

Tell Me What I See: Recognize RFID Tagged Objects in Augmented Reality Systems

Lei Xie[†], Jianqiang Sun[†], Qingliang Cai[†], Chuyu Wang[†], Jie Wu[‡], and Sanglu Lu[†]

[†]State Key Laboratory of Novel Software Technology, Nanjing University, China

[‡]Department of Computer Information and Sciences, Temple University, USA

{lxie,sanglu}@nju.edu.cn, {SunJQ,caiqingliang,wangcyu217}@dislab.nju.edu.cn, jiewu@temple.edu

ABSTRACT

Nowadays, people usually depend on augmented reality (AR) systems to obtain an augmented view in a real-world environment. With the help of advanced AR technology (e.g. object recognition), users can effectively distinguish multiple objects of different types. However, these techniques can only offer limited degrees of distinctions among different objects and cannot provide more inherent information about these objects. In this paper, we leverage RFID technology to further label different objects with RFID tags. We deploy additional RFID antennas to the COTS depth camera and propose a continuous scanning-based scheme to scan the objects, i.e., the system continuously rotates and samples the depth of field and RF-signals from these tagged objects. In this way, by pairing the tags with the objects according to the correlations between the depth of field and RF-signals, we can accurately identify and distinguish multiple tagged objects to realize the vision of “tell me what I see” from the augmented reality system. For example, in front of multiple unknown people wearing RFID tagged badges in public events, our system can identify these people and further show their inherent information from the RFID tags, such as their names, jobs, titles, etc. We have implemented a prototype system to evaluate the actual performance. The experiment results show that our solution achieves an average match ratio of 91% in distinguishing up to dozens of tagged objects with a high deployment density.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords

RFID; Augmented Reality System; Prototype Design

INTRODUCTION

As the proliferation of augmented reality technology, people nowadays start to leverage augmented reality (AR) systems (e.g. Microsoft Kinect, Google Glass) to obtain an augmented view in a real-world environment. For example, devices like the Microsoft Kinect [13], i.e., a depth camera, can effectively



Figure 1. Tell me what I see from the augmented reality system

conduct object recognition based on pattern recognition technology. Therefore, users can effectively distinguish multiple objects of different categories, e.g., a specified object in the camera can be recognized as a vase, a laptop, or a pillow based on its natural features. However, these techniques can only offer a limited degree of distinctions among different objects, since multiple objects of the same type may have very similar features, e.g., the system cannot effectively distinguish between two laptops of the same brand, even if they are owned by different people. Moreover, they cannot provide more inherent information about these objects, e.g., the specific configurations, the manufacturers, and production date of the laptop. Therefore, it is rather difficult to provide these functions by purely leveraging the AR technology.

Fortunately, the rise of RFID technology has brought new opportunities to meet the new demands [27, 31, 15]. The RFID tags can be used to label different objects, and store inherent information of these objects in their onboard memory. Moreover, in comparison to the optical markers such as QR code, the COTS RFID tag has an onboard memory with up to 4K or 8K bytes, and it can be effectively identified even if it is hidden in/under the object. This provides us with an opportunity to effectively distinguish these objects, even if they belong to the same brand and have the same features of appearance. Figure 1 shows a typical application scenario of the above vision. In this scenario, multiple people are standing or sitting together in the cafe, while they are wearing the RFID tagged badges. From the camera’s view, the depth camera can recognize multiple objects, or rather human subjects, as well as the depth from its embedded depth sensor, which is associated with the distance to the camera. The RFID reader can identify multiple tags within the scanning range, moreover, it is able to extract the signal features like the received signal strength (RSSI)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

UbiComp '16, September 12–16, 2016, Heidelberg, Germany
©2016 ACM. ISBN 978-1-4503-4461-6/16/09...\$15.00

DOI: <http://dx.doi.org/10.1145/2971648.2971661>

and phase from the RFID tags. By effectively pairing these information together, the system can realize the vision of “tell me what I see from the augmented reality system”. For example, as shown in Figure 1, the inherent information extracted from the RFID tags, such as their names, jobs and titles can be directly associated with the corresponding human subjects in the camera’s view. This provides us more opportunities to communicate with unknown people by leveraging this novel RFID assisted augmented reality.

Although many schemes for RFID-based localization [32, 28, 34] have been proposed, they mainly focus on the absolute object localization, and usually require anchor nodes like reference tags for accurate localization. They are not suitable for distinguishing multiple tagged objects because of two reasons. First, we only require to distinguish the relative location instead of absolute location of multiple tagged objects, by pairing the tags to the objects based on the correlation between the depth of field and RF-signals. Second, the depth camera cannot effectively use the anchor nodes, and it is impractical to deploy multiple anchor nodes in conventional AR applications.

In this paper, we leverage the RFID technology [33, 16] to further label different objects with RFID tags. We deploy additional RFID antennas to the COTS depth camera and propose a continuous scanning-based scheme to scan the objects, i.e., the system continuously rotates and samples the depth of field and RF-signals from these tagged objects. In this way, we can accurately identify and distinguish multiple tagged objects, by sufficiently exploring the inherent correlations between the depth of field and the received RF-signal. Specifically, we respectively extract the RSSI and phase value from RF-signal, and pair the tags with the objects according to the correlation between the depth value and RSSI/phase value.

However, there are several challenges in distinguishing multiple tagged objects in AR systems. The first challenge is to conduct accurate pairing between the objects and the tags. In real applications, the tagged objects are usually placed in very close proximity, and the number of objects is usually in the order of dozens. In this situation, it is difficult to realize accurate pairing due to the large cardinality and mutual interference. The second challenge is to mitigate the interference from the issues like the multi-path effect and object occlusion in realistic settings. These issues can lead to nonnegligible interference to pair the tags with the objects, such as the missing tags or objects which fail to be identified. The third challenge is in devising an efficient solution without any additional assistance, like the anchor nodes. It is impractical to intentionally deploy anchor nodes in real applications due to intensive deployment costs on manpower and time.

This paper represents the first study of using RFID technology to precisely distinguish multiple objects in augmented reality systems. Specifically, we make three key contributions in this paper. 1) To the best of our knowledge, we are the first to consider identifying and distinguishing multiple tagged objects with RFID systems, it provides a key supporting technology for the augmented reality systems to realize the vision “tell me what I see from the AR system”. 2) We conduct an extensive experimental study to explore the inherent correlations

between the depth of field and RF-signals from the tagged objects. We thus propose continuous scanning-based solutions and respectively leverage the RSSI and phase value from RF-signals to accurately distinguish the multiple tagged objects. 3) We implemented a prototype system and evaluated the actual performance with case studies. Our solution achieves an average match ratio of 91% in distinguishing up to dozens of RFID tagged objects with a high deployment density.

RELATED WORK

Depth camera-based pattern recognition: Depth camera-based pattern recognition aims at using the depth and RGB captured from the camera to recognize objects in a more accurate approach. Based on the depth processing [11, 18], a number of technologies are proposed in object recognition [23] and gesture recognition [5, 21, 8, 30, 22]. Nirjon et al. solve the problem of localizing and tracking household objects using depth-camera sensors [20]. Kinect-based pose estimation method [21] is proposed in the context of physical exercise, examining the accuracy of joint localization and robustness of pose estimation with respect to the orientation and occlusions.

RFID in Ubiquitous Applications: RFID has been investigated in various ubiquitous applications, including indoor localization [34, 24], activity sensing [2], tabletop interaction [9], physical object search [19], etc. Prior work on RFID-based localization primarily relied on Received Signal Strength [34, 24] or Angle of Arrival [1] to acquire the absolute location of an object. The state-of-the-art systems use the phase value to estimate the absolute or relative location of an object with higher accuracy [33, 27, 17, 25]. RF-IDraw uses a 2-dimensional array of RFID antennas to track the movement trajectory of one finger attached with an RFID tag so that it can reconstruct the trajectory shape of the specified finger [29]. Tagoram exploits tag mobility to build a virtual antenna array, and uses differential augmented hologram to facilitate the instant tracking of a mobile RFID tag [32]. Find My Stuff (FiMS) provides search support for physical objects inside furniture, on room level, and in multiple locations [19].

Combined use in augmented reality environment: Recent works further consider using both depth camera and RFID for indoor localization and object recognition in augmented reality environment [26, 14, 6, 3]. Wang et al. propose an indoor real-time location system combined with active RFID and Kinect by leveraging the positioning feature of identified RFID and the object extraction ability of Kinect. Klompmaker et al. use RFID and depth-sensing cameras to enable personalized authenticated tangible interactions on a tabletop [14]. Galatas et al. propose a multimodal context-aware localization system, by using RFID and 3-D audio-visual information from 2 Kinect sensors deployed at various locations [6]. Cerrada et al. present a method to improve the object recognition by combining the vision-based techniques applied to the range-sensor captured 3D data, and object identification obtained from RFID tags [3].

SYSTEM OVERVIEW

Design Goals

We aim to implement a supporting technology for the AR systems to realize the vision of “tell me what I see from the

augmented system”, by leveraging RFID tags to label different objects. In order to achieve this goal, we need to collect the responses from multiple tags and objects, and then pair the RFID tags to the corresponding objects, according to the correlations between the depth of field and RF-signals. Therefore, we need to consider the following metrics in regard to system performance: 1) *Accuracy*: Since the objects are usually placed in very close proximity, there is a high accuracy requirement in distinguishing these objects, i.e., the average match ratios should be greater than a certain value, e.g., 85%. 2) *Time-efficiency*: Since the AR applications are usually executed in a real-time approach, it is essential to reduce the time delay in identifying and distinguishing the multiple objects. 3) *Robustness*: The environmental factors, like the multi-path effect and partial occlusion, may cause the responses from the tagged objects to be missing or distorted. Besides, the tagged objects could be partially hidden behind each other due to the randomness in the deployment. The solution should be robust to these noises and distractions.

System Framework

We design a prototype system as shown in Figure 2(a). We deploy one or two additional RFID antennas to the COTS depth camera. The RFID antenna(s) and the depth camera are fixed to a rotating shaft so that they can rotate simultaneously. For the RFID system, we use the COTS ImpinJ R420 reader [10], one or two Laird S9028 antennas, and multiple Alien 9640 general purpose tags; for the depth camera, we use the Microsoft Kinect for windows. They are both connected to a laptop placed on the mobile robot. The mobile robot can perform a 360 degree rotation along with the rotation axis. In the following sections, without loss of generality, we evaluate the performance using the above configurations. By attaching the RFID tags to the specified objects, we propose a continuous scanning-based scheme to scan the objects, i.e., the system continuously rotates and samples the depth of field and RF-signals from these tagged objects. In this way, we can obtain the depth of specified objects from the depth sensor inside the depth camera, we can also extract the signal features such as the RSSI and phase values from the RF-signals of the RFID tags. By accurately pairing these information, the tags and the objects can be effectively bound together.

Figure 2(b) further shows the software framework. The system is mainly composed of three layers: the sensor data collection layer, the middleware layer, and the application layer. For the sensor data collection layer, the depth camera recognizes multiple objects and collects the corresponding depth distribution, while the RFID system collects multiple tag IDs and extracts the corresponding RSSIs or phases from the RF-signals of RFID tags. For the middleware layer, we aim to sample and extract some features from the raw sensor data, and conduct an accurate matching among the objects and RFID tags. For the application layer, the AR applications can use the matching results directly to realize various objectives.

FEATURE SAMPLING AND EXTRACTION

In this section, we investigate the feature sampling and extraction based on the observations from empirical studies. Without loss of generality, in the following each experiment observation is summarized from the statistical properties of 100 re-

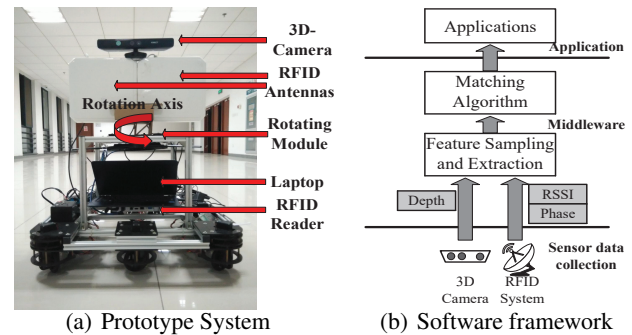


Figure 2. System Framework

peatable experiments. We set a typical indoor environment, i.e., a 10m × 8m lobby, as the testing environment.

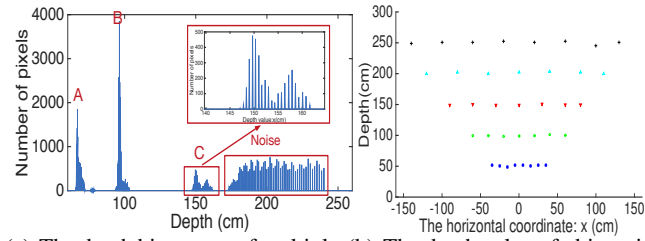
Extract the Depth of Field from Depth-Camera

Depth cameras, such as the Microsoft Kinect, are a kind of range camera, which produces a 2D image showing the distance to points in a scene from a specific point, normally associated with a depth sensor. Therefore, the depth camera can effectively estimate the distance to a specified object according to the depth, because the depth is linearly increasing with the distance. If multiple objects are placed in different positions in the scene, they are usually at different distances away from the depth camera. Therefore, it is possible to distinguish among different objects according to the depth values from the depth camera.

Experiment Observations

We first conduct experiment to evaluate the characteristics of the depth. We arbitrarily place three objects *A*, *B*, and *C* in front of the depth camera, i.e., Microsoft Kinect, object *A* is a box at distance 68cm, object *B* is a can at distance 95cm and object *C* is a tripod at distance 150cm. We then collect the depth histogram from the depth sensor. As shown in Figure 3(a), the *X*-axis denotes the depth value, and the *Y*-axis denotes the number of pixels at the specified depth. We find that, as *A* and *B* are regular-shaped objects, there are respective peaks in the depth histogram for object *A* and *B*, meaning that many pixels are detected from this distance. Therefore, *A* and *B* can be easily distinguished according to the distance. However, there exist two peaks in the corresponding distance of object *C*, because object *C* is an irregularly-shaped object (the concave shape of the tripod), there might be a number of pixels at different distances. Moreover, we can also find some background noises past the distance of 175 cm, which can be produced by background objects, such as the wall and floor. This implies that, for the object with a continuous surface, the depth sensor usually detects a peak in the vicinity of its distance, for an irregularly-shaped object, the depth sensor detects multiple peaks with intermittent depths. Nevertheless, we find that these peaks are usually very close in distance.

In order to further validate the relationship between the depth and distance, we set multiple horizontal lines with different distances to the Kinect (from 500 mm to 2500 mm). For each horizontal line, we then move a certain object along the line and respectively obtain the depth value from the Kinect. We show the experiment results in Figure 3(b). Here we find



(a) The depth histogram of multiple objects (b) The depth value of objects in different horizontal lines

Figure 3. Experiment results of depth value

that, for each horizontal line, the depth values of the object keep nearly constant, with rather small deviation; for different horizontal lines, these depth values have obvious variations. Due to the limitation of the Kinect's view, the Kinect has smaller view angle in closer distance. This observation implies that, the depth value collected from the depth cameras depicts the vertical distance rather than the absolute distance between the objects and the depth camera.

Depth Feature Extraction

To extract the depth of specified objects from the depth histogram of multiple objects, we set a threshold t to detect the peaks in regard to the number of pixels. We thus iterate from the minimum depth to the maximum depth in the histogram, if the number of pixels for a certain depth is larger than t , we identify it as a peak $p(d_i, n_i)$ with the depth d_i and the number of pixels n_i . In order to address the multiple-peaks problem of irregularly-shaped objects, we set another threshold Δd . If the differences of these peaks' depth values are smaller than Δd , we then combine them as one peak. Both the value of t and Δd are selected based on the empirical value from a number of experimental studies ($t=200$ and $\Delta d=10\text{cm}$ in our implementation). Then, each peak actually represents a specified object. For each peak, we respectively find the leftmost depth d_l and the rightmost depth d_r with the number of pixels $n_r > 0$. We then compute the average depth for the specified object as follows: $d = \sum_{i=l}^r (d_i \times \frac{n_i}{\sum_{i=l}^r n_i})$. The average depth is calculated in a weighted average approach according to the number of pixels for each depth around the peak.

Extract the Received Signal Strength from RF-signals

The received signal strength (RSSI) measures the power of received radio signal, which is inversely proportional to the distance between the tag and the reader. However, according to the previous study [34], the RSSI is impacted by various issues like multi-path effect, and path loss, etc. This indicates that the RSSI does not always have a monotonic relationship with the distance. Therefore, with the RSSI from a specified tag, the RFID system can roughly estimate the distance between the reader and the tag.

Experiment Observations

It is found that, inside the RFID antenna's effective scanning range, the RSSI from the tag is also impacted by its position offset from the center of the antenna beam. In order to validate the above judgment, we separate the RFID reader and the tag with a distance d , and then we evaluate the average RSSI value by gradually rotating the antenna from an offset degree of -40° to $+40^\circ$. Figure 4 shows the experiment results. We find

that, as the distance between the tag and the reader increases from 50 cm to 150 cm, the RSSI decreases rapidly; when the distance further increases, the RSSI then decreases slowly. Moreover, in regard to a certain distance, the RSSI from the tag always reaches the maximum value when the antenna is directly facing towards the tag. As we further increase the offset degree in rotation, the RSSI gradually decreases. This is because the antenna outputs the maximum transmitting power in the central area of the beam, and thus the RSSI of the backscattered RF-signals reaches the maximum value when the tag is in the center. As the tag's position is deviated from the center of the antenna beam, the RSSI of the backscattered RF-signals thus decreases. We call the position of achieving the peak value in RSSI the *perpendicular point*, since the perpendicular bisector of the RFID antenna crosses this point.

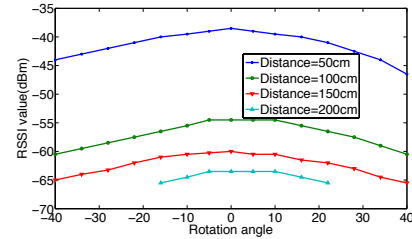


Figure 4. The variation of RSSI via rotating the RFID antenna

Although the RSSI can only be used to measure the vertical distance between the tag and the antenna in a coarse granularity, nevertheless, with different offset degrees from the tag to the center of antenna beam, the RSSI changes in a convex curve with the peak value at the *perpendicular point*. We can further leverage this property to differentiate the positions of various objects in the horizontal aspect.

Extract the Phase Value from RF-Signals

Background

Phase is a basic attribute of a signal along with amplitude and frequency. The phase value of an RF signal describes the degree that the received signal offsets from the sent signal, ranging from 0 to 360 degrees. Let d be the distance between the RFID antenna and the tag, the signal traverses a round-trip with a distance of $2d$ in each backscatter communication. Therefore, the phase value θ output by the RFID reader can be expressed as [25, 4]:

$$\theta = \left(\frac{2\pi}{\lambda} \times 2d + \mu \right) \bmod 2\pi, \quad (1)$$

where λ is the wave length. Besides the RF phase rotation over distance, the reader's transmitter, the tag's reflection characteristic, and the reader's receiver will also introduce some additional phase rotation, denoted as θ_T , θ_R and θ_{TAG} respectively. We use $\mu = \theta_T + \theta_R + \theta_{TAG}$ to denote this diversity term in Eq. (1). Since μ is rather stable according to the previous results [32], and it is only related to the physical properties of the specified tag-antenna pair, we can record μ for different tags in advance. Then, according to each tag's response, we can calibrate the phase value by offsetting the diversity term. Thus, the phase value can be used as an accurate and stable metric to measure distance.

Estimate the Vertical Distance from Phase Value

According to the definition in Eq. (1), the phase value is a periodical function of the distance. Hence, given a specified

phase value from the RF-signal, there can be multiple solutions for estimating the distance between the tag and antenna. Therefore, we can deploy an RFID antenna array to scan the tags from slightly different positions, so as to figure out the unique solution of the distance. Without loss of generality, in this paper, we separate two RFID antennas with a distance of d , and use them to scan the RFID tags and respectively obtain their phase values from the RF-signals, as shown in Figure 5.

Since the depth value from the depth cameras like Kinect measures the vertical distance, instead of the absolute distance between the objects and the depth camera, in order to achieve a perfect match between the collected RF-signals and the depth of field, it is essential to measure the vertical distance between the tags and RFID antennas. However, it is rather difficult to directly measure the vertical distance via the phase value. Figure 5 shows the relationship between the vertical distance and the absolute distance. In regard to a specified RFID tag, suppose its absolute distances to Antenna 1 and Antenna 2 are respectively d_1 and d_2 , then we need to derive its vertical distance h to the antenna pairs.

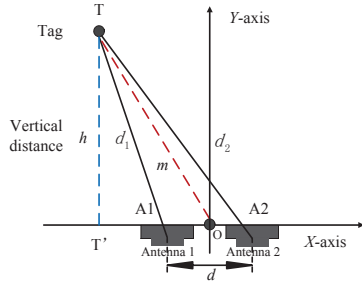


Figure 5. Compute the (x, y) coordinate of the tag

If we respectively use A_1 and A_2 to denote the midpoint of Antenna 1 and Antenna 2, and use T to denote the position of the tag, as a matter of fact, the three sides of $\langle T, A_1 \rangle$, $\langle T, A_2 \rangle$, and $\langle A_1, A_2 \rangle$ form a triangle. Since Antenna A_1 and Antenna A_2 are separated with a fixed distance d , according to *Heron's formula* [12], the area of this triangle is $A = \sqrt{s(s-d_1)(s-d_2)(s-d)}$, where s is the semiperimeter of the triangle, i.e., $s = \frac{(d_1+d_2+d)}{2}$. Moreover, since the area of this triangle can also be computed as $A = \frac{1}{2}h \times d$, we can thus compute the vertical distance $h = \frac{2\sqrt{s(s-d_1)(s-d_2)(s-d)}}{d}$. Then, according to the Apollonius' theorem [7], for a triangle composed of point A_1, A_2 and T , the length of median TO bisecting the side A_1A_2 is equal to $m = \frac{1}{2}\sqrt{2d_1^2 + 2d_2^2 - d^2}$. Hence, the horizontal distance between the tag and the midpoint of the two antennas, i.e., $T'O$, should be $\sqrt{m^2 - h^2}$. Therefore, if we build a local coordinate system with the origin set to the the midpoint of the two antennas, the coordinate (x', y') is computed as follows:

$$x' = \begin{cases} \sqrt{\frac{1}{2}d_1^2 + \frac{1}{2}d_2^2 - \frac{1}{4}d^2 - h^2} & d_1 \geq d_2 \\ -(\sqrt{\frac{1}{2}d_1^2 + \frac{1}{2}d_2^2 - \frac{1}{4}d^2 - h^2}) & d_1 < d_2 \end{cases} \quad (2)$$

$$y' = h. \quad (3)$$

Therefore, the next problem we need to address is to estimate the absolute distance between the tag and antenna according to

the extracted phase value from RF-signals. Suppose the RFID system respectively obtains two phase values θ_1 and θ_2 from two separated RFID antennas, then, according to the definition in Eq. (1), the possible distances from the tag to the two antennas are: $d_1 = \frac{1}{2} \cdot (\frac{\theta_1}{2\pi} + k_1) \cdot \lambda$, and $d_2 = \frac{1}{2} \cdot (\frac{\theta_2}{2\pi} + k_2) \cdot \lambda$. Here, k_1 and k_2 are integers ranging from 0 to $+\infty$. Due to the multiple solutions of k_1 and k_2 , there could be multiple candidate positions for the tag. However, since the difference of the lengths of two sides is smaller than the length of the third side in a triangle, i.e., $|d_1 - d_2| < d$, we can leverage this constraint to effectively eliminate many infeasible solutions of k_1 and k_2 . Besides, due to the limited scanning range of the RFID system (the maximum scanning range l is usually smaller than 10 m), the value of k_1 and k_2 should be upper bounded by a certain threshold, i.e., $\frac{2l}{\lambda}$.

Figure 6 shows an example of feasible positions of the target tag according to the obtained phase values θ_1 and θ_2 . The feasible solutions include multiple positions like $A \sim D$, which respectively belong to two hyperbolas H_1 and H_2 . Due to the existence of multiple solutions, we can use these hyperbolas to denotes a superset of these feasible positions in a straightforward approach.

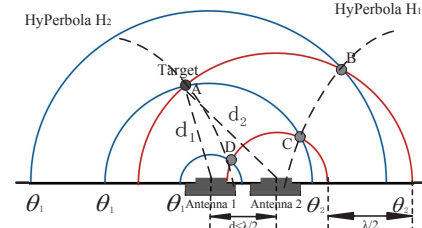


Figure 6. Estimate the distance from phase value of RF signals

MATCHING ALGORITHM VIA CONTINUOUS SCANNING

Motivation

To identify and distinguish the multiple tagged objects, a straightforward solution is to scan the tags in a static approach, where both the depth camera and RFID antenna(s) are deployed in a fixed position without moving. The system scans the objects and tags simultaneously and respectively collect the depth value and RF-signals from these tagged objects. We can further pair the tags with the objects accordingly. However, when multiple tagged objects are placed at close vertical distance to the system, this solution cannot effectively distinguish multiple tagged objects in different horizontal distances.

To address this problem, we propose a continuous scanning-based solution as follows: we continuously rotate the scanning system (including the depth camera and RFID antennas), and simultaneously sample the depth of field and RF-signals from multiple tagged objects. Hence, we are able to collect a continuous series of features like depth, RSSI and phase values during continuous scanning. While the scanning system is rotating, the vertical distances between multiple objects and the scanning system are continuously changing, from which we can further derive the differences of multiple tagged objects in different horizontal distances. In this way, we are able to further distinguish multiple tagged objects with close vertical distance but in different positions.

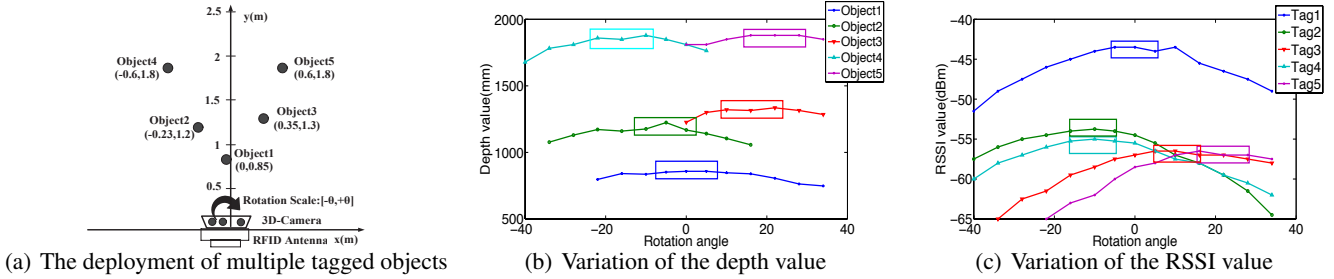


Figure 7. The experiment results of continuous scanning

Extract Depth via Continuous Scanning

In this section, we present our approach to extract the depth series via continuous scanning, so as to derive both the vertical distance and the horizontal distance of the tagged objects.

During the continuous scanning, we continuously rotate the depth camera from the angle of $-\theta$ to $+\theta$ and use it to scan the multiple tagged objects. During this process, as the vertical distance between the specified objects and the depth camera is continuously changing, the depth values collected from these objects are also continuously changing. We conduct experiments to validate this judgment. As shown in Figure 7(a), we arbitrarily deploy multiple tagged objects within the effective scanning range, the coordinates of these objects are also labeled. We continuously rotate the depth camera from the angle of -40° to $+40^\circ$ and collect the depth values from multiple tagged objects for every $5\sim 6$ degrees. Figure 7(b) shows the experiment results. We use the method of quadratic curve fitting to connect the depth values as a curve for a certain object. We find that the series of depth values for each object actually form a convex curve with a peak value. This peak value denotes the snapshot when the vertical distance reaches the maximum value. It appears only when the perpendicular bisector of the depth camera crosses the specified object, since the vertical distance reaches the value of the absolute distance between the object and the depth camera, which is the theoretical upper bound it can achieve. In other words, the peak value appears when the depth camera is right facing towards the object, we call this *perpendicular point*.

In this way, according to the peak value of depth, we are able to further distinguish multiple objects with the same vertical distance but different positions. The solution is as follows: After the system finishes continuous scanning, it extracts the peak value from the curve of each object's depth value. Then, we label each object with the coordinate of its peak value, i.e., $\langle \theta, d \rangle$, where θ represents the rotation angle and d represents the depth value. Therefore, as the depth d denotes the vertical distance of objects, we can use the depth to distinguish the objects in the vertical dimension; as the rotation angle θ denotes the angle for the camera to meet the *perpendicular point*, we can use the angle to distinguish the objects in the horizontal dimension. They can be easily distinguished from the horizontal dimension.

Pair the Tags with Objects according to Depth and RSSI

It is known that the RSSI is not a very reliable metric to accurately measure the distance between the tags and the antennas, as it is easy to be impacted by the environmental factors like

multi-path fading and path loss. However, since most mid- and low-end COTS RFID systems can only extract the RSSI from RF-signals, we need to figure out a solution based on RSSI. In this section, we present our approach to pair the tags with objects according to the correlations between the depth and RSSI in continuous scanning.

According to the observations from Figure 4, with different offset degrees from the tag to the center of antenna beam, the RSSI changes with a fixed variation pattern. This implies that, if we conduct the continuous scanning to the tagged objects, the RSSI from the tag always reaches the maximum value when the antenna is right facing towards the tag. This variation pattern of RSSI is quite similar to the depth value, since they both reach the peak value when the tagged object is at the *perpendicular point* of depth camera/RFID antenna. We further conduct experiments to validate the above judgment. Using the deployment in Figure 7(a), we continuously rotate the RFID antenna from the angle of -40° to $+40^\circ$ and collect the RSSI from multiple tags. As shown in Figure 7(c), the variation of RSSI for each tag has very similar features as depth: during the continuous scanning, the RSSI first increases to a maximum value, and then further decreases to a certain value. The only difference is that the RSSI is inversely corresponding with the depth value for any specified object, i.e., the larger the RSSI, the smaller the depth. Therefore, we can also label each tag with the coordinate of its peak value, i.e., $\langle \theta, r \rangle$, where θ represents the rotation angle and r represents the RSSI. We can respectively use the RSSI r and the rotation angle θ to distinguish the tags in vertical and horizontal dimensions.

Therefore, in order to pair multiple tags with multiple objects, we propose a matching solution in Algorithm 1. Our goal is to find a matching between two disjoint sets O and T according to the correlation of their measurements. After we extract the vector from the measured data, for any object O_i with vector $\langle \theta_i, d_i \rangle$, we first select the candidate tags for pairing according to the angle θ_i . We set all tags as pairing candidates with their angles in the range $[\theta_i - \delta, \theta_i + \delta]$ ($\delta = 5^\circ$ in our implementation). Then we further compare their values in RSSI and depth. As the RSSI and the depth are measured in different dimensions, e.g., the depth value is linearly correlated to the distance, while the RSSI is nonlinearly correlated to the distance, it is not reasonable to compare them directly. We thus match each object to a candidate tag based on their relative rank in RSSI and depth. After that, since multiple objects may be matched to one tag, we make the tag select the object with the closest rank as the final pair. This process then iterates until all the objects and tags are paired.

Algorithm 1 Match multiple objects to multiple tags

- 1: **Extract the vector:** After continuous scanning, we respectively identify the peak value from the quadratic fitting curve of depth and RSSI. For each object O_i , we label it with a vector $\langle \theta_i, d_i \rangle$, and add the vector to a set O ; for each tag T_j , we label it with a vector $\langle \theta_j, r_j \rangle$, and add the vector to a set T .
- 2: **while** $O \neq \emptyset$ and $T \neq \emptyset$ **do**
- 3: **Match the objects and tags:** For each object $O_i \in O$ with vector $\langle \theta_i, d_i \rangle$, respectively add those objects $O_j \in O$ and those tags $T_j \in T$ with $\theta_j \in [\theta_i - \delta, \theta_i + \delta]$ into the set O_c and T_c . In regard to the depth value d_i , compute the rank of O_i in the set O_c as k . Select the tag $T_j^* \in T_c$ with the rank of k in regard to the RSSI r_j , and pair the object O_j with the tag T_j^* .
- 4: **Calibrate the matching results:** For any tag $T_j \in T$ paired with multiple objects, select the object O_i from these objects with the closest rank similarly, and pair the object O_i with the tag T_j^* . Respectively remove the object O_i and the tag T_j from set O and T .
- 5: **end while**
- 6: Output the matched pairs of objects and tags.

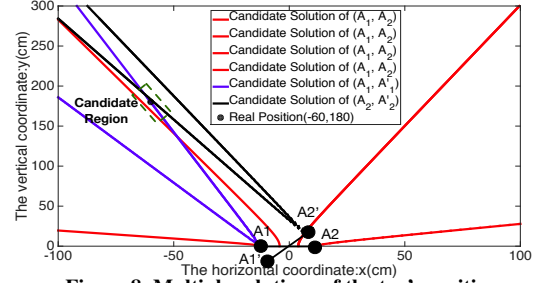
Pair the Tags with Objects according to Depth and Phase

Since a new brand of COTS RFID systems, like the ImpinJ, are able to extract the phase value from the RF-signals of tags, it provides us a new opportunity to differentiate the positions of the tagged objects with a more accurate approach. In this section, we present our approach to pair the tags with objects according to the correlations between the depth and phase in continuous scanning.

According to the analysis shown in Figure 6, given the two phase values of RF-signals extracted from two antennas separated with a distance d ($d=25\text{cm}$ in our implementation), there could be multiple solutions for the tag's position, which could be represented with multiple hyperbolas in the two-dimensional space. In fact, we can leverage continuous scanning to figure out a unique solution by filtering out those unqualified solutions. The idea is as follows: for each snapshot t_i ($i = 1 \sim m$) of the continuous scanning, for a specified tag T , we can respectively extract the phase values (θ_1, θ_2) from the two antennas, then compute the feasible distances (d_1, d_2) between the tag and two antennas. We further compute the set of feasible positions in a global coordinate system as S_i . Then, by computing the intersection of different sets S_i for all snapshots, we are able to figure out a unique solution for the tag's position as follows: $S = \bigcap_{i=1}^m S_i$.

As a matter of fact, as long as two pairs of phase values are obtained, we are able to further derive the unique solution of the tag's position by computing the intersection of multiple feasible solutions. Figure 8 shows an example of deriving the unique solution. Suppose a target tag is deployed at the coordinate $(-60, 180)$. We first obtain the phase values $(2.58, 5.81)$ from the two antennas when they are respectively at the position of A_1 and A_2 . After the antenna pairs are rotated with a degree of 40° , we then obtain the phase values $(5.56, 2.49)$ from the two antennas when they are respectively at the position of A'_1 and A'_2 . In this way, we can obtain three pairs of phase values $(2.58, 5.81)$, $(2.58, 5.56)$ and $(5.81, 2.49)$, which are respectively collected from antenna pairs $\langle A_1, A_2 \rangle$, $\langle A_1, A'_1 \rangle$, and $\langle A_2, A'_2 \rangle$. We can respectively use them to compute the feasible solutions in a unified coordinate system. As shown in Figure 8, we use different colors to label the hyperbolas

of multiple feasible solutions according to different pairs of phase values. It is found that the multiple hyperbolas of different feasible solutions all intersect at a small area which is very close to the target tag's real position. We thus set the central point of the intersection region as the estimate value of the tag's position.

**Figure 8. Multiple solutions of the tag's position**

After deriving the target tag's position, we can further derive the angle when the tag is at the *perpendicular point* of the RFID antennas, that is the moment when the perpendicular bisector of the midpoint of the antenna pairs crosses the tag. We use the pair $\langle \theta, \delta \rangle$ to denote this situation, here θ denotes the offset angle of the antenna, and δ denotes the vertical distance. The pair $\langle \theta, \delta \rangle$ is computed as follows: $\theta = \arctan \left| \frac{x}{y} \right|$, and $\delta = \sqrt{x^2 + y^2}$. Therefore, we can further leverage an algorithm like Algorithm 1 to match multiple tags to multiple objects. The only difference is that we can directly pair the objects $O_i \in O$ with the tags $T_j \in T$ according to the distance between the vector $\langle \theta_i, d_i \rangle$ and the vector $\langle \theta_j, \delta_j \rangle$, since they can accurately estimate the positions of the objects/tags.

Discussion*Robustness*

Due to the environmental issues like the multi-path fading and object occlusion, the system may fail to identify some of the objects and the tags. Moreover, in some situations, it is essential to isolate the recognizable object with non-recognizable ones. Hence, it is possible that the cardinality of objects identified by the depth camera is not equal to the cardinality of tags identified by the RFID antenna. This leads to imperfect matching between the objects and tags. Our solution is able to tackle this problem by using the *two-dimensional matching* method with regression analysis. By means of continuous scanning via rotation, in regard to the tags and objects, we can derive their vertical distances in the *horizontal dimension* and horizontal distances in the *horizontal dimension*, respectively from the depth camera and the RFID antenna. Then we perform the regression analysis on the vertical distances and horizontal distances between the tags and objects, and filter out those outliers according to the regression model. After that, we pair the tags with objects according to their two dimensional positions. This approach effectively mitigates the interference from those tags and objects which fail to be identified and isolates the recognizable object with non-recognizable ones.

Scalability

Our technical solution is primarily based on the distinction of depth (vertical distance) from the tagged objects. However, even if multiple objects are of the same depth, our solution is still able to distinguish these objects via continuous

scanning. By leveraging continuous scanning, our solution is able to effectively distinguish the tagged objects in terms of both vertical distance and horizontal distance. Moreover, in real applications, since the tagged objects are deployed in 3-dimensional space, two tagged objects can be at the same coordinate (x, y) but at different heights in z . Our solution can effectively scale to this situation by conducting continuous scanning in a 2-dimensional space, i.e., continuously scan the tagged objects by rotating up and down, and from left to right. In this way, the system is able to distinguish multiple objects of different positions in 3-dimensional space.

Time Delay

Since the number of tagged objects cannot be too large in real applications, the computation complexity of our algorithm is fairly low, hence the time delay of our solution mainly lies in the process of continuous scanning. Therefore, we can reduce the time delay via the following two approaches: 1) Increase the rotation speed of the continuous scanning system, such that the time delay in rotation is reduced. 2) Appropriately decrease the number of samples during the continuous scanning, without too much loss in the accuracy of distinguishing multiple tagged objects, such that the time delay in sampling is reduced. In fact, our depth-phase-based pairing approach only requires to sample twice during the continuous scanning, which greatly reduces the time-delay in scanning.

PERFORMANCE EVALUATION

Experiment Settings

We evaluated our system using one Microsoft Kinect for windows, one ImpinJ R420 reader, two Laird S9028 RFID antennas, and multiple ImpinJ E41-B general purpose tags. We deploy multiple objects in an area of about $3\text{m} \times 3\text{m}$, and attach each tag to an object. We use the Kinect as the depth-camera and use the RFID reader to scan the tags. The average distance between the tagged objects and the system is 2 m. We implement four schemes for performance comparison:

- 1) *Static Scanning via Depth-RSSI Pairing (SS-RSSI)*: The system scans the tagged objects once at a fixed position, and pairs the tags with the objects according to their partial orders respectively in collected depth and RSSI.
- 2) *Hybrid Scanning via Depth-Phase Pairing (HS-Phase)*: The depth camera continuously rotates and scans the tagged objects, while the RFID antennas scan the tagged objects once at a fixed position, and pairs the tags with the objects according to the extracted depth and phase.
- 3) *Continuous Scanning via Depth-RSSI Pairing (CS-RSSI)*: The system continuously scans the tagged objects while it is rotating, and pairs the tags with the objects according to the extracted series of depth and RSSI.
- 4) *Continuous Scanning via Depth-Phase Pairing (CS-Phase)*: The system continuously scans the tagged objects while it is rotating, and pairs the tags with objects according to the extracted series of depth and phase.

Evaluate the Accuracy in Pairing the Tags with Objects

We run experiments to evaluate the accuracy in pairing the tags with the objects. Without loss of generality, by default we deploy 10 tagged objects in the scanning area. We vary the settings of the average horizontal/vertical distance, and the cardinality of tagged objects. For each setting, we randomly

generate 10 types of deployments for the tagged objects, and evaluate the average match ratio for successful pairing in the above four schemes.

Accuracy for different cardinalities of tagged objects

Our solution achieves good performance in accuracy when the cardinalities of tagged objects are varied from 3 to 15. We first deploy 10 tagged objects in the scanning area, and set the average horizontal/vertical distance among the objects to 30 cm, thus the average density is 11 objects/ m^2 , which is a fairly large density for conventional applications. As shown in Figure 9(a), we find that both CS-RSSI and CS-Phase achieve much better performance than SS-RSSI and HS-Phase, e.g., the match ratios of CS-RSSI and CS-Phase are respectively 75% and 91%, while the match ratios of SS-RSSI and HS-Phase are only 25% and 40%. We further evaluate the match ratio for pairing different cardinalities of tagged objects, by varying the cardinality of tagged objects from 3 to 15. As shown in Figure 9(b), as the cardinality increases from 3 to 15, the match ratios of SS-RSSI and HS-Phase decrease in a rapid approach, whereas the match ratios of CS-RSSI and CS-Phase decrease slowly. Nevertheless, CS-RSSI and CS-Phase respectively achieve a match ratio of 60% and 77% when the cardinality of tagged objects is 15.

Accuracy for different vertical/horizontal distances

Our solution achieves good performance in accuracy when the vertical/horizontal distances are varied from 10 cm to 50 cm. We respectively vary the average vertical distances and horizontal distances among the tagged objects, thus to further evaluate the performance in accuracy. We fix the average horizontal (vertical) distance among the objects to 30 cm, and vary the average vertical (horizontal) distance from 10 cm to 50 cm. Figure 9(c) and Figure 9(d) show the match ratios with different vertical distances and horizontal distances, respectively. We find that as the average vertical/horizontal distance decreases, the match ratios of all schemes gradually decrease. Besides, for the same vertical distance and horizontal distance, the match ratio of the former situation is apparently less than the latter situation, since the vertical distance is more difficult to estimate than the horizontal distance. Nevertheless, CS-Phase respectively achieves a match ratio of 68% and 72% when the average vertical/horizontal distance is 10cm, which is corresponding to a rather high density for the tagged objects, i.e., 33 objects/ m^2 .

Evaluate the Robustness in Pairing the Tags with Objects

Robustness to missing tags/objects

Our solution achieves good performance in robustness with different ratios of missing objects/tags ranging from 10% to 50%. We run experiments to evaluate the robustness to missing tags/objects, when there exist several objects or tags which fail to be identified. Here we measure the match ratio for the remaining objects or tags. Figure 9(e) and Figure 9(f) show the experiment results for different ratios of missing objects and tags, respectively. As the ratio of missing objects/tags increases from 10% to 50%, the match ratios for all schemes decrease in most cases, except that in some cases, the match ratio of SS-RSSI and HS-Phases slightly increase, since the number of objects/tags for pairing is reduced. Nevertheless, CS-RSSI and CS-Phase respectively achieve a match ratio of

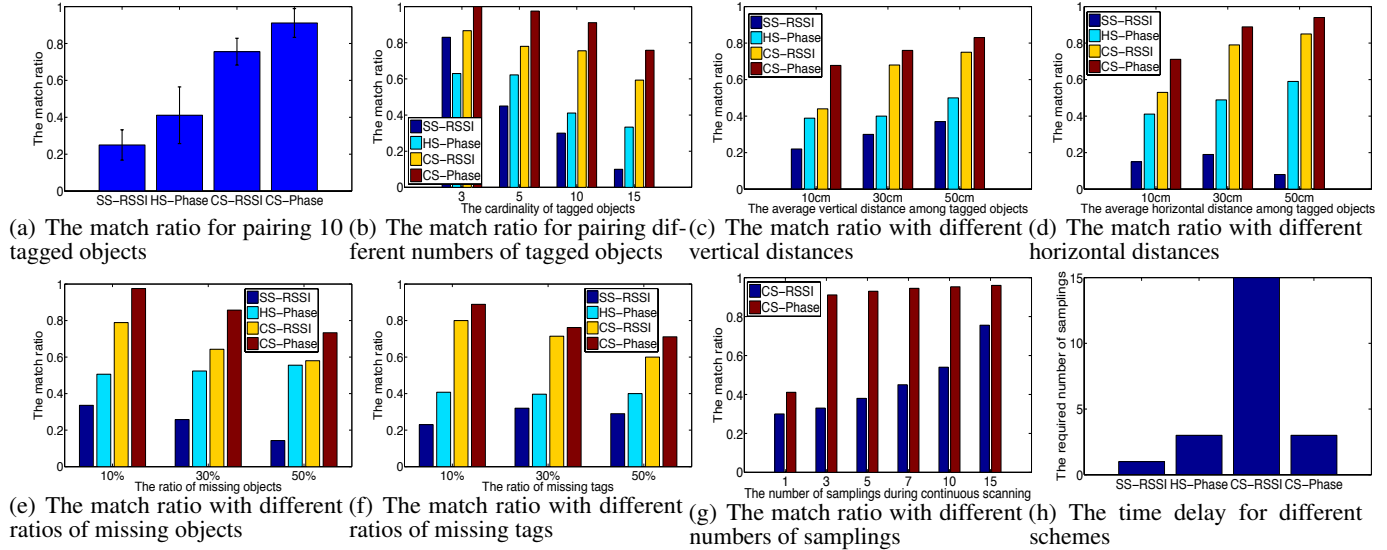


Figure 9. The experiment results

near 60% and 72% when the ratio of missing objects/tags is even 50%.

Robustness to different numbers of samplings

Our solution achieves good performance in robustness with different numbers of samplings ranging from 3 to 15 during continuous scanning. Figure 9(g) shows the experiment results. As the number of samplings increases from 1 to 15, we find that the match ratio of CS-RSSI rapidly increases from 30% to 75%, while the match ratio of CS-Phase first rapidly increases to 91% when the number of samplings is 3, then slowly increases to 96% when the number of samplings is 15. This implies that CS-Phase is more robust to the low sampling situation than CS-RSSI, since CS-Phase requires only a few phase-pair samples to figure out the position according to the intersections of multiple hyperbolas.

Evaluate the Time Efficiency

The time efficiency mainly depends on the number of samplings and the rotation speed in continuous scanning. Since the rotation speed is device-dependent, we thus evaluate the time-efficiency via the number of samplings. As shown in Figure 9(h), SS-RSSI achieves the least time delay, as it only requires to scan once, whereas CS-RSSI achieves the most time delay, as it requires to scan multiple times to find the peak point via continuous scanning, HS-Phase and CS-Phase achieve the medium time delay, as basically 3~4 samplings is enough for them to estimate the position of tagged objects.

CASE STUDY: RECOGNIZE MULTIPLE TAGGED HUMAN SUBJECTS IN THE CAFE

In order to further evaluate the real performance of our system by considering more practical issues (e.g. indoor multi-path and energy absorption), we do more thorough experiments in a more realistic setting.

User Interface: In this case study, a major task of our system is to recognize multiple tagged human subjects in the cafe and further show their inherent information in the camera's view,

as shown in Figure 10. Thus, we have implemented an application which were executed on a SAMSUNG PC equipped with an Intel(R) Core(TM) I5 1.4GHz CPU and 4G RAM. The PC is remotely connected to the system via WiFi. Figure 11 shows an example user interface of our application. The left window shows the camera's view from the Kinect, while the right window shows the detailed description of the specified object. Once the button "Scan" is pressed, the system runs our continuous scanning-based algorithm to match the objects with the tags, and then draws multiple bounding boxes on the camera's view based on the scanning results. Each bounding box (in blue color) is actually a rectangle which distinguishes the object from the background based on the color/depth gradient between the object and background. When the specified bounding box is further clicked, the detailed information such as the ID photo, name, age, job and interests are displayed in the right window.

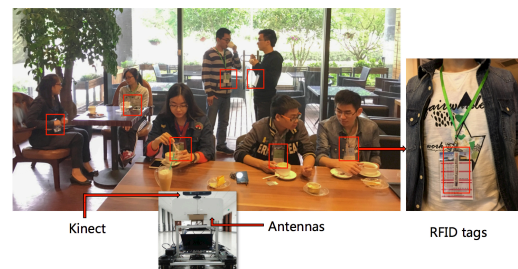


Figure 10. Example deployment of multiple human subjects wearing RFID badges in the cafe



Figure 11. Example application interface

Experiment Settings: As shown in Figure 10, we let multiple human subjects (4~8 people) stand or sit freely in the cafe, while wearing the RFID tagged badges. These “tagged” human subjects are thus different in terms of heights, horizontal distance and vertical distance. Besides, they can be slightly moving or turning with a limited speed or angle. It raises more challenges than the free-space testing, since the human body may lead to many interferences like multi-path effect and energy absorption. We conducted experiments to evaluate the performance of match ratios, by varying the factors like the number of human subjects, the spacing between human subjects, and the moving state. We deploy our system in front of the human subjects with a distance of 1.5~3m. The default number of human subjects and the default average spacing is respectively 6 and 60 cm.

Performance Evaluation: *Our solution can achieve fairly good matching accuracy to recognize multiple tagged human subjects of different factors like the height, spacing, moving state, etc.* Figure 12(a)-(d) respectively shows the match ratios with different configurations. Without loss of generality, we show the matching results of 5 randomly generated deployments with different spacing and heights of the human subjects. In the first experiment, we let the human subjects remain stationary, i.e., standing or sitting still, and evaluate the match ratios. As shown in Figure 12(a), our solution achieves a match ratio of 50% and 80% respectively with CS-RSSI and CS-Phase. In the second experiment, we let the human subjects keep in slightly moving state, i.e., they may be moving or turning with a limited speed (<40cm/s) or angle (<30°/s). As shown in Figure 12(b), our solution achieves a match ratio of 60% and 74% respectively with CS-RSSI and CS-Phase. In the third experiment, we vary the average spacing between the human subjects from 60cm to 90cm. As shown in Figure 12(c), our solution achieves an average match ratio of over 50% and 75% respectively with CS-RSSI and CS-Phase. In the fourth experiment, we vary the number of human subjects from 4 to 8. As shown in Figure 12(d), our solution achieves an average match ratio of over 45% and 70% respectively with CS-RSSI and CS-Phase. The performance reduction of CS-RSSI in the above experiments is mainly due to the energy absorption of human bodies, which distracts the conventional distribution of RSSI in RF-signals. Nevertheless, CS-Phase always achieves fairly good performance since the phase in RF-signals is irrelevant to the energy absorption problems.

User Experience Evaluation: We invite a total of 44 people (28 males and 16 females with different technical backgrounds, their ages range from 20 to 58) to use our system in the augmented reality applications, and evaluate their user experience via the questionnaire surveys, including 1) application meaning, 2) technical complexity, 3) accuracy and 4) friendliness of user interface. Figure 13 shows the evaluation results. For the application meaning, most of the people have positive/very positive evaluation, they believe it is a promising approach for future augmented reality application. For the technical complexity, several people have some negative evaluation, this is mainly because the current prototype system is fairly huge in size, and the RSSI/phase-based continuous scanning method may not be so intuitive for users with various technical

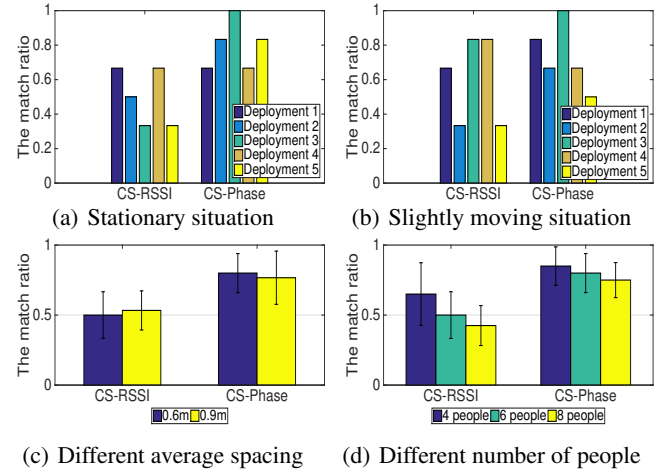


Figure 12. Evaluate the match ratios backgrounds. For the accuracy, most of the people have positive/very positive evaluation, since in most cases our solution can achieve very good performance in accuracy. For the friendliness of user interface, most of the people have positive/very positive evaluation. due to the interesting vet simple design.

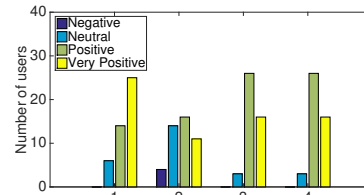


Figure 13. Evaluation of the user experience: 1) application meaning, 2) technical complexity, 3) accuracy, 4) friendliness of user interface.

CONCLUSION AND FUTURE WORK

In this paper, we design an RFID-based system to identify and distinguish multiple RFID tagged objects in an augmented reality system. We deploy additional RFID antennas to the COTS depth camera, and propose a continuous scanning-based scheme to distinguish multiple tagged objects. The current implementation is a proof-of concept prototype for the “tell me what I see” vision. The size of the system is huge for wearable usages, and the battery usage is high for conventional applications. In the future design, we consider to miniaturize the technical solution and integrate it into the wearable devices. For example, we can miniaturize the RFID antennas and the 3D camera, and integrate them into the wearable helmets/glasses for augmented reality applications. In this way, in order to perform the continuous scanning, the user only need to continuously turn her head from one side to the other side with a certain angle. All the inherent information of the detected objects can be shown on the screen of glasses.

Acknowledgments

This work is supported in part by National Natural Science Foundation of China under Grant Nos. 61472185, 61373129, 61321491, 91218302, 61502224; JiangSu Natural Science Foundation, No. BK20151390; EU FP7 IRSES MobileCloud Project under Grant No. 612212; CCF-Tencent Open Fund. This work is partially supported by Collaborative Innovation Center of Novel Software Technology and Industrialization. The work of Jie Wu was supported in part by NSF grants CNS 1449860, CNS 1461932, CNS 1460971, CNS 1439672, CNS 1301774, and ECCS 1231461.

REFERENCES

1. S. Azzouzi, M. Cremer, U. Dettmar, R. Kronberger, and T. Knie. 2011. New measurement results for the localization of UHF RFID transponders using an Angle of Arrival (AoA) approach. In *Proc. of IEEE RFID*.
2. M. Buettner, R. Prasad, M. Philipose, and D. Wetherall. 2009. Recognizing Daily Activities with RFID-based Sensors. In *Proc. of UbiComp*. 51–60.
3. C. Cerrada, S. Salamanca, E. Perez, J. A. Cerrada, and I. Abad. 2007. Fusion of 3D vision techniques and RFID technology for object recognition in complex scenes. In *IEEE International Symposium on Intelligent Signal Processing*.
4. H. Ding, L. Shangguan, Z. Yang, J. Han, Z. Zhou, P. Yang, W. Xi, and J. Zhao. 2015. FEMO: A Platform for Free-weight Exercise Monitoring with RFIDs. In *Proc. of ACM SenSys*.
5. B. Drost, M. Ulrich, N. Navab, and S. Ilic. 2010. Model globally, match locally: Efficient and robust 3D object recognition. In *Proc. of IEEE CVPR*.
6. G. Galatas and F. Makedon. 2013. A system for multimodal context-awareness. *International Journal of Advanced Computer Science and Applications* (2013).
7. C. Godfrey and A. W. Siddons. 1908. Apollonius' Theorem. *Modern Geometry* (1908), 20.
8. J. Han, L. Shao, D. Xu, and J. Shotton. 2013. Enhanced computer vision with microsoft kinect sensor: A review. *IEEE Transactions on Cybernetics* 43, 5 (2013), 1318–1334.
9. J.D. Hincapié-Ramos, A.n Tabard, and J. E. Bardram. 2011. Mediated Tabletop Interaction in the Biology Lab: Exploring the Design Space of the Rabbit. In *Proc. of UbiComp*. 301–310.
10. ImpinJ 2016. ImpinJ. (2016). <http://www.impinj.com>.
11. S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, and A. Davison. 2011. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In *Proc. of ACM UIST*.
12. K. Kendig. 2000. Is a 2000-Year-Old Formula Still Keeping Some Secrets? *Amer. Math. Monthly* 107 (2000), 402–415.
13. Kinect 2016. Kinect. (2016). <http://www.microsoft.com/en-us/kinectforwindows>.
14. F. Klompaker, H. Fischer, and H. Jung. 2012. Authenticated Tangible Interaction using RFID and Depth-Sensing Cameras. In *Proc. of ACHI*.
15. J. Liu, B. Xiao, K. Bu, and L. Chen. 2014. Efficient Distributed Query Processing in Large RFID-enabled Supply Chains. In *Proc. of IEEE INFOCOM*. 163–171.
16. J. Liu, B. Xiao, S. Chen, F. Zhu, and L. Chen. 2015. Fast RFID grouping protocols. In *Proc. of IEEE INFOCOM*. 1948–1956.
17. T. Liu, L. Yang, Q. Lin, Y. Guo, and Y. Liu. 2014. Anchor-free backscatter positioning for RFID tags with high accuracy. In *Proc. of IEEE INFOCOM*.
18. A. Mian, M. Bennamoun, and R. Owens. 2010. On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes. *International Journal of Computer Vision* (2010).
19. J. Nickels, P. Knierim, B. Königings, F. Schaub, B. Wiedersheim, S. Musiol, and M. Weber. 2013. Find My Stuff: Supporting Physical Objects Search with Relative Positioning. In *Proc. of UbiComp*.
20. S. Nirjon and J. Stankovic. 2012. Kinsight: Localizing and tracking household objects using depth-camera sensors. In *Proc. of IEEE DCOSS*.
21. S. Obdrzalek, G. Kurillo, F. Ofli, R. Bajcsy, E. Seto, H. Jimison, and M. Pavel. 2012. Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population. In *Proc. of IEEE EMBC*.
22. I. Oikonomidis, N. Kyriazis, and A. A Argyros. 2011. Efficient model-based 3D tracking of hand articulations using Kinect.. In *British Machine Vision Conference*, Vol. 1. 3.
23. Z. Ren, J. Yuan, and Z. Zhang. 2011. Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera. In *Proc. of ACM Multimedia*.
24. L. Shangguan, Z. Li, Z. Yang, M. Li, and Y. Liu. 2013. Otrack: Order tracking for luggage in mobile RFID systems. In *Proc. of IEEE INFOCOM*.
25. L. Shangguan, Z. Yang, A. X. Liu, Z. Zhou, and Y. Liu. 2015. Relative Localization of RFID Tags using Spatial-Temporal Phase Profiling. In *Proc. of NSDI*.
26. C. S. Wang, C. L. Chen, and Y. M. Guo. 2013. Real-Time Indoor Positioning System Based on RFID and Kinect. In *Proc. of Information Technology Convergence*.
27. J. Wang, F. Adib, R. Knepper, D. Katabi, and D. Rus. 2013. RF-compass: Robot Object Manipulation Using RFIDs. In *Proc. of ACM MobiCom*.
28. J. Wang and D. Katabi. 2013. Dude, where's my card?: RFID positioning that works with multipath and non-line of sight. In *Proc. of the ACM SIGCOMM*.
29. J. Wang, D. Vasisht, and D. Katabi. 2014. RF-IDraw: Virtual Touch Screen in the Air Using RF Signals. In *Proc. of ACM SIGCOMM*.
30. L. Xia, C. C. Chen, and J. K. Aggarwal. 2011. Human detection using depth information by Kinect. In *Proc. of IEEE CVPRW*. 15–22.
31. L. Xie, Q. Li, X. Chen, S. Lu, and D. Chen. 2013. Continuous scanning with mobile reader in RFID systems: an experimental study. In *Proc. of ACM MobiHoc*.

32. L. Yang, Y. Chen, X. Li, C. Xiao, M. Li, and Y. Liu. 2014. Tagoram: Real-time Tracking of Mobile RFID Tags to High Precision Using COTS Devices. In *Proc. of ACM MobiCom*.
33. L. Yang, Q. Lin, X. Y. Li, T. Liu, and Y. Liu. 2015. See Through Walls with COTS RFID System. In *Proc. of MobiCom*.
34. L. Yang, Y. Qi, J. Fang, X. Ding, T. Liu, and M. Li. 2014. Frogeye: Perception of the slightest tag motion. In *Proc. of IEEE INFOCOM*.