum_{i=1}^{n} \left( (\Delta x_{i,t} - \Delta \hat{x}_{i,t})^2 + (\Delta y_{i,t} - \Delta \hat{y}_{i,t})^2 \right).
$$

5.3 Movement Calibration

Due to the issues such as the multi-path effect and mutual interferences in the ambient environment, the phase values of some tags can be distorted to a certain extent, which might further impact the movement tracking of the rigid body via the tag array. Specifically, the multi-path effect mainly comes from the signal reflections from the hand movement while using the tagged object as the HCI device; whereas the mutual interference mainly comes from the interference of inductive coupling among adjacent tags.

Therefore, it is essential to detect the outliers of the phase values, and further eliminate them to calibrate the estimated rigid transformation.

5.3.1 Outlier Detection

Our solution is based on the observation that during the rigid transformation of the tagged object, if the phase values of one or more tags are severely distorted, then, for the tag with outlier phase values, the estimated displacement derived from the phase variations of this tag should be greatly different from the estimated displacement derived from the rigid transformation of the tag array. Specifically, for any specified tag $T_i$, assume that the estimated displacement is denoted with a vector $\hat{s}_i = (\Delta \hat{x}_i, \Delta \hat{y}_i)$, whereas the estimated displacement derived from the rigid transformation is denoted with a vector $s_i = (\Delta x_i, \Delta y_i)$. By comparing the cosine value of the vectorial angle $\epsilon_i = \frac{\hat{s}_i \cdot s_i}{\|\hat{s}_i\| \|s_i\|}$, we can normalize the difference between the two vectors in the range $[0, 1]$. If $\epsilon_i$ is less than a threshold $\tau$, then we can identify tag $T_i$ as a candidate outlier.
5.3.2 Outlier Elimination

As there might exist one or more outliers for the phases of tags, one or more outliers can distort the overall rigid transformation to a certain extent, to effectively eliminate the outliers, our solution is designed as follows: We first compute $\epsilon_i$ for each tag $T_i$ in the tag array. Among the candidate outliers with $\epsilon_i < \tau$, we select the tag with the minimum value of $\epsilon_i$. We then eliminate this tag and re-calculate the rigid transformation $[R, S]$ according to the phase values of the remaining tags. We repeat the above process until all the remaining tags have $\epsilon_i \geq \tau$ or half of the total tags are eliminated. At last, we use the value of $[R, S]$ after the outlier elimination as the final calibrated result for the rigid transformation.

The MMSE function can be solved quickly with advanced algorithms, i.e., Gauss-Newton iteration method. When the multi-path is very serious, it is possible that the residual of the MMSE function is still very large, thereby, the motion tracking fails or has the poor accuracy. Here, the serious multi-path effect often happens when there are moving people around the monitoring area within 60cm. The user’s operating hand is usually considered to have much influence on the RF-signals, but during the operation, its influence on different tags is similar. As we leverage the phase variation of a tag or phase difference between tags, as long as the signal distortion is steady, we can still achieve the accurate motion tracking. While for the serious multi-path effect, it is the main drawback for almost all wireless sensing solutions, including our work. It seems a pity that the wireless sensing is sensitive to the environment, but many researchers have put much effort into extracting the clean signal combating the environment interference [23], which is beyond the scope of our work. In this paper, we aim to provide a light-weight and functional human-computer interaction solution, so we simply have the assumption that there is no moving object within the monitoring area except the tagged object. In this manner, the received signals will not too bad and can be used for the motion tracking. However, it does not mean that our solution cannot tolerate the environmental factors at all. We tested the heavy multi-path situation when there are many reflection things within the monitoring area, and results show that our solution can work well in this situation, details are shown in Section 9.2.

6 Touch Gesture Detection

Different from the rigid motion detection where we view a tag as particle and small tags are preferred, in this section, we use a linear dipole tag as the fine-grained touch interface.

6.1 RSSI Deviation during Touch Gesture

We use a linear dipole tag E51 as the touch interface to explore the signal characteristics due to the touch gesture. We put a tag on a table in front of an antenna. The antenna has the same height as the table. We perform the swipe on the tag from left to right with different settings. We leverage OptiTrack to track the finger position. By synchronizing the finger trace from OptiTrack and the received RF-signals based on timestamps, we can derive the RSSI variation of the tag with different touch positions.

Observation 2: The received signal strength (RSSI) of a tag changes significantly when a user touches the tag. During the swipe from left to right on the tag, the RSSI increases from left to center and then decreases from center to right, the RSSI variation of the swipe process is relatively symmetrical around the tag’s center and forms an $\Omega$-wave pattern. As shown in Fig. 10(a), we perform the swipe with two settings, the RSSI is plotted in Fig. 10(b), where the top figure is the raw RSSI from RFID, the bottom figure is the RSSI values with different touch positions by combining OptiTrack data and RFID data. When the finger presses down on the left of the tag, RSSI decreases about 20dB. RSSI has larger values when the touch position approaches the tag center, the maximum RSSI value is just 1dB smaller than the untouched situation. The whole RSSI variation forms an $\Omega$-wave. This is because that the human skin can be modeled as an equivalent impedance, when the user touches the tag, the capacitive coupling happens that influences the backscattered signals from the tag. As our tag has the dipole antenna, theoretically there exists a mirror position on the other antenna of the tag that has the same coupling effect for any touch position. Note that, the multi-path effect is different at different positions, so the signals of two mirror positions will not be exactly the same, but the RSSI variation pattern seems symmetrical on the whole.

Observation 3: The RSSI deviation keeps steady regardless of the tag’s position and orientation or even the user. Here, RSSI deviation refers to the RSSI difference between the RSSI value when a tag is touched and the RSSI value when the tag is untouched, denoted as $\Delta$RSSI. The absolute RSSI value is related to the position or orientation of the tag relative to the antenna as shown in Fig. 10, but the RSSI variation patterns are similar. Let the stable RSSI value just before touch be the reference value (denoted as $C$), so the RSSI deviation is calculated as: $\Delta$RSSI = RSSI$-C$, the $\Delta$RSSI pattern during the swipe is relatively steady despite the settings. In particular, we explore more about the $\Delta$RSSI pattern. As illustrated in Fig. 11(a), a tag is in front of the antenna’s center about 80cm with its orientation parallel to the antenna plane ($\gamma$=0°), a user faces the antenna and stays behind the tag of 40cm ($d$=40cm) by default. Through adjusting different positions, orientations or users, we collect tag signals during the swipe, results are plotted in Fig. 11(b)-11(d). It is found that the RSSI deviation patterns at different settings have the high similarity. The $\Omega$-waves almost coincide with each other, especially in terms of the distance and users. For different angles, the similarity of $\Omega$-wave patterns is a bit smaller than that of distance or user, but it is still very high. It is reasonable that the capacitive coupling effect accounts for the major signal variation when a tag is touched as the impedance of different users is similar for the tag, so the RSSI deviation keeps steady for different
6.2 Phase Deviation during Touch Gesture

Compared with the RSSI, the phase is a more sensitive parameter of backscattered signals. Similarly, define phase deviation as the phase difference between the phase values when a tag is or not touched, denoted as $\Delta \phi$. When the tag orientation gets changed, the phase deviation does not keep steady. Fig. 14 plots the phase variation patterns during the swipe when putting the tag with different angles, and these patterns are different from each other. It means that if using the pattern of $0^\circ$ as the template, we cannot determine the accurate touch position when the angle is $45^\circ$. This phenomenon is likely to be caused by the difference of the reflection signals from the moving hand. As shown in Fig. 15, the signal received by the tag contains not only the transmitted signals from the antenna, but also the reflection signals from the surrounding objects, among which the hand touching the tag is the main reflector. When the tag is not parallel to the antenna plane, the distance between the hand and the antenna changes during the swipe process, so the reflection signal from the hand to the tag changes a bit as well. The larger angle between the tag orientation and the antenna plane, the larger distance difference of the reflection signal from the moving hand. As the range of the distance variation is just several centimeters, the strength of the reflection signal can be viewed unchanged. However, as mentioned above, the phase is very sensitive to the distance, so the slight distance difference of the moving hand can lead to the significant phase difference of the reflection signal, which further affects the tag signals collected by the antenna. If using the phase, we need to determine the tag orientation first, then generate the pattern templates for each tag orientation, adding much burden for the training and detection. Meanwhile, the amplitude of the phase variance due to the touch is much smaller than the RSSI variance, so the RSSI deviation is more significant and more robust to the environment interference. Therefore, we select the RSSI deviation to detect the touch gesture.

6.3 Model of RSSI Deviation due to Touch Gesture

The power received by the receiver from the transmitter can be described by the Friis equation [32]. Denote the received and transmit power as $P_R$ and $P_T$, the receiver gain and transmitter gain as $G_R$ and $G_T$, the polarization angle...
between the antenna and the tag as $\theta_{\text{pol}}$, so the received power can be represented as:

$$P_R = P_T G_T G_R \cos^2(\theta_{\text{pol}})(\frac{\lambda}{4\pi r})^2,$$

where $\lambda$ is the wavelength, $r$ is the distance between the receiver and the transmitter. However, when the reader’s antenna transmits signals to the tag, the received power of the tag is probable not be power delivered to the tag IC to activate itself, there is a chance of power lost in the transfer.

Fig. 16 shows the equivalent circuit of a tag. The total impedance includes two parts, one is from the tag antenna ($Z_{\text{ant}} = R_{\text{rad}} + jX_{\text{ant}}$), and the other is from the tag IC ($Z_{\text{load}} = R_{\text{load}} + jX_{\text{IC}}$). So the dissipated power by the tag’s load is:

$$P_{\text{load}} = \frac{|I|^2R_{\text{load}}}{2} = \frac{V_{\text{oc}}^2R_{\text{load}}}{2|Z_{\text{ant}} + Z_{\text{load}}|^2},$$

here $V_{\text{oc}}$ is the open-circuit voltage, $I$ is the current, $Z = R + jX$, $R$ refers to the resistance, a real part caused by the circuit resistivity, $X$ refers to the reactance, a complex part caused by inductance or capacitance or both. Based on Eq. (15), when the impedance of the tag antenna and that of the tag IC are conjugate matching, as $R_{\text{ant}} = R_{\text{load}}$, $X_{\text{ant}} + X_{\text{load}} = 0$, the power transfer reaches the maximum:

$$P_{\text{load}}^{\text{max}} = \frac{V_{\text{oc}}^2}{8R_{\text{rad}}}. \quad (16)$$

Normally, it is assumed that received power calculated by the Friis equation in Eq. (14) is totally delivered the tag IC when conjugate matching exists, as $P_R = P_{\text{load}}^{\text{max}}$. If the perfect matching does not exist, the actual power received by the tag IC is less than that predicted by the Friis equation:

$$P_{\text{load}} = P_R \frac{4R_{\text{load}}R_{\text{rad}}}{|Z_{\text{ant}} + Z_{\text{load}}|^2}. \quad (17)$$

Correspondingly, the reported RSSI will change due to different mismatching conditions. When a user swipes on a tag, the position and polarization angle of the tag is fixed, i.e., $P_R$ remains unchanged in Eq. (17), the main changing parameter is about the impedance of the tag due to the capacitive coupling, denoted as $\hat{R} = \frac{4R_{\text{load}}R_{\text{rad}}}{|Z_{\text{ant}} + Z_{\text{load}}|^2}$. When touching the center of the tag, both the IC loop and the two tag antennas are touched at the same time, it is probable that the new conjugate matching is built, so the power transferred to the tag IC is similar to that when the tag is untouched. As the tag impedance accounts for the power transfer, pick up a power reference at time $t_0$, then the received power ratio of a tag only depends on the impedance change due to the touch position at time $t$ relative to time $t_0$, as:

$$\frac{P_{\text{load}}(t)}{P_{\text{load}}(t_0)} = \frac{R(t)}{R(t_0)}. \quad (18)$$

Moreover, in COTS RFID systems, the unit of the RSSI by Impinj R420 is $\text{dBm}$, while the unit of the received power is $\text{Watt}$, they can be transformed as:

$$P_{\text{load}} = \sqrt{\frac{10^{\text{RSSI}/10} \times 1000}{1000}}. \quad (19)$$

Hence, the power ratio in Eq. (18) can be represented by the RSSI deviation, as:

$$\text{RSSI}(t) - \text{RSSI}(t_0) = 20 \log \frac{P_{\text{load}}(t)}{P_{\text{load}}(t_0)} = 20 \log \frac{R(t)}{R(t_0)}. \quad (20)$$

According to Eq. (20), when the tag is fixed during the swipe, the RSSI difference is only related to the impedance change due to the touch position at time $t$ relative to time $t_0$. As the swipe is a continuous process, the user touches the tag first, and then moves the finger to different positions on the tag, so we consider the untouched state before touching the tag as the reference time $t_0$, which is easy to derive during the interaction. That is:

- The RSSI deviation due to touch on the tag is independent of the position or orientation of the tag, which supports our findings in Observation 3.
- Based on the RSSI deviation, we are able to derive the absolute candidate touch positions regardless of the position or polarization angle of the tag.

The above key insight also works with different tag types more than E51, experiment results are shown in Fig. 28.

7 Design of Touch Gesture Detection

We use a single linear dipole tag E51 as the touch interface. As the RSSI deviation pattern is symmetrical for the swipe on the two antennas of a linear dipole tag in Observation 2, we leverage the monotonic trend of the half tag to use one half as the slider and the other half as buttons.

Generally, the touch tag plays a role of “touch bar”, supporting three touch gestures as the click, the press and hold, and the swipe. Before identifying which gesture is performed, it is necessary to generate the RSSI deviation template for estimating the touch position in advance, as illustrated in Fig. 2. Based on the RSSI deviation template, the touch gesture estimation can be performed as follows.

1) Data Preprocessing: First, determine the reference value for computing the RSSI deviation by recalling the RSSI value just before touch. Then, smooth the received signals to filter outliers caused by the ambient noise or the multi-path effect, and calculate the real-time RSSI deviation.

2) Gesture Detection: Based on the RSSI variation in consecutive time intervals, we can determine which touch gesture is actually performed, including the click, the press and hold, and the swipe. Click means touching one position of the tag with short time; Press and hold means touching one position of the tag with long time; Swipe means touching different positions of the tag continuously and smoothly.

3) Touch Estimation: With the extracted RSSI deviation and the estimated touch gesture, we can derive the absolute touch position on the tag referring to the template. For the swipe, we can further estimate the distance, direction and speed with consecutive touch positions.
7.1 Template Generation
The template is generated in advance as the reference for the touch estimation, especially for figuring out the real-time touch position according to the RSSI deviation. To collect the actual position of the finger on the tag, we utilize OptiTrack [33] to track the finger position as the ground-truth. The generation can be divided into three steps: re-sampling, feature extraction, and curve fitting. 1) Re-sampling: By synchronizing the finger trace from OptiTrack and the received RF-signals of touch tag, we are able to capture the ground-truth of the finger position on the tag. Note that, the two kinds of data have different sampling rates, i.e., 120Hz of OptiTrack and around 100Hz of RF-signals, so we re-sample the data to the same sampling rate of 100Hz to correspond RSSI values to each touch point. 2) Feature Extraction: As for a touch gesture, the touch gesture usually starts with the finger hangs over the tag, then the finger presses down to the tag. Considering this continuous process, it is not hard to take the RSSI value just before touching the tag as the reference value for calculating the RSSI deviation, we give an example in Section 7.5. With the reference RSSI value, we can extract the RSSI deviation by subtracting the reference value from the received RSSI values. 3) Curve Fitting: Apply the spline fitting to the derive RSSI deviation pattern, so as to obtain the continuous relationship between the touch point and the RSSI deviation with the discrete RSSI values at finite positions. Considering the uncertain noise during once swipe, a user is required to repeat the swipe across the tag from left to right three times, then we average the three fitting curves to extract the RSSI deviation template.

7.2 Data Preprocessing for Touch Gesture Detection
To eliminate the data fluctuation due to the ambient noise in the environment, before performing the fine-grained touch gesture detection, we first use the moving average filter to smooth the data. Specifically, for each sample \( i \), take the average value of the nearest five samples, from \( i - 2 \) to \( i + 2 \), to replace sample \( i \). Meanwhile, similar to the template generation, we determine the reference RSSI value and then calculate the RSSI deviation for the touch position estimation. As we update the reference RSSI value for each touch operation, we can eliminate the major difference of raw RSSI values due to different positions or orientations. Therefore, the derived RSSI deviation mainly reflects the impedance change caused by the finger touch on the tag’s antenna, so it is robust to different positions and orientations.

7.3 Gesture Detection
Each touch gesture includes three periods: finger hanging over the tag, finger touching on the tag, finger leaving the tag. Fig. 17 plots the RSSI variations of different touch gestures, beginning with pressing down to the tag and ending with lifting the finger from the tag. Here, the swipe is performed on the half of the tag, i.e., from left to center. It is found that the finger vertical movement, including finger up and finger down, will bring the sudden RSSI variation due to the quick action. Specifically, for click, the RSSI decreases suddenly, then lasts for a short period and recovers to original value suddenly. For press and hold, it is similar to “click”, the only difference is its longer lasting time of the touch situation than “click”. The threshold of the lasting period is set to be 0.7 seconds. While for swipe, the RSSI value decreases suddenly, then changes continuously with different touch positions in several time intervals. If the RSSI changes significantly over successive time intervals, it is identified as the swipe operation. Empirically, the time interval is set to be 0.1 seconds, the number of successive time intervals is 3. Note that, the tag center is a fallible area where the absolute sudden RSSI variation is a bit small, it is easy to mistake the click or press and hold for the noise. According to the yellow swipe data in Fig. 17(b), when the finger leaves the tag from the tag center, the RSSI changes only about 2dB. It is because that when the touch position approaches the tag center, the conjugate mismatching condition gets better, so the power loss gets less. Especially when the finger just locates at the tag center, touching both the two tag antennas, new conjugate matching will form, and the power loss could be less than 1dB. As the RSSI of a tag in the static environment will change about 1dB due to the ambient noise, a too small RSSI variation threshold will mistake the noise for the touch gesture, so if the range of RSSI values in a short time interval exceeds 2dB, we think the RSSI changes significantly and suddenly, it probably indicates the finger vertical movement or the swipe, instead of the noise. The size of the fallible area is related to the RSSI variation threshold, and it is about 1cm with our settings. Considering the width of the finger, the effective operation size on the used touch tag is around 8.5cm. Thankfully, the size of this fallible area is not large, we suggest to avoid to design a button in this fallible area. Also, as depicted above, the RSSI deviation pattern is symmetric during the swipe across the whole tag, so in our design, we utilize the half part of the tag as the button interface and the other half as the slider interface, an example is illustrated in Fig. 29. This may seem to be the limitation to our work, but in fact, such design is natural and friendly, e.g., the touch bar on Apple MacBook Pro contains both buttons and the slider. How to extend the operation area of the tag and further improve the interaction experience will be studied in our future work.

7.4 Touch Estimation
With the known touch interface design, as which part functions as buttons and which part functions as the slider, we can estimate the absolute touch position on the tag by comparing the extracted RSSI deviation with the template. For swipe, by combining the consecutive touch positions, we are able to determine the swipe direction, and calculate the swipe distance as well as the swipe speed, which are the key factors for analyzing the friendly interaction commands.
Note that, it is possible that there are several candidate positions corresponding to the same RSSI value, considering the swipe is a continuous process, the position change should be coherent, so we select the one that lets the position change smooth and reasonable as the estimated result.

7.5 Touch or Motion: Segmentation and Identification

Last but not least, how to segment the received RF-signals is a key issue during the practical operation. The above system design of either the motion tracking or the touch gesture detection is based on the extracted segment of a single command. But in fact, the consecutive signals contain several commands as well as invalid variations due to the multi-path effect, like a moving hand or a moving person around the tagged object. Hence, it is significant to segment signals and extract clean signal segments for each valid command. Taking the tagged box in Fig. 29, we first moved hands around the object, then let it move on the table, after that we performed a swipe on the tag, the collected signals of one movement tag and one touch tag are illustrated in the top two figures of Fig. 18. It is observed that the motion mainly brings the phase variation and the touch causes obvious RSSI variation. While for the invalid hand movement, both phase and RSSI values have slight fluctuations. The observation is consistent with Table 1. Thus, we use these characteristics to segment signals and identify commands, including invalid command, motion command, and touch command. The workflow is as follows:

1) Calculate the differential coefficient (diff) of phase, and extract the candidate segment. As the phase is more sensitive than the RSSI, we use the phase diff for segmentation. The signals with continuous non-zero phase diff values are the candidate valid command segment.

2) Calculate the feature vector for the segment. We extract two features: range and standard deviation (std). As shown in Fig. 18, the motion and touch commands usually have the larger range and std values than the invalid hand movement. Thus, we calculate the range and std values of either phase or RSSI for each tag (movement tags and the touch tag) within the segment, and form a feature vector.

3) Classify the segment and identify the command, i.e., invalid, motion or touch. With the feature vector, we classify the segment using the decision tree algorithm. The classifier only needs to be trained once in advance, and can be used for any position or orientation of the tagged object within the monitoring area. The classifier is light-weight, and the delay of classification can be ignored. If the segment is estimated as caused by the motion or touch, the system inputs the segment to the motion or touch module to identify the detailed command, as shown in Fig. 2.

Moreover, for the touch gesture detection, it is important to determine the reference RSSI value. The reference value is used to degrade effects of changing environmental factors. Theoretically, the range of the RSSI variation during the swipe is steady, it could be better to select the extreme values as the reference, but we cannot obtain the extreme value before the user touches the tag. Considering the user’s habit, there usually exists a short pause when the user prepares to perform the touch gesture, so it is a basic idea to choose the stable RSSI value in the short pause. Note that, when the finger hangs over the tag with different heights, the RSSI value has a small changing range of 1-2dB, so we need to update the template when detecting the non-outlier value beyond the range of the template for the higher accuracy.

8 Discussion

The 2D motion tracking method can extend to the 3D tracking: The model of the rigid motion decomposition in Section 4.4 can be rewritten for the 3D motion with the 3D translation and rotation matrices. To realize the 3D motion tracking, one more antenna orthogonal to the antenna pair in Fig. 20 is required. The additional antenna can be deployed on the ceiling to capture the signal change mainly along the vertical direction. Also, to ensure there are always enough tags read by the reader antenna, more tags with different orientations should be attached on the tag, just like Tagball [16].

The tag type has influence on the touch gesture recognition: Not all kinds of tags can be used as the touch tag. For example, the antenna of tag AZ9654 is in a broadband structure, its RSSI variation due to the swipe is totally different from the Ω-wave observed with tag E51, as shown in Fig. 19. Our method is more suitable for the tag with meandered lines in the tag antenna, i.e., E51, AZ9640 and AZ9662. Due to the different length of meandered lines and chip types, the performance of different tag types differs, details are shown in Section 9.3. Meanwhile, as the swipe along different lines on the tag could incur different RSSI variation patterns, to ensure the swipe can be performed along the same line, we prefer the linear tag instead of the square tag.

The touch gesture detection module works as long as the touch tag can be activated wherever the touch position is: For motion tracking, we split the scanning area into two regions based on the relationship between tag movement and phase variation. However, the touch gesture detection relies on the RSSI deviation, it views the whole area as the same. As shown in Observation 3, as long as the touch tag can be activated wherever the touch position is, the touch method works. The operating area mainly depends on the smaller readable range of movement tags and the touch tag. Empirically, the maximum distances of movement tag AZ9629 and touch tag E51 in our settings are 1.8m and 2m, respectively.
9 PERFORMANCE EVALUATION

9.1 Experimental Setup

We built RF-Dial using an ImpinJ R420 reader, two Laird S9028 RFID antennas and a board eraser with multiple tags, i.e., AZ9629 tags as movement tags, and E51 / AZ9640 / AZ9662 as the touch tag. For the rigid motion tracking, as shown in Fig. 20, we deployed two antennas around a table in a mutually orthogonal manner. The two antennas are separated about 0.8m away from the table, and the antennas and the table are at the same height of 0.8m. The size of the board eraser is $13 \times 5.5$cm$^2$, and we designed three layouts of the tag array with different numbers of tags, i.e., 2, 4, 6. We operated the tagged board eraser on the table within the range of $100 \times 50$cm$^2$, including the linear region and non-linear region. For the touch gesture detection, only one antenna is required in theory. According to Observation 3, the RSSI deviation is insensitive to the angle below 45°. Considering that the board eraser can be with any rotation on the table and there are one orthogonally deployed antenna pair, we select the antenna whose polarization angle relative to the tag is smaller to collect data for analysis. We tested three kinds of linear tags as touch tag, including E51 / AZ9640 / AZ9662. OptiTrack was used to provide the ground truth.

9.2 Evaluation of Rigid Motion Tracking

We first evaluated the performance in tracking the motion of translation and rotation. We respectively used two metrics to evaluate the performance in tracking accuracy, i.e., the translation error and rotation error. The translation error refers to the difference between the ground-truth translation and the estimated translation, which is measured in the unit of cm. The rotation error refers to the difference between the ground-truth rotation and the estimated rotation, which is measured in the unit of degree. We evaluated the performance by varying the following factors: 1) the number of tags in the tag array, 2) the moving distance of the tagged object, 3) the light multi-path and heavy multi-path situations, and 4) the linear region and non-linear region.

9.2.1 Evaluate the translation error

To evaluate the translation error of the rigid transformation, we selected a start point and moved the tagged eraser to an end point with only translation. We randomly selected 100 samples with different moving distances, i.e., 0~3cm, 3~6cm, and 6~9cm. Fig. 21(a) shows the translation error with different multi-path effect in the linear region. Here the light multi-path situation refers to the situation where the user does not hold the eraser, whereas the heavy multi-path situation refers to the situation where the user holds the eraser. The multi-path propagation mainly comes from the RF-signal reflection from the human hand. It is observed that as the number of tags increases, the translation error monotonically decreases for the light multi-path situation, because phase samplings from more tags help reduce the estimation error; For the heavy multi-path situation, the translation error does not monotonically decrease, sometimes it even slightly increases, because the phase samplings from some tags become outliers due to the multi-path from the hand, further increasing the estimation error. Besides, as the moving distance gradually increases from 0~3cm to 6~9cm, the translation error monotonically increases, because the translation error should be mainly linear to the moving distance. Nevertheless, the average translation errors are all less than 1cm for the light multi-path situation, and less than 2cm for the heavy multi-path situation. Fig. 21(b) shows the translation error for the linear region in the heavy multi-path situation. It is observed that the translation error for the non-linear region is a bit larger than the one for the linear region, nevertheless, very close performance is achieved for the two regions. It implies that our solution is able to efficiently tackle the translation estimation in both the linear region and non-linear region. The above solutions all used the movement calibration method, we further evaluated the performance for the movement calibration method. Fig. 21(c) shows the translation error with/without calibration in the heavy multi-path situation. The number of tags is 4, and the translation is randomly selected from 0~9cm. It is observed that the calibrated solution can further reduce 44% and 47% of the translation error, respectively, for the linear region and non-linear region.

9.2.2 Evaluate the rotation error

To evaluate the rotation error of the rigid transformation, we set the central point of the eraser as the rotation center, and rotated the eraser around the center with a random angle from 0° to 30°. We randomly selected 100 samples with different rotation angles, i.e., 0°~10°, 10°~20°, and 20°~30°. Fig. 21(d) shows the rotation error with different multi-path effect in the linear region. The results are very similar to the translation error for different multi-path situations. The average rotation errors are all less than 2.2° for the light multi-path situation, and less than 6° for the heavy multi-path situation. Fig. 21(e) shows the rotation error for the non-linear region in the heavy multi-path situation. The rotation error for the non-linear region is a bit larger than the linear region, nevertheless, very close performance is achieved for the two regions. It implies that our solution can efficiently tackle the rotation in both regions. The above solutions all used the calibration method, we further evaluated the performance for the calibration method. Fig. 21(f) shows the rotation error with/without calibration in the heavy multi-path situation. The number of tags is 4, and the rotation is randomly selected from 0~30°. It is observed that the calibrated solution can further reduce 25% and 26% of the rotation error for the linear region and non-linear region.

9.2.3 Macro benchmark of rigid motion tracking

We further compare RF-Dial with the state-of-art absolute localization-based solution (AbsLoc), which is also adopted in Tagball [16]. For AbsLoc, we estimated the absolute
position of each tag in the tag array, and then used these estimated absolute positions to compute the rigid motion. We continuously moved the tagged eraser on the table, and used RF-Dial and AbsLoc to track the motion, respectively.

RF-Dial achieves much better performance than AbsLoc in term of the translation error. Fig. 22(a) plots the Cumulative Distribution Function (CDF) of the translation error. For RF-Dial, we evaluated the translation error in the linear region and non-linear region, respectively. For the linear region, the average translation error is 0.6cm, and the translation error is controlled in the range of 0~1.2cm for 90% of the test cases. For the non-linear region, the average translation error is 0.66cm, the translation error is controlled in the range of 0~1.4cm for 90% of the test cases. As AbsLoc uses the absolute positioning method, which is not sensitive to whether it is the linear region or nonlinear region, we thus plot the translation error in the vast region. The average translation error is 2.8cm, and the translation error is controlled in the range of 0~3.7cm for 90% of the test cases.

RF-Dial achieves much better performance than AbsLoc in term of the rotation error. Fig. 22(b) plots the CDF of the rotation error. For RF-Dial, we evaluated the rotation error in the linear region and non-linear region, respectively. For the linear region, the average rotation error is 1.9°, the rotation error is controlled in the range of 0~3.9° for 90% of the test cases. For the non-linear region, the average rotation error is 2.8°, the rotation error is controlled in the range of 0~5.3° for 90% of the test cases. For AbsLoc, we evaluated the rotation error in the vast region, the average rotation error is 12.9°, the rotation error is controlled in the range of 0~24.6° for 90% of the test cases.

9.3 Evaluation of Touch Gesture Detection
The core of touch gesture is the touch position of the touching finger on the tag, thus, we use the metric of error of touch position to reflect the touch accuracy. Error of touch position refers to the absolute difference between the estimated touch position from RF-signals and the captured touch position from OptiTrack. By default, we put the tagged object in front of the antenna at the distance of 80cm, and utilized three times swipe data to generate the RSSI deviation template. Users were required to repeat the swipe ten times. With each swipe data, we selected five touch positions during the swipe randomly, and recorded the error of touch position. As we deployed one orthogonal antenna pair, the polarization angle between the tag and one antenna of the pair is within 45°, which antenna we would like to utilize for the touch gesture detection. We evaluated the performance by varying the following factors: 1) the depth from the tag to the antenna, 2) the relative polarization angle between the tag and the antenna, 3) the wearing jewelry of the user, 4) the user, 5) the swipe speed, and 6) the tag type.

Depth: RF-Dial can track the touch position accurately with the same template when the depth from the tag to the antenna changes within several tens of centimeters. We put the tagged object in front of the center of antenna, and adjusted the distance from 60cm to 110cm. Fig. 23 shows the error of touch position with different depth. It is observed that we can use the template generated with data at depth of 80cm to accurately track the real-time touch position at depth within [60, 110]cm, whose average error is no more than 1.7mm. This is because that the extracted RSSI deviation mainly relies on the coupling effect due to the added human impedance when touching on the tag, which should be position-independent. Thus, the same template can be used when the depth changes several tens of centimeters.

Polarization Angle: RF-Dial can tolerant the polarization angle variation within 45° to track the touch position with error less than 5mm. We collected the training data when the tag plane parallel to the antenna plane, denoted the angle as 0°, and rotated the tag within 45° clockwise and counterclockwise. Fig. 24 shows the error of touch position with different angles. We can find that with the increasing angle,
the error increases as well and achieves 4.92mm when the angle is $-45^\circ$, which is still acceptable compared with the half tag length of around 4cm. This is because that although the extracted RSSI deviation should be independent of the polarization angle, the tag’s received signal caused by the multi-path changes during the rotation, such that, the extracted RSSI deviation contains influence of both the coupling effect and the multi-path effect. Despite our solution is not as robust to angle as to distance, the average accuracy is satisfying of less than 4mm within rotation angle of $45^\circ$.

**Wearing Jewelry:** The touch detection is insensitive to the wearing jewelry of the user. We let the users wear different jewelry on the operation hand to evaluate the effect of wearing jewelry, including the crystal bracelet, metal ring, smart wristband, and mechanical watch. The template is generated with the data without jewelry. The results are shown in Fig. 25, the average errors for different jewelry are within [1.09, 1.56]mm, so wearing jewelry will not bring huge influence on accuracy. It is probably because that the jewelry will not change the human impedance significantly.

**User:** Different users can share the same template with the average error below 2.7mm. We selected six users, i.e., two females (user1 and user3) and four males, to evaluate whether a general template can be shared with different users. User1 is required to perform the swipe more three times than other users, based on which we generated the template. Fig. 26 plots the error of touch position with different users. We can find that the template is robust to different users and unrelated to the gender. The accuracy of user1 herself is the highest, with the average distance error of 1.1mm. Even in the worst case (user6), we can still achieve the high accuracy with the error less than 2.7mm. It is because that different people have different impedance, but such difference is small enough, allowing us to use a general template for monitoring the touch gesture of different users.

**Speed:** The swipe speed has little influence on the accuracy of touch detection. We swiped on the tag with three different speeds: slow (less than 15mm/s), normal (15-20mm/s), fast (larger than 20mm/s). The generated template is based on the swipe data of normal speed. Fig. 27 plots the CDF of the touch position error with different speeds. 90 percent of errors with slow, normal and fast speed are 2.83, 2.09, and 4.52mm, respectively. It is because that the finger gesture has small difference for different speeds, i.e., the touch angle of the finger on the tag, the touch area. Meanwhile, our results are calculated with the assumption that the captured position from OptiTrack is absolutely accurate, but in fact, when we swipe fast, the finger tracking accuracy will decrease, which degrades our accuracy to some extent.

**Tag Type:** Different kinds of linear dipole tags can be used as the touch tag. We tried three kinds of tags: E51, AZ9640, and AZ9662, which are all linear dipole tag as shown in Fig. 20. They vary in terms of the length of meandered lines and the chip type. For the meandered line length, E51 $\approx$ AZ9640 $> A Z9662$. The chip of either AZ9640 or AZ9662 is H3/H4, while the chip of E51 is M5. The CDF of touch position error in Fig. 28 indicates that these three tag types have the potential to be a reliable touch interface. Although E51 almost has the same size as AZ9640, their chips are different, so AZ9640 performs better. Tag AZ9662 does not perform as well as others, which is likely because of its shorter meandered line, but it still has a high accuracy of error less than 5.7mm for 90 percentile estimations. Overall, a linear dipole tag with the suitable chip and long meandered lines in the antenna could be a good choice for the touch tag.

### 9.4 Case Study: A Functional Drawing Application

In this section, we implemented a prototype system of RF-Dial to provide an functional drawing sample application. Fig. 29 shows the experiment setup and user interface of RF-Dial. This way, we turned the ordinary candy box into an intelligent interaction device. The size of the used candy box is about $13 \times 4.5 \times 10cm^3$, so it is comfortable for the user to operate with the tagged box. We attach four AZ9640 tags on two side faces together with an E51 tag on the top face. The tag separation distance on the same side face is about 10cm, and the height difference between the tags on the top face and the side face is about 8cm. Technically, RF-Dial tracks...
the translation of the object to draw lines, and adjusts the line color automatically using the rotation of the object just during the drawing process. Meanwhile, there is a simple touch interface on the top face of the object, including two buttons at the left part and a slider at the right part. The touch interface is used to receive the user’s commands of adjusting parameters like the line color and the line width. The right figure of Fig. 29 shows the user interface of our drawing application, here we tried to draw a colorful circle clockwise with radius of 15cm based on the rigid motion tracking. That is, we used the translation of the object to restore the drawing trace, and adjusted the line color based on the rotation at the same time. At the beginning, we moved the object without rotation, so the line color kept black. Later, we moved the object with rotation, so the line color got changing with the trace. Totally, the drawing trace matched the baseline circle well. We also tested the touch gesture by clicking buttons to choose the item and performing the swipe to adjust the parameter value. Experiments show that we can filter the invalid gesture and extract data segment of motion or touch accurately. Specifically, in terms of the rigid motion, RF-Dial can achieve the average accuracy of 0.92cm in the translation tracking and of 2.88° in the rotation tracking. The latency of motion tracking mainly comes from the calculation of the MMSE method. We use the Gauss-Newton iteration method to optimize the calculation process, so the latency is about 0.85ms, which is acceptable. While in terms of the touch gesture, RF-Dial can identify the three gestures with high accuracy of 98.5% and track the touch position with 90 percent of errors less than 2.65mm. The latency mainly comes from the window size to identify which command is performed, and it is about 0.23s when the window size is 0.2s. Therefore, we can realize the real-time touch detection. This sample application shows that RF-Dial has the potential to be a promising HCI device.

10 Conclusion

In this paper, we propose RF-Dial, a light-weight, battery-free and functional HCI solution via COTS RFID devices. By attaching a tag array on the side face of an ordinary object together with a dipole tag on the top face, an ordinary object can be easily turned into an intelligent interaction device. In particular, RF-Dial can not only track the rigid motion of the tagged object, including the translation and the rotation simultaneously, but also detect the touch gesture of a user on the surface of the object, such as the click, the press and hold, and the swipe. For the rigid motion tracking, we build a phase-based model that tracks the translation and the rotation of the tagged object simultaneously. For the touch gesture detection, we build an RSSI-based model to accurately estimate the real-time touch position of the touching finger on the tag, in spite of the position or orientation of the tag. We implemented a prototype system of RF-Dial, and evaluated its performance in real settings. Extensive experiments show that RF-Dial achieves an accurate rigid motion tracking, with a small error of 0.6cm for the translation tracking, and a small error of 1.9 degrees for the rotation estimation. Besides, RF-Dial can also detect the touch gesture accurately, as the 90 percent of touch position errors are less than 2.09mm.

Acknowledgments

This work is supported in part by National Natural Science Foundation of China under Grant Nos. 61827174, 61832008, 61902175, 61802169, 61702257; JiangSu Natural Science Foundation under Grant Nos. BK20190293, BK20180325, BK20170648. This work is partially supported by Collaborative Innovation Center of Novel Software Technology and Industrialization. The research of Lei Yang is supported by NSFC General Program (NO. 61972331) and NSFC Key Program (NO. 61932017). Lei Xie is the corresponding author.

References


[33] https://optitrack.com/, OptiTrack.

Yanling Bu received her M.S. degree in computer technology from Nanjing University, China in 2018. She is currently a Ph.D. student in the Department of Computer Science and Technology at Nanjing University. Her research interests include RFID sensing and localization.

Lei Xie received his B.S. and Ph.D. degrees from Nanjing University, China in 2004 and 2010, respectively, all in computer science. He is currently an associate professor in the Department of Computer Science and Technology at Nanjing University. He has published over 80 papers in IEEE Transactions on Mobile Computing, ACM/IEEE Transactions on Networking, IEEE Transactions on Parallel and Distributed Systems, ACM Transactions on Sensor Networks, ACM MobiCom, ACM UbiComp, ACM MobiHoc, ACM INFOCOM, IEEE ICNP, IEEE ICDCS, etc.

Yinyin Gong received her M.S. degree in computer technology from Nanjing University, China in 2019. Her research interests include RFID systems and Internet of Things.

Chuyu Wang received his Ph.D. degree in computer science from Nanjing University, China in 2016. He is currently an assistant professor in the Department of Computer Science and Technology at Nanjing University. His research interests include RFID systems, Software-Defined Radio, activity sensing, indoor localization, etc.

Lei Yang received the B.S. and Ph.D. degrees from the School of Software and the Department of Computer Science and Engineering at Xi’an Jiaotong University. Previously, he was a postdoc fellow at the School of Software of Tsinghua University. He is currently working as a Research Assistant Professor with the Department of Computing, The Hong Kong Polytechnic University.

Jia Liu received his Ph.D. degree from Nanjing University in 2016 and B.S. degree from Xidian University in 2010. He is currently a research assistant professor in the Department of Computer Science and Technology at Nanjing University. His research interests include RFID identification and passive sensing.

Sanglu Lu received her B.S., M.S. and Ph.D. degrees from Nanjing University, China in 1992, 1995 and 1997, respectively, all in computer science. She is currently a professor in the Department of Computer Science and Technology at Nanjing University. Her research interests include distributed computing and pervasive computing.