This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2857812, IEEE Transactions on Mobile Computing

TaggedAR: An RFID-based Approach for Recognition of Multiple Tagged Objects in Augmented Reality Systems

Lei Xie, *Member, IEEE*, Chuyu Wang, *Student Member, IEEE*, Yanling Bu, *Student Member, IEEE*, Jianqiang Sun, Qingliang Cai, Jie Wu, *Fellow, IEEE*, and Sanglu Lu, *Member, IEEE*

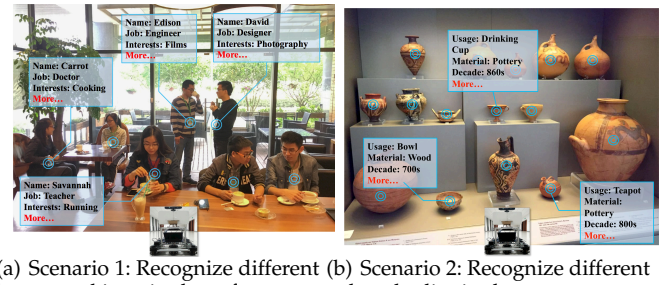
Abstract—With computer vision-based technologies, current Augmented reality (AR) systems can effectively recognize multiple objects with different visual characteristics. However, only limited degrees of distinctions can be offered among different objects with similar natural features, and inherent information about these objects cannot be effectively extracted. In this paper, we propose *TaggedAR*, i.e., an RFID-based approach to assist the recognition of multiple tagged objects in AR systems, by deploying additional RFID antennas to the COTS depth camera. By sufficiently exploring the correlations between the depth of field and the received RF-signal, we propose a rotate scanning-based scheme to distinguish multiple tagged objects in the stationary situation, and propose a continuous scanning-based scheme to distinguish multiple tagged human subjects in the mobile situation. 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 AR system. We have implemented a prototype system to evaluate the actual performance with case studies in real-world environment. 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.

Index Terms—Passive RFID; Augmented Reality System; Object Recognition; Prototype Design

1 Introduction

Augmented Reality (AR) systems (e.g., Microsoft Kinect, Google Glass) are nowadays increasingly used to obtain an augmented view in a real-world environment. For example, by leveraging the computer vision and pattern recognition, depth camera-based devices like the Microsoft Kinect [1] can effectively perform object recognition. Hence, the users can 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 visual characteristics. However, these techniques can only offer a limited degree of distinctions, since multiple objects of the same type may have very similar physical features, e.g., the system cannot effectively distinguish between two laptops of the same brand, even if they belong to different product models. Moreover, they cannot indicate more inherent information about these objects, e.g., the specific configurations, the manufacturers, and production date of the laptop. It is rather difficult to provide these functions by purely leveraging the computer vision-based technology.

- *Lei Xie, Chuyu Wang, Yanling Bu, Jianqiang Sun, Qingliang Cai and Sanglu Lu are with the State Key Laboratory for Novel Software Technology, Nanjing University, China.*
  E-mail: lxie@nju.edu.cn, wangcyu217@dislab.nju.edu.cn, yanling@smail.nju.edu.cn, {SunJQ,caiqingliang}@dislab.nju.edu.cn, sanglu@nju.edu.cn.
- *Jie Wu is with the Department of Computer Information and Sciences, Temple University, USA.*
  E-mail: jiewu@temple.edu.
  Lei Xie and Sanglu Lu are the co-corresponding authors.



(a) Scenario 1: Recognize different human subjects in the cafe (b) Scenario 2: Recognize different cultural relics in the museum

Fig. 1. Typical scenarios of "Tell me what I see" from the AR system

Nevertheless, the RFID technology has brought new opportunities to meet the new demands [2, 3]. The RFID tags can be used to label different objects, and store inherent information of these objects in their onboard memory. 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 have very similar natural features from the visual sense. Fig. 1 shows two typical application scenarios. The first scenario is to recognize different human subjects in the cafe, as shown in Fig. 1(a). 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 such as Kinect can recognize multiple human subjects, and capture 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

1536-1233 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2857812, IEEE Transactions on Mobile Computing

phase and RSSI from the RFID tags. By pairing these information together, the vision of "tell me what I see" can be effectively realized in the AR system. In comparison to the *pure* AR system, which can only show some basic information like the gender and race according to the vision-based pattern recognition, by leveraging this novel RFID assisted AR technology, the inherent information such as their names, jobs and titles can be directly extracted from the RFID tags and associated with the corresponding human subjects in the camera's view. For example, when we are meeting multiple unknown people wearing RFID badges in public events, the system can effectively help us recognize these people by illustrating the detailed information on the camera's view in a smart glass. The second scenario is to recognize different cultural relics in the museum, as shown in Fig. 1(b). In this scenario, multiple cultural relics like the ancient potteries are placed on the display racks. Due to the same craftsmanship, they might have very similar natural features like the color and shape from the visual sense. This prohibits the *pure* AR system from distinguishing different objects when they have very similar physical features. In contrast, using our RFID assisted AR technology, these objects can be easily distinguished according to the differences in the labeling tags. In summary, the advantages of RFID assisted AR systems over the *pure* AR systems lie in the essential capability of *identification* and *localization* in RFID.

Although many schemes for RFID-based localization [4, 5] 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 distinguishing 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 most AR applications.

In this paper, we leverage the RFID technology [6, 7] to further label different objects with RFID tags. We deploy additional RFID antennas to the COTS depth camera. To recognize the stationary tagged objects, we propose a rotate 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. We extract the phase value from RF-signal, and pair the tags with the objects according to the correlation between the depth value and phase value. Similarly, to recognize the mobile tagged human subjects, we propose a continuous scanning-based scheme to scan the human subjects, i.e., the system continuously samples the depth of field and RF-signals from these tagged human subjects. In this way, we can accurately identify and distinguish multiple tagged objects, by sufficiently exploring the correlations between the depth of field and the RF-signal.

However, there are several challenges in distinguishing multiple tagged objects in AR systems. The first challenge is conducting accurate paring 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. It is difficult to realize accurate paring due to the large cardinality and mutual interference. The second challenge is mitigating the interferences from the multi-path effect, object occlusion in real settings. These issues lead to nonnegligible interference to pair the tags with the objects, such as the missing tags/objects which fail to be identified as well as extra objects which are untagged. The third challenge is designing an efficient solution without any additional assistance, like the anchor nodes. It is impractical to intentionally deploy anchor nodes in real AR applications due to intensive deployment costs on manpower and time.

This paper presents the first study of using RFID to assist recognizing multiple objects in AR systems (a preliminary version of this work appeared in [8]). Specifically, we make three key contributions : 1) We propose *TaggedAR* to realize the vision "tell me what I see" from AR systems. By sufficiently exploring the correlations between the depth of field and the RF-signal, we propose a rotate scanning-based scheme to distinguish multiple tagged objects in the stationary situation, and propose a continuous scanning-based scheme to distinguish multiple tagged human subjects in the mobile situation. 2) We efficiently tackle the interference from the multi-path effect, object occlusion in real settings, by reducing this problem to a stable marriage problem and propose a stable-matching-based solution to mitigate the interferences from the outliers. 3) We implemented a prototype system and evaluated the performance with case studies in real-world environment. Our solution achieves an average match ratio of 91% in distinguishing up to dozens of RFID tagged objects with a high deployment density.

2 Related Work

Pattern recognition via depth camera: Pattern recognition via depth camera mainly leverages the depth and RGB captured from the camera to recognize objects in a computer vision-based approach. Based on the depth processing [9], a number of technologies are proposed in object recognition [10] and gesture recognition [11, 12]. Nirjon et al. solve the problem of localizing and tracking household objects using depth-camera sensors [13]. The Kinect-based pose estimation method [11] 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.

Batteryless sensing via RFID: RFID has recently been investigated as a new scheme of *batteryless sensing*, including indoor localization [14] , activity sensing [15], physical object search [16], etc. Prior work on RFID-based localization primarily relied on Received Signal Strength [14] or Angle of Arrival [17] 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 [6, 18–20]. 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 [21]. 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 [4].

Combined use in augmented reality environment: Recent works further consider using both depth camera and RFID for indoor localization and object recognition in

1536-1233 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2857812, IEEE Transactions on Mobile Computing

augmented reality environment [22–26]. 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 [23]. Galatas et al. propose a multimodal context-aware localization system, by using RFID and 3D audio-visual information from 2 Kinect sensors deployed at various locations [24]. 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 [25]. Li et al. present a hybrid computer vision and RFID system ID-Match, it uses a novel reverse synthetic aperture technique to recover the relative motion paths of RFID tags worn by people, and correlate that to physical motion paths of individuals as measured with a 3D depth camera [26]. Duan et al. present TagVision, a hybrid RFID and computer vision system for fine-grained localization and tracking of tagged objects [27]. Instead of simply performing indoor localization or object recognition, in this paper, we aim to identify and distinguish multiple tagged objects with depth camera and RFID antennas. Our solution does not require any anchor nodes for assistance, and only leverages at most two RFID antennas for rotate/continuous scanning, which greatly relieves the intensive deployment cost and makes our solution more practical in various scenarios.

3 System Overview

3.1 Design Goals

To realize the vision of "tell me what I see " from the augmented system, we aim to propose an RFID-based approach to use RFID tags to label different objects. Therefore, 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, such that the information stored in the RFID tag can be used to illustrate the specified objects in a detailed approach. Hence, 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) *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.

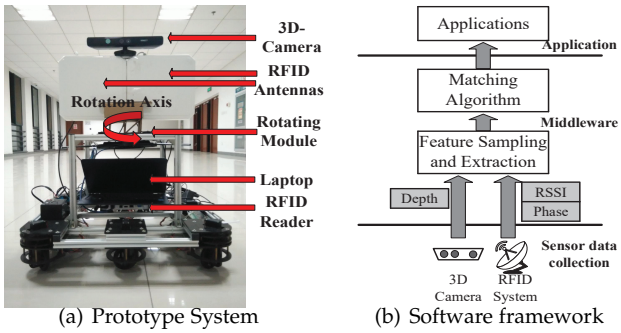
3.2 System Framework

3.2.1 System Prototype

We design a system prototype as shown in Fig. 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 [28], 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. By attaching the RFID tags to the specified objects, to recognize the stationary tagged objects, we propose a rotate 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 the 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. Similarly, to recognize the mobile tagged human subjects, we propose a continuous scanning-based scheme to scan the human subjects, i.e., the system continuously samples the depth of field and RF-signals from these tagged human subjects. By accurately pairing these information, the tags and the objects can be effectively bound together.

3.2.2 Software Framework

The software framework is mainly composed of three layers, i.e., the sensor data collection layer, the middleware layer, and the application layer, as shown in Fig. 2(b). 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. In the following sections, without loss of generality, we evaluate the performance using the Microsoft Kinect for windows, the ImpinJ R420 reader, two Laird S9028 RFID antennas, and multiple Alien 9640 general purpose tags. We attach each tags to one object, and use the Kinect as the depth-camera and use the RFID reader to scan the tags.



(a) Prototype System (b) Software framework

Fig. 2. System Framework

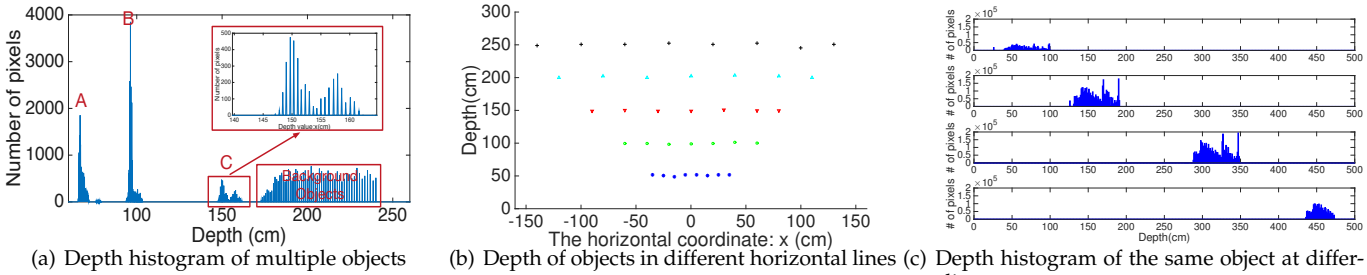
4 Feature Sampling and Extraction

4.1 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. The depth sensor usually consists of an infrared laser projector combined with a monochrome CMOS sensor, which captures the depth.

1536-1233 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2857812, IEEE Transactions on Mobile Computing



(a) Depth histogram of multiple objects (b) Depth of objects in different horizontal lines (c) Depth histogram of the same object at different distances

Fig. 3. Experiment results of depth value

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 at 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.

In order to understand the characteristics of the depth information collected from the depth camera, we conduct real experiments to obtain more observations. We first conduct an experiment to evaluate the characteristics of the depth. Without loss of generality, each experiment observation is summarized from the statistic properties of 100 repeatable observations. 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 Fig. 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 objects 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. 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. If multiple objects are placed with a rather close proximity, it may increase the difficulty to distinguish these objects.

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 Fig. 3(b). Here we find that, for each horizontal line, the depth values of the object keep nearly constant, with rather small deviations; for different horizontal lines, these depth values have obvious variations. Due to the limitation of the Kinect's view, the Kinect has a smaller view angle in a 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.

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. It is found that for an irregularly-shaped object, the depth sensor usually detects multiple peaks with intermittent depths. 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 Δd=10cm 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.

Moreover, in Fig. 3(a), we also find some background noises past the distance of 175 cm, which are produced by background objects, such as the wall and floor. To address the background noise problem, we note that these background noises always lead to a continuous range of depth value, with a very close amount of pixels in the depth histogram. Therefore, we can use a specified pattern to detect and eliminate this range of depth values. Specifically, we respectively set a threshold t_l for the length of the continuous range and a threshold t_p for the number of pixels corresponding to each depth (t_l=50cm and t_p=500 in our implementation). Then, for a certain range of depth value in the depth histogram, if the range is greater than t_l and the number of pixels for each depth value is greater than t_p, we can determine this range as background noise.

The effective scanning distance of the depth camera is very important to the potential range of AR applications, otherwise the potential application scenario should be very limited. In fact, the effective scanning distance of the depth camera, such as Kinect, can be as far as 475cm. To validate that, we perform a set of experiments in regard to the effective scanning distance of the depth camera, e.g., Kinect. We deploy a cardboard of size 20cm×20cm×5cm on the top of a tripod, and evaluate the corresponding depth histogram when the cardboard is separated from the depth camera (i.e., Kinect) with the distance of 50cm, 150cm, 300cm and 450cm, respectively. We plot the experiment results in Fig. 3(c). Note that, when the object is deployed at different

1536-1233 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2857812, IEEE Transactions on Mobile Computing

distances, the profiles of the correspond depth histogram are very similar to each other in most cases. In particular, when the object is deployed at a distance very close to the depth camera, e.g., 50cm, the profile may be distorted to a certain degree. When the object is deployed at a distance of 450cm, the depths over 475cm are no longer illustrated since they are out of the effective scanning distance. Therefore, the experiment results show that the depth camera is able to extract the depth information of the objects at a distance as far as 475cm.

4.2 Extract the Phase Value from RF-Signals

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 [20, 29]:

$$\theta = (\frac{2\pi}{\lambda} \times 2d + \mu) \mod 2\pi, \tag{1}$$

where λ is the wave length. μ is a diversity term which is related with additional phase rotation introduced by the reader's transmitter/receiver and the tag's reflection characteristic. According to the previous study [4], as μ is rather stable, we can record μ for different tags in advance. Then, according to each tag's response, we can calibrate the phase by offsetting the diversity term. Thus, the phase value can be used as an accurate and stable metric to measure distance.

According to the definition in Eq. (1), the phase 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, use them to scan the RFID tags, and respectively obtain their phase values from the RF-signals, as shown in Fig. 4.

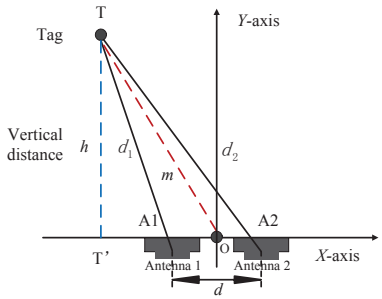


Fig. 4. Compute the (x, y) coordinate of the tag

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* [30], 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 [31], 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 midpoint of the two antennas, the coordinate (x', y') is computed as follows:

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

$$y' = h. \tag{3}$$

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., $\frac{2l}{\lambda}$.

Fig. 5 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 denote a superset of these feasible positions in a straightforward approach.

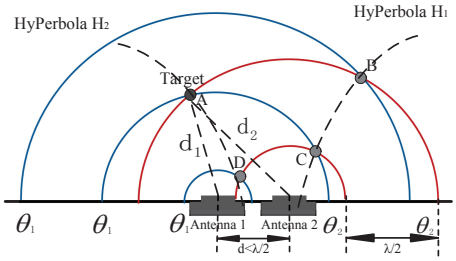


Fig. 5. Estimate the distance from phase values

5 Match the Stationary Tagged Objects via Rotate Scanning

5.1 Motivation

To identify and distinguish the multiple tagged objects, a straightforward solution is to scan the tags in a static

1536-1233 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2857812, IEEE Transactions on Mobile Computing

approach, where both the depth camera and the RFID antenna(s) are deployed in a fixed position without moving. The system scans the objects and tags simultaneously and respectively collects 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 a close vertical distance to the system, this solution cannot effectively distinguish multiple tagged objects in different horizontal distances.

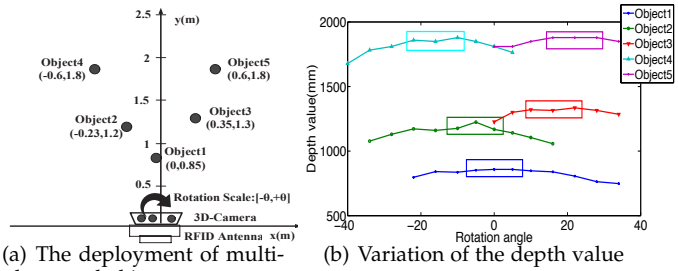
To address this problem, we propose a rotate scanning-based solution as follows: we continuously rotate the scanning system (including the depth camera and RFID antennas), and simultaneously sample the depth of field and RF-signals from multiple tagged objects. Hence, we are able to collect a continuous series of features like depth, RSSI and phase values during rotate 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 a close vertical distance but in different positions.

5.2 Pair the Tags with Objects via Rotate Scanning

5.2.1 Extract Depth via Rotate Scanning

During the rotate 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 Fig. 6(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. Fig. 6(b) shows the experiment results. Note that the series of depth values for each object actually form a convex curve with a peak value. For each depth value obtained at a certain rotation angle, we can use k-NearestNeighbor(kNN) to classify it into a corresponding curve according to the distance between the depth value and the other depth values in the curve, and then use the method of quadratic curve fitting to connect the corresponding depth values as a curve. In this way, we are able to continuously identify and track these depth values for a specified object. The peak value of the convex curve 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 it the *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 rotate 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 the 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. For example, in Fig. 6(a), we deploy the object 4 and object 5 with the same vertical distance to the depth camera, according to the results in Fig. 6(b), these two objects can be distinguished since the peak values of their depth exist in different angles, i.e., -17° and $+22^\circ$ respectively. They can be easily distinguished from the horizontal dimension.



(a) The deployment of multiple tagged objects (b) Variation of the depth value

Fig. 6. The experiment results of rotate scanning

5.2.2 Estimate the tag's position with hyperbolas

According to the analysis shown in Fig. 5, given the two phase values of RF-signals extracted from two antennas separated with a distance d (d=25cm 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 rotate 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 rotate 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: $S = \cap_{i=1}^{m} S_i$.

5.2.3 Estimate the tag's position with angle of arrival

In some situations, it could be difficult to directly derive the tag's candidate position using the intersections of *multiple hyperbolas*, since the hyperbolas must be exactly plotted in the two-dimensional space, which might be computationally expensive for some mobile devices. Nevertheless, it is found that, as long as the tagged objects are relatively far from the antenna pair, we can use the method of *angle of arrival at antenna pair* [21] to simplify the solution. Specifically, suppose the distance between the antenna pair A_1 and A_2 is d, the distances between the tag and the antenna pair A_1/A_2 are respectively d_1 and d_2. As Fig. 7 shows, when the distance between the tag and the antenna pair is significantly larger than the distance between the antenna

1536-1233 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2857812, IEEE Transactions on Mobile Computing

pair, i.e., $d_1 \gg d$ and $d_2 \gg d$, suppose that the angle of arrival of the tagged objects is α, then

$$\Delta d = d_1 - d_2 = d\cos\alpha. \tag{4}$$

Furthermore, when the distance between the antenna pair is less than half of the wavelength, i.e., $d \leq \frac{\lambda}{2}$, we can figure out a pair of symmetric solutions for the angle of arrival of the tagged object. In this regard, we can further use the phase difference between the two antennas to depict the value of Δd, i.e., $\Delta d = |d_1 - d_2| = \Delta\theta = |\theta_1 - \theta_2|$. Therefore, we can figure out the angle of arrival of the tagged objects using the equation:

$$\alpha = \arccos(\frac{\Delta\theta}{d}). \tag{5}$$

As a matter of fact, by leveraging the method of *angle of arrival at antenna pair*, we are able to use the *asymptotic lines of the hyperbolas* to approximate the candidate position of the tagged object, as long as the tagged object is relatively far from the antenna pair.

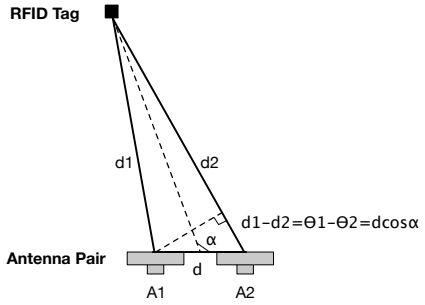


Fig. 7. Angle of arrival at antenna pair

Fig. 8 shows an example of deriving the unique solution of tag's position from the intersections. 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 pair is 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. We use different colors to label the hyperbolas of multiple feasible solutions according to different pairs of phase values. By using the method of *angle of arrival*, we use the *asymptotic lines* to approximate the corresponding hyperbolas. For example, as the distance between A_1 and A_2 is greater than half the wave length, two pairs of symmetric directions of the tagged object are derived, marked with red color; as the distance between A_1 and A_1' is less than half the wave length, one pair of symmetric directions of the tagged object are derived, marked with blue color; similarly, as the distance between A_2 and A_2' is less than half the wave length, one pair of symmetric directions of the tagged object are derived, marked with black color. Moreover, 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.

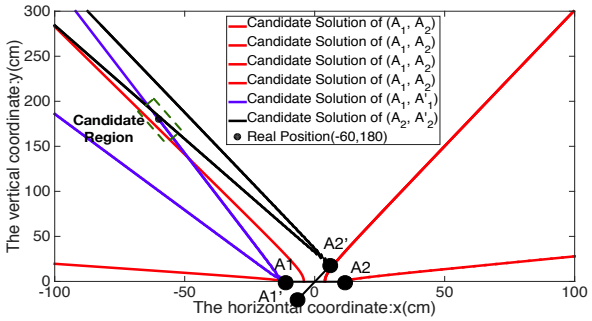


Fig. 8. Figure out the unique solution of the tag's position

5.2.4 *Deriving the angle-distance pair*

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: $\theta = \arctan\frac{x}{y}$, and $\delta = \sqrt{x^2 + y^2}$. Therefore, we can further leverage an algorithm like Algorithm 2 to match multiple tags to multiple objects.

Algorithm 1 Match multiple objects to multiple tags

1: **Extract the vector**: After continuous scanning, identify the peak value from the depth curve and the crossing point of multiple hyperbolas derived from phase pairs. For each object O_i, label it with a vector $\langle\theta_i, d_i\rangle$, respectively normalize the angle and depth into the interval $[0, 1]$ by dividing the maximum value of angle and depth, and add the vector to a set O; For each tag T_j, label it with a vector $\langle\theta_j, \delta_j\rangle$, normalize it and add the vector to a set T.
2: **while** $O \neq \varnothing$ or $T \neq \varnothing$ **do**
3: **Match the objects and tags**: For each object $O_i \in O$ with vector $\langle\theta_i, d_i\rangle$, compute the distance with each tag $T_j \in T$ with vector $\langle\theta_j, \delta_j\rangle$ as follows:

$$\Delta_{i,j} = \sqrt{(\theta_i - \theta_j)^2 + (d_i - \delta_j)^2}.$$

Select the tag T_{j*} with the minimum distance and pair the object O_i 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 minimum distance $\Delta_{i,j}$, 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.

5.3 Tackle the Issues of Interferences

5.3.1 *Impact of Interferences*

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. For example, the multi-path fading may cause the line-of-sight RF-signal and the reflected RF-signals to offset each other at the tag's position, such that

1536-1233 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2857812, IEEE Transactions on Mobile Computing

the tag cannot be effectively activated due to the reduced incident power from the RF-antennas. Besides, the object occlusion may cause one specific object to be blocked by another object placed in front of it, such that this object cannot be effectively identified from its depth histograms. This leads to the issue of *missing tags or objects*. Moreover, in some situations, it is essential to isolate the recognizable object with non-recognizable ones, e.g., a number of tagged objects are placed on an untagged table, and the tagged objects are expected to be recognized instead of the table. However, the table might have effects on the depth-camera reading, but not in RFID-based scanning. This leads to the issue of *extra objects*. The above two issues further lead to imperfect matching between the objects and tags.

5.3.2 Tackle the Outliers in Bipartite Graph Matching

Since we need to find a matching between the set of tags and the set of objects according to their estimated positions, it is similar to finding a matching in a weighted bipartite graph, where the weight refers to the distance between the tag-object pairs. However, due to the existence of the above interferences, they actually form the outliers in addition to the regular points of the tag set and object set. Specifically, these outliers are not essentially far from the regular points in regard to their relative distance, e.g., the extra objects can be fairly close to the regular tagged objects. Therefore, traditional solutions for the *weighted bipartite graph matching* such as the *Hungarian algorithm* [32] cannot effectively tackle the outlier issues in matching, since they seek to find a matching in a weighted bipartite graph which minimizes the overall weight (i.e., the distance between points). They aim to pursue the overall benefits of all members while sacrificing the benefits of individuals. In this regard, to avoid huge value in the overall weight, some specific regular points can be mismatched to the outliers for trade off, then a cascade of mismatches between the regular points could appear frequently.

Hence, in order to tackle the outliers, we reduce this problem to the *stable marriage problem* [33]. Specifically, we aim to find a *stable matching* between the set of tags and the set of objects, given an ordering of preferences for each element. The ordering of preferences can be computed according to the distance between each object-tag pair. We aim to achieve the *stable* property for the matching, i.e., there does not exist any match (A, B) by which both A and B would be individually better off than they are with the element to which they are currently matched. The basic intuition for using the idea of *stable matching* is that, for any tagged object, the distance between the positions of the tag and the object is usually much smaller than the distance from the outliers. So we can give priority to matching the specific pair of the tag and object according to their best preferences in terms of distance. By considering the individual benefits rather than the overall benefits of the object-tag pairs, we can mitigate the impact from the outliers.

We use the *Gale-Shapley algorithm* [33] to solve this problem, as shown in Algorithm 2. It involves a number of iterations. Initially all objects and tags are set to *free*. In the first round, each *free* object *proposes to* the tag it prefers most, and then each tag replies "maybe" to the object it most prefers and gets temporarily *engaged* to the object if it is *free*. In each subsequent round, each *free* object *proposes to* the most-preferred tag to which it has not yet proposed, and each tag replies "maybe" if it is currently *free* or if it prefers this object over its current partner object. This scheme preserves the right of an already-*engaged* tag to *trade up* for better choice. This process is repeated until all objects/tags are *engaged* or have no candidate partner to *propose to*.

Algorithm 2 Stable Matching-based Solution

1: Initialize all $O_i \in O$ and $T_j \in T$ to *free*.
2: Set the weight $w_{i,j}$ as the distance between each pair of object O_i and tag T_j. Compute the ordering of preferences for each object/tag according to $w_{i,j}$. If the weight is greater than a threshold t, remove the corresponding tag/object from the object/tag's preference list.
3: **while** $\exists$ free object o which still has a candidate tag t to *propose to* **do**
4: t=first tag on o's list to which o has not yet *proposed*.
5: **if** t is free **then**
6: (o, t) become *engaged*.
7: **else**
8: some pair (o', t) already exists.
9: **if** t prefers o to o' **then**
10: o' becomes *free*, (o, t) become *engaged*.
11: **else**
12: (o', t) remain *engaged*.
13: **end if**
14: **end if**
15: **end while**

We further illustrate the above idea with an example, as shown in Fig. 9. Fig. 9 shows a scenario where 5 tagged objects are randomly placed in the 2-dimensional space. Due to the impact of interferences, there exist some outliers such as the missing tags/objects and extra interference objects. In this case, the Hungarian Matching (HM)-based solution can mismatch the tag T_2 to the extra objects rather than the object O_2, since it considers the overall benefit to make the tag T_4 to be paired with its only adjacent object O_2. Nevertheless, our Stable Marriage Matching (SMM)-based solution is able to effectively tackle the outliers, by giving priority to matching the tag-object pairs with best preference in distance, e.g., it matches the tag T_2 to the object O_2 rather than the extra object, since O_2 is in higher order of T_2's preference than the extra object, and T_2 is in higher order of O_2's preference than the tag T_4.

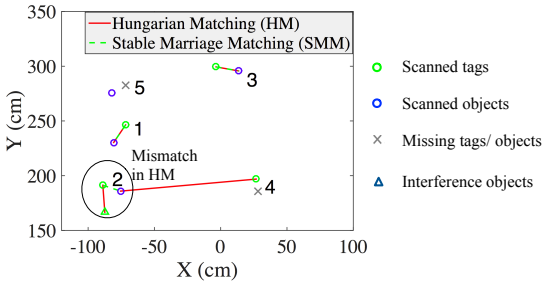


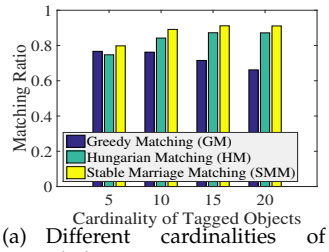
Fig. 9. Tackle the outliers with stable marriage matching

We further compare the performance of different solutions under different settings, i.e., the Greedy Matching in Algorithm 1 (GM), the Hungarian Matching (HM), and the Stable Marriage Matching (SMM), as shown in Fig. 10. By default, the average cardinality and spacing of tagged

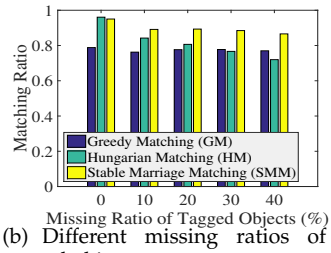
1536-1233 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2857812, IEEE Transactions on Mobile Computing

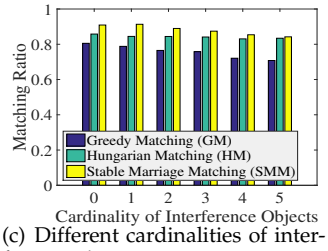
objects are set to 10 and 50cm, respectively, the average missing ratio of tags/objects is set to 10%, the average cardinality and distance of extra interference objects are set to 2 and 50cm, respectively. In Fig. 10(a), we evaluate the match ratios by varying the cardinalities of tagged objects. It is found that SMM always achieves the best performance than the other two solutions. In Fig. 10(b), we evaluate the match ratio by varying the missing ratio of tags or objects for the tagged objects. As the missing ratio increases from 0% to 40%, the matching ratio of HM gradually decreases from 98% to 73%. This implies that HM cannot effectively tackle the outliers due to the missing tags/objects. Nevertheless, SMM always achieves the match ratios greater than 92%, it effectively tackles the outliers of missing tags/objects. In Fig. 10(c) and Fig. 10(d), we evaluate the match ratios by varying the cardinalities of the extra interference objects, and the average distance between the interference objects and tagged objects, respectively. In all situations, SMM achieves the best performance over the other solutions.



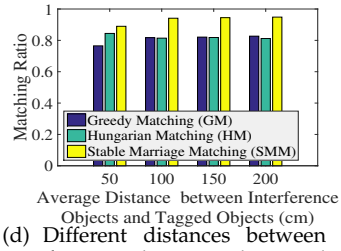
(a) Different cardinalities of tagged objects



(b) Different missing ratios of tagged objects



(c) Different cardinalities of interference objects



(d) Different distances between interference objects and tagged objects

Fig. 10. Performance Evaluation

6 Match the Mobile Tagged Human Subjects via Continuous Scanning

6.1 Motivation

In most cases, the AR systems are designed towards a mobile scenario, e.g., multiple human subjects wearing RFID badges are continuously moving around. For this mobile situation, the rotate scanning-based solution for recognizing multiple stationary tagged objects is no longer suitable. Since the locations of the tagged human subjects are continuously changing, the scanning frequency of the rotate scanning-based solution cannot be high enough to locate the positions of the tags and human subjects in a real-time manner. Nevertheless, we observe that, when multiple tagged human subjects are continuously moving, their moving traces in the two-dimensional space can be distinguishable among each other. Hence, according to the depth information and the phase information extracted from multiple tagged human subjects, we are able to derive some metric to depict the moving traces for the tags and human subjects, respectively. In this way, by matching the moving traces of tags to the corresponding human subjects, we are able to match the mobile tagged human subjects. Therefore, to recognize multiple tagged human subjects in the mobile situation, in this section, we propose a continuous scanning-based solution to pair the mobile tags with moving human subjects via trace matching.

6.2 Pair the Tags with Mobile Human Subjects via Trace Matching

When deploying our system in front of multiple human subjects, where the human subjects wearing RFID badges are moving around, it is known that the state-of-art depth camera such as Kinect is able to extact the skeleton models from the human subjects. Based on the skeleton model, we can further extract the *spinemid* point [1] from the skeleton to represent the human subject, which is also very close to the place of RFID badge worn by the human subject, as shown in Fig.11(a). According to the two-dimensional coordinate of the *spinemid* point in the horizontal plane, we can figure out the moving traces of different human subjects from the depth camera, as shown in Fig.11(b). Moreover, suppose the reader/depth camera is deployed in the origin O, for any *spinemid* point P, we can use the *angle profile* to denote the angle between the vector **OP** and the X-axis **OX**, as shown in Fig.11(b).

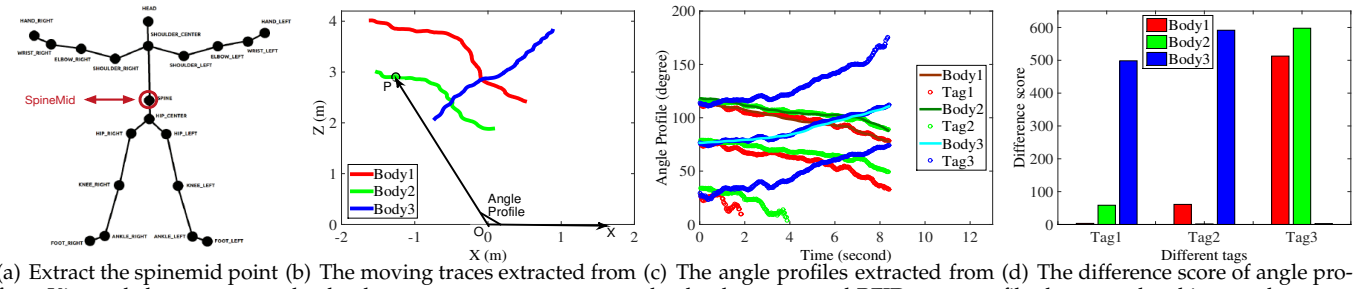
As aforementioned, using the RFID antenna pair, our system can estimate the Angle of Arrival (AoA) of the RFID tag in the horizontal plane. Then, we can similarly use the *angle profile* to denote the angle between the AoA direction of the tag and the X-axis. Recall that according to the phase values collected from the RFID antenna pair, there could be multiple solutions for the angle of arrival of the RFID tag. Hence, there could be multiple *angle profiles* corresponding to the specified tag. Therefore, while the tagged human subjects are moving from time to time, we can plot the angle profiles for both the human subjects and the tags over time. Fig.11(c) shows the corresponding angle profiles for the human subjects and RFID tags over time, where the Tag i is worn on the Body i. Note that for a specified tag, there are multiple solutions for its angle profile, we use the same color to label them. We can observe that the angle profile of the specified body has very close variation trend to one of the angle profiles of the corresponding RFID tag, as they share very similar moving traces in the horizontal plane. Therefore, in order to evaluate the correlation of the angle profile between the bodies and tags, we use the *difference score* to denote this correlation. Specifically, in a specified sliding window W with length L, for the lth snapshot ($1 \leq l \leq L$), suppose the angle profiles of the body O_i and the tag T_j are $\alpha_i(l)$ and $\{\alpha'_j(l)\}$, respectively. Then, the *difference score* $s_{i,j}$ between α_i and α'_j in W is as follows:

$$s_{i,j} = \min_{\alpha'_j \in \{\alpha'_j\}} \frac{1}{L} \sum_{l=1}^{L} \left(\alpha_i(l) - \alpha'_j(l)\right)^2. \quad (6)$$

Here we enumerate all feasible angle profiles α'_j for the tag T_j to compare with the angle profile of α_i for the body O_i, and obtain the minimum value as the *difference score* $s_{i,j}$. Fig.11(d) shows the difference scores in angle profiles between various pairs of tags and bodies, it is found that

1536-1233 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2857812, IEEE Transactions on Mobile Computing



(a) Extract the spinemid point from Kinect skeleton (b) The moving traces extracted from the depth camera (c) The angle profiles extracted from the depth camera and RFID system (d) The difference score of angle profiles between the objects and tags

Fig. 11. An example to illustrate the idea of matching the mobile tagged human subjects via continuous scanning

the least difference score is achieved only for the correct tag-body pair. Based on the above analysis, we further propose Algorithm 3 to pair the tags with mobile human subjects via trace matching.

Algorithm 3 Pair the Tags with Mobile Human Subjects via Trace Matching

1: Perform continuous scanning on the human subjects and the tags, respectively, with the depth camera and RFID antenna pair. Add the human subjects into set O and the tags into set T within a sliding window W.
2: **for** each tag $T_j \in T$ **do**
3: For each snapshot in W, extract the phases of T_j from the antenna pair, and figure out the angle of arrival of T_j. Compute the feasible angle profiles $\{\alpha'_j(l)\}$ corresponding to T_j .
4: **end for**
5: **for** each human subject $O_i \in O$ **do**
6: For each snapshot in W, capture the *spinemid* point from the skeleton of O_i, and calculate the angle profile $\alpha_i(l)$ of O_i.
7: **end for**
8: **while** $O \neq \varnothing$ or $T \neq \varnothing$ **do**
9: **Match the objects and tags**:
10: **for** each human subject $O_i \in O$ **do**
11: **for** each tag $T_j \in T$ **do**
12: Calculate the difference score $s_{i,j}$ between α_i and $\{\alpha'_j\}$.
13: **end for**
14: Select the tag T_{j*} with the minimum difference score and pair the object O_i with the tag T_{j*}.
15: **end for**
16: **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 minimum difference score $s_{i,j}$, 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.
17: **end while**

Note that we use the *angle profiles* to depict the human movement in this paper, whereas the previous works such as *TagVision*[27] and *ID-Match*[26] mainly use the metric *radial distance*, i.e., the Euclidean distance between the reader and the tagged human subject, to depict the human movement. When multiple tagged human subjects are moving around in a large range, e.g., greater than 100cm, which we call *large movement*, the *angle profiles* can depict the human movement in a more sensitive manner than *radial distance*, especially when the human subjects are moving in the scanning range close to the RFID reader/ depth camera, since the *angle profiles* change more rapidly than the *radial distance* when the human subject is performing large movement.

However, when multiple tagged human subjects are close to each other in position, e.g., the distances between adjacent human subjects are less than 20∼30cm, and they only have *slight movements*, e.g., shaking body or turning around, for this situation, our trace-matching-based solution cannot further distinguish these tagged human subjects purely based on the *angle profiles*, since the changes of angle profiles from the tagged human subjects are rather small, which could be less than the inherent errors of the trace-matching-based solution in usual multi-path environment. In this situation of *slight movement*, we can use the *radial distance* to distinguish the multiple tagged human subjects, by referring to the previous solutions [27][26], since the *radial distance* still has some sensitivities to the human movement.

7 Discussion

7.1 Robustness to Environmental Variances

In the real-world environment, besides the environmental interferences such as the multi-path effect, path loss fading, the environmental variances like the material variances and deployment variances could also impact the system performance. E.g., when the RFID tags are deployed to different materials like beverage can, human body or plastic toys, the RF-signal features like RSSI could be totally different. This could greatly impact the performance of the Depth-RSSI pairing-based solution [8]. Nevertheless, the RF-signal features like phase are irrelevant to these factors like different materials, the phase describes the degree that the received signal offsets from the sent signal, which is only correlated to the relative distance and orientation between the antenna and tag. Moreover, the RSSI variation is very sensitive to the orientation change of the tag, whereas the phase variation is relatively insensitive to the orientation change of the tag. In other words, as the tag orientation changes, the RSSI might be changing sharply, whereas the phase is changing relatively gently. Thus, based on the stability of the phase, our Depth-Phase pairing-based solution is able to effectively address the variability of these environmental factors.

Hence, we further evaluate the RSSI and phase values with different orientations of the tag and different materials of the tagged objects. First, we continuously rotate the tag to measure the RSSI and phase values with different tag orientations. Fig.12 shows the diagram of tag rotation. We rotate the tag on two different axes, i.e., the X-axis and Y-axis. While the tag is rotating on the Y-axis, we use

1536-1233 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2857812, IEEE Transactions on Mobile Computing

α to define the angle between the antenna plane and the tag plane, and continuously change the value of α from $-90°$ to $90°$. We find that both the RSSI and the phase keep stable, as shown in Fig.13(a) and Fig.13(b). It implies that they are insensitive to the tag orientation change on the Y-axis. While the tag is rotating on the X-axis, we fix the value of α to $0°, 30°, 60°, 70°, 80°$ and $90°$, respectively, and rotate the tag on the X-axis. We find that the phase is changing relatively gently, whereas the RSSI is changing very sharply in most situations, as shown in Fig.13(c) and Fig.13(d). It implies that the phase is relatively insensitive to the tag orientation change on the X-axis, whereas the RSSI is very sensitive to the tag orientation change on the X-axis. We further evaluate the RSSI and phase values by attaching the RFID tag to different materials like the carton, plastic bottle (with/without water) and metal can (with/without water). As shown in Fig.13(e) and Fig.13(f), while we vary the distance between the tag and the reader from 50cm to 300cm, we find that the RSSI is rather sensitive to the attached materials, whereas the phase is relatively insensitive to the attached materials. Therefore, the phase is a more stable metric than RSSI in regard to the robustness to environmental variances.

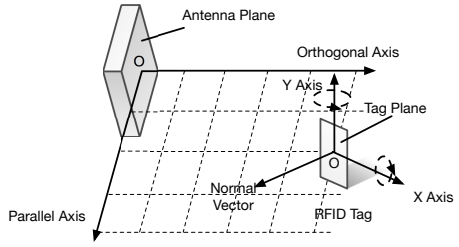
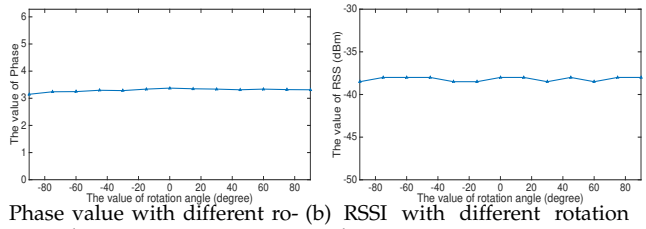
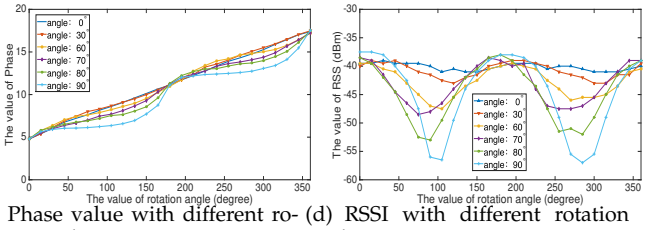


Fig. 12. The diagram of tag rotation



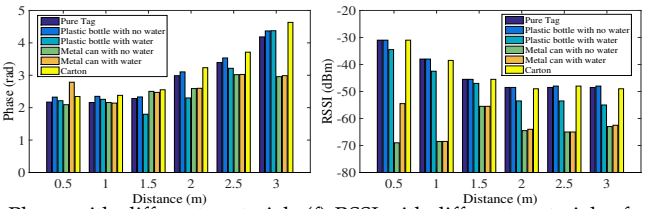
(a) Phase value with different rotation angles on Y-axis

(b) RSSI with different rotation angles on Y-axis



(c) Phase value with different rotation angles on X-axis

(d) RSSI with different rotation angles on X-axis



(e) Phase with different materials of the tagged objects

(f) RSSI with different materials of the tagged objects

Fig. 13. RSSI and phase values with different orientations of the tag and different materials of the tagged objects

7.2 Robustness to Severe Tag Missing

In real application environment, due to the environmental factors like the multi-path effect and energy absorption of different materials, the miss reading of RFID tags can be frequent. As a matter of fact, in the real-world environment, tag missing seems to be unavoidable based on current RFID design. As aforementioned, it may further lead to mismatch in the pairing process of the system. In the situation of severe tag missing, the actual performance can be greatly degraded, even if the stable matching-based solution is applied. In order to tackle this issue, we can mitigate the negative effect of severe tag missing by deploying multiple tags on the objects. Specifically, we can attach a tag array of multiple tags onto the surface of the specified object. The tags in the tag array can be deployed in a one-dimensional manner or two-dimensional manner, where the tags are separated with a certain distance, e.g., 3∼10cm. As the tags are placed at different positions and different orientations on the object's surface, the environmental factors can be different, this sufficiently reduces the possibility of miss reading of all tags simultaneously. As long as the RF-signals of at least one tag can be resolved from one object, the pairing process can be executed. Therefore, the robustness to severe tag missing can be guaranteed by deploying multiple tags on the objects.

8 Performance Evaluation

8.1 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 deployed multiple objects in an area of about 3.5m$\times$ 3.5m, and attach each tag to an object. We used the Kinect as the depth-camera and use the RFID reader to scan the tags.

8.2 Evaluate the Performance in Stationary Situation

We implemented four schemes for performance comparison (Readers can refer to a preliminary version of this work in [8] for the detailed solution):

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.

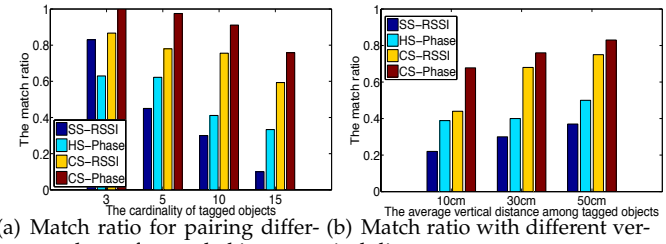
4) *Continuous Scanning via Depth-Phase Pairing (CS-Phase)*: The system continuously scans the tagged objects while it is rotating, and pairs the tags with objects according to the extracted series of depth and phase.

Without loss of generality, by default we deployed 10 tagged objects in the scanning area. We varied the settings of the average horizontal/vertical distance, and the cardinality

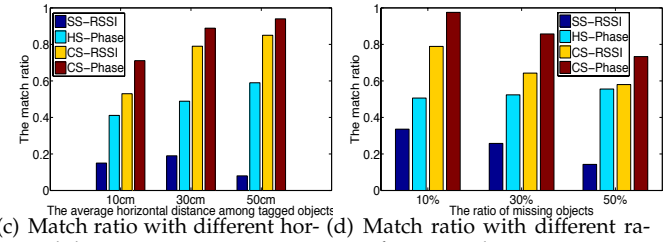
1536-1233 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2857812, IEEE Transactions on Mobile Computing

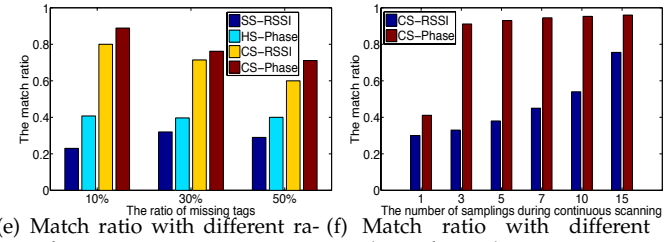
of tagged objects. For each setting, we randomly generated 10 types of deployments for the tagged objects, and evaluated the average match ratio for successful pairing in the above four schemes.



(a) Match ratio for pairing different numbers of tagged objects (b) Match ratio with different vertical distances



(c) Match ratio with different horizontal distances (d) Match ratio with different ratios of missing objects



(e) Match ratio with different ratios of missing tags (f) Match ratio with different numbers of samplings

Fig. 14. The experiment results in stationary situation

8.2.1 Evaluate the Accuracy

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 evaluated the match ratio for pairing different cardinalities of tagged objects, by varying the cardinality of tagged objects from 3 to 15. As shown in Fig. 14(a), 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 varied the average vertical distances and horizontal distances among the tagged objects, thus to further evaluate the performance in accuracy. We fixed the average horizontal (vertical) distance among the objects to 30 cm, and varied the average vertical (horizontal) distance from 10 cm to 50 cm. Fig. 14(b) and Fig. 14(c) 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.

8.2.2 Evaluate the Robustness

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 ran experiments to evaluate the robustness to missing tags/objects, when there exist several objects or tags which fail to be identified. Here we measured the match ratio for the remaining objects or tags. Fig. 14(d) and Fig. 14(e) 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 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 rotate scanning*. Fig. 14(f) 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.

8.3 Evaluate the Performance in Mobile Situation

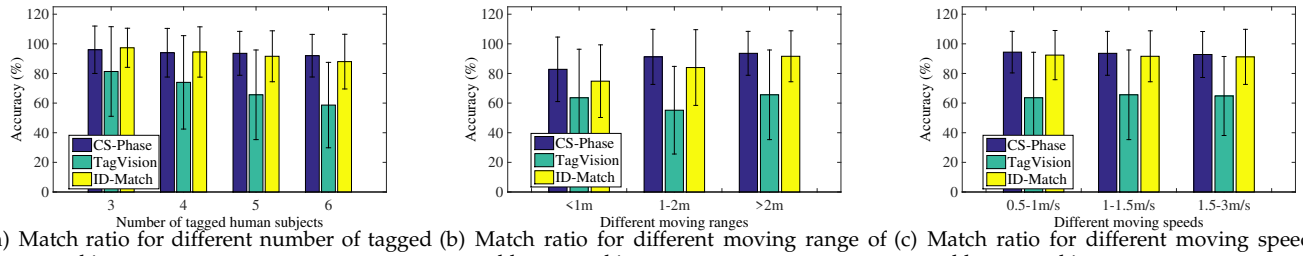
We implemented three schemes of recognizing the tagged moving objects for performance comparison, i.e., *ID-Match* [26], *TagVision* [27], and our solution *CS-Phase*. Without loss of generality, by default we let 5 tagged human subjects move around in the scanning area. We varied the cardinality, moving range, and moving speed of the tagged human subjects. For each setting, we randomly generated 50 different moving traces for the tagged human subjects, and evaluated the average match ratio for successful pairing.

Accuracy for different cardinality of tagged human subjects *Our solution achieves good performance in accuracy when the number of tagged human subjects is varied from 3 to 6*. As shown in Fig. 15(a), as the cardinality increases from 3 to 6, the match ratios of *CS-Phase* and *ID-Match* decrease slightly, whereas the match ratios of *TagVision* decrease rapidly. The reason is probably that *TagVision* is originally designed to track the moving objects like the toy trains rather than the human subjects, it is not robust to tackle the heavy multi-path effect from human subjects. Nevertheless, *CS-Phase* achieves a match ratio greater than 92% in all cases.

Accuracy for different moving range of tagged human subjects *Our solution achieves good performance in accuracy when the average moving range of tagged human subjects is*

1536-1233 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2857812, IEEE Transactions on Mobile Computing



(a) Match ratio for different number of tagged human subjects (b) Match ratio for different moving range of tagged human subjects (c) Match ratio for different moving speed of tagged human subjects

Fig. 15. The experiment results in mobile situation

varied from 0.5 to 3m. As shown in Fig. 15(b), as the moving range increases from 0.5 to 3m, the match ratios of *CS-Phase* and *ID-Match* increase slightly, whereas the match ratios of *TagVision* fluctuate around 65%. The increase of the match ratio is mainly because the larger moving ranges help to further distinguish different moving traces of the tags and human subjects. Nevertheless, *CS-Phase* achieves a match ratio greater than 83% in all cases.

Accuracy for different moving speed of tagged human subjects *Our solution achieves good performance in accuracy when the moving speed of tagged human subjects is varied from 0.5m/s to 3m/s*. As shown in Fig. 15(c), as the moving speed increases from 0.5m/s to 3m/s, i.e., low speed (0.5-1m/s), medium speed (1-1.5m/s), and high speed (1.5-3m/s), the match ratios of all three solutions mainly keep unchanged. The reason is mainly because the sampling rate of both the RFID reader and the depth camera is large enough to accurately capture and distinguish the moving trace of the tags and the human subjects, whatever the moving speed is. Nevertheless, *CS-Phase* achieves a match ratio greater than 92% in all cases.

9 Case Study: Recognize Tagged Human Subjects in the Cafe

To further evaluate the real performance of our system by considering more practical issues (e.g., indoor multi-path and energy absorption), we did more thorough experiments in a more realistic setting. We deployed our system in a typical application scenario, i.e., recognizing multiple tagged human subjects in the cafe, as shown in Figure 16. In the case study, a major task of our system is to effectively identify and distinguish these real-world objects and further show their inherent information in the camera's view. Thus, we implemented an application which was executed on a SAMSUNG PC equipped with an Intel(R) Core(TM) I5 1.4GHz CPU and 4G RAM. The PC was remotely connected to the system via WiFi. We deployed our system in front of these human subjects with the distance from 1.5m to 4.5m.

9.1 Stationary Situation

Experiment Settings: As shown in Fig. 16, 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 height, horizontal distance and vertical distance. 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

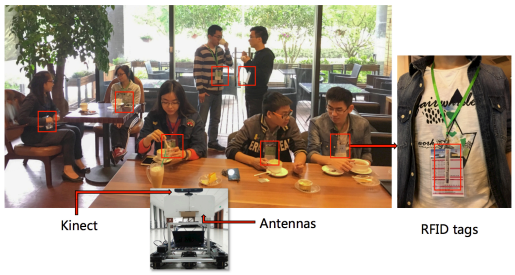


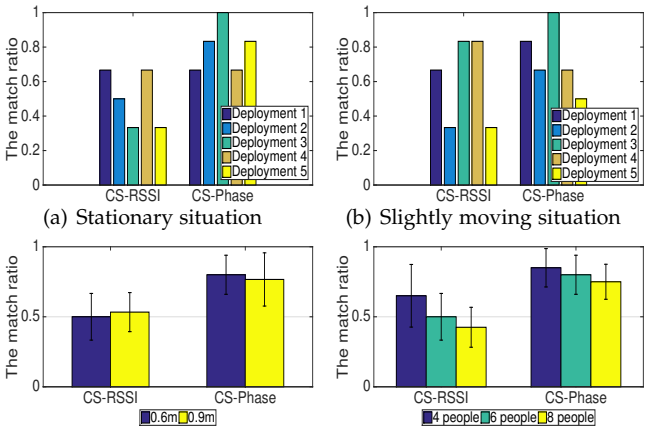
Fig. 16. Example deployment of multiple human subjects wearing RFID badges in the cafe

the performance of match ratios, by varying the factors like the number of human subjects, the spacing between human subjects, and the moving state. The default number of human subjects and the default average spacing are 6 and 60 cm, respectively.

Performance Evaluation: *In the stationary situation, our solution achieves good matching accuracy to recognize multiple tagged human subjects of different factors like the height, spacing, etc.* Fig. 17(a)-(d) respectively show 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 Fig. 17(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 ($\leq$40cm/s) or angle ($\leq$ 30°/s). As shown in Fig. 17(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 Fig. 17(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 Fig. 17(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 as the phase in RF-signals is irrelevant to the energy absorption problems.

1536-1233 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2857812, IEEE Transactions on Mobile Computing



(a) Stationary situation (b) Slightly moving situation

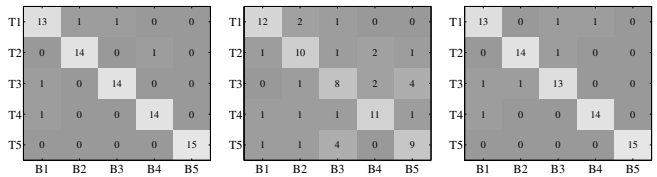
(c) Different average spacing (d) Different number of people

Fig. 17. Performance evaluation in stationary situation

9.2 Mobile Situation

Experiment Settings: We let 5 human subjects walking around in the cafe, wearing the RFID tagged badges. For each experiment setting, we randomly generated 15 sets of traces for these human subjects. For each moving trace, these human subjects can be moving around with different speeds and different ranges. The dynamic movement of human bodies leads to more dynamics and uncertainties in the multi-path environment. We conducted experiments to evaluate the performance of match ratios by investigating the confusion matrix.

Performance Evaluation: *In the mobile situation, our solution achieves good matching accuracy to recognize multiple tagged human subjects of different factors like the moving speed, moving range, etc.* Fig. 18(a)-(c) respectively show the example confusion matrix for pairing the tag $(T_1, T_2, \cdots, T_5)$ with the human subject $(B_1, B_2, \cdots, B_5)$ with different solutions, i.e., our solution *CS-Phase*, *TagVision*[27], *ID-Match*[26]. According to the confusion matrix, it is found that both *CS-Phase* and *ID-Match* are able to accurately pair the tag with the human subject for most of the 15 sets of traces, *CS-Phase* achieves an average accuracy of 93.3%, whereas *ID-Match* achieves an average accuracy of 92%. *TagVision* achieves the worst performance with the average accuracy of 66.7%. The reason is probably that *TagVision* is originally designed to track the moving objects like the toy trains rather than the human subjects, it is not robust to tackle the heavy multi-path effect from human subjects.



(a) CS-Phase (b) Tag-Vision (c) ID-Match

Fig. 18. Confusion matrix in mobile situation

10 Conclusion

In this paper, we propose *TaggedAR*, i.e., an RFID-based system to recognize multiple RFID tagged objects in AR system, by deploying additional RFID antennas to the COTS depth camera. By sufficiently exploring the correlations between the depth of field and the RF-signal, we propose a rotate scanning-based scheme to distinguish multiple tagged objects in the stationary situation, and propose a continuous scanning-based scheme to distinguish multiple tagged human subjects in the mobile situation. The experiment results show that we achieve an average accuracy of 91% in distinguishing up to dozens of tagged objects.

Acknowledgments

This work is supported in part by National Natural Science Foundation of China under Grant Nos. 61472185, 61321491, 61702257; JiangSu Natural Science Foundation under Grant No. BK20151390, BK20170648. This work is partially supported by Collaborative Innovation Center of Novel Software Technology and Industrialization. This work is partially supported by the program A for Outstanding PhD candidate of Nanjing University. The work of Jie Wu was supported in part by NSF grants CNS 1449860, CNS 1461932, CNS 1460971, CNS 1439672, CNS 1301774, CNS 1629746, CNS 1564128, and ECCS 1231461.

References

[1] "Kinect," 2016, http://www.microsoft.com/en-us/kinectforwindows.
[2] L. Xie, Q. Li, X. Chen, S. Lu, and D. Chen, "Continuous scanning with mobile reader in rfid systems: an experimental study," in *Proc. of ACM MobiHoc*, 2013.
[3] J. Liu, M. Chen, B. Xiao, F. Zhu, S. Chen, and L. Chen, "Efficient RFID grouping protocols," *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 3177–3190, 2016.
[4] L. Yang, Y. Chen, X. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile rfid tags to high precision using COTS devices," in *Proc. of ACM MobiCom*, 2014.
[5] J. Wang and D. Katabi, "Dude, where's my card?: Rfid positioning that works with multipath and non-line of sight," in *Proc. of the ACM SIGCOMM*, 2013.
[6] L. Yang, Q. Lin, X. Y. Li, T. Liu, and Y. Liu, "See through walls with cots rfid system," in *Proc. of MobiCom*, 2015.
[7] J. Liu, B. Xiao, X. Liu, and L. Chen, "Fast RFID polling protocols," in *Proc. of ICPP*, 2016, pp. 304–313.
[8] L. Xie, J. Sun, Q. Cai, C. Wang, J. Wu, and S. Lu, "Tell me what i see: Recognize rfid tagged objects in augmented reality systems," in *Proc. of ACM UbiComp*, 2016.
[9] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, and A. Davison, "Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera," in *Proc. of ACM UIST*, 2011.
[10] Z. Ren, J. Yuan, and Z. Zhang, "Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera," in *Proc. of ACM Multimedia*, 2011.
[11] S. Obdrzalek, G. Kurillo, F. Ofli, R. Bajcsy, E. Seto, H. Jimison, and M. Pavel, "Accuracy and robustness of kinect pose estimation in the context of coaching of elderly population," in *Proc. of IEEE EMBC*, 2012.
[12] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: A review," *IEEE Transactions on Cybernetics*, vol. 43, no. 5, pp. 1318–1334, 2013.
[13] S. Nirjon and J. Stankovic, "Kinsight: Localizing and tracking household objects using depth-camera sensors," in *Proc. of IEEE DCOSS*, 2012.
[14] L. Shangguan, Z. Li, Z. Yang, M. Li, and Y. Liu, "Otrack: Order tracking for luggage in mobile RFID systems," in *Proc. of IEEE INFOCOM*, 2013.
[15] M. Buettner, R. Prasad, M. Philipose, and D. Wetherall, "Recognizing daily activities with rfid-based sensors," in *Proc. of UbiComp*, 2009, pp. 51–60.
[16] J. Nickels, P. Knierim, B. Könings, F. Schaub, B. Wiedersheim, S. Musiol, and M. Weber, "Find my stuff: Supporting physical objects search with relative positioning," in *Proc. of UbiComp*, 2013.
[17] S. Azzouzi, M. Cremer, U. Dettmar, R. Kronberger, and T. Knie, "New measurement results for the localization of UHF RFID transponders using an Angle of Arrival (AoA) approach," in *Proc. of IEEE RFID*, 2011.

1536-1233 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2857812, IEEE Transactions on Mobile Computing

[18] J. Wang, F. Adib, R. Knepper, D. Katabi, and D. Rus, "Rf-compass: Robot object manipulation using RFIDs," in *Proc. of ACM MobiCom*, 2013.
[19] T. Liu, L. Yang, Q. Lin, Y. Guo, and Y. Liu, "Anchor-free backscatter positioning for RFID tags with high accuracy," in *Proc. of IEEE INFOCOM*, 2014.
[20] L. Shangguan, Z. Yang, A. X. Liu, Z. Zhou, and Y. Liu, "Relative localization of RFID tags using spatial-temporal phase profiling," in *Proc. of NSDI*, 2015.
[21] J. Wang, D. Vasisht, and D. Katabi, "RF-IDraw: Virtual touch screen in the air using RF signals," in *Proc. of ACM SIGCOMM*, 2014.
[22] C. S. Wang, C. L. Chen, and Y. M. Guo, "Real-time indoor positioning system based on RFID and kinect," in *Proc. of Information Technology Convergence*, 2013.
[23] F. Klompmaker, H. Fischer, and H. Jung, "Authenticated tangible interaction using RFID and depth-sensing cameras," in *Proc. of ACHI*, 2012.
[24] G. Galatas and F. Makedon, "A system for multimodal context-awareness," *International Journal of Advanced Computer Science and Applications*, 2013.
[25] C. Cerrada, S. Salamanca, E. Perez, J. A. Cerrada, and I. Abad, "Fusion of 3D vision techniques and RFID technology for object recognition in complex scenes," in *IEEE International Symposium on Intelligent Signal Processing*, 2007.
[26] H. Li, P. Zhang, S. Al Moubayed, S. N. Patel, and A. P. Sample, "Id-match: A hybrid computer vision and rfid system for recognizing individuals in groups," in *Proc. of the CHI Conference on Human Factors in Computing Systems (CHI '16)*, 2016, pp. 4933–4944.
[27] C. Duan, X. Rao, L. Yang, and Y. Liu, "Fusing RFID and computer vision for fine-grained object tracking," in *Proc. of IEEE INFOCOM*, 2017.
[28] "ImpinJ," 2016, http://www.impinj.com.
[29] H. Ding, L. Shangguan, Z. Yang, J. Han, Z. Zhou, P. Yang, W. Xi, and J. Zhao, "Femo: A platform for free-weight exercise monitoring with RFIDs," in *Proc. of ACM SenSys*, 2015.
[30] K. Kendig, "Is a 2000-year-old formula still keeping some secrets?" *Amer. Math. Monthly*, vol. 107, pp. 402–415, 2000.
[31] C. Godfrey and A. W. Siddons, "Apollonius' theorem," *Modern Geometry*, p. 20, 1908.
[32] H. W. Kuhn, "Variants of the hungarian method for assignment problems," *Naval Research Logistics Quarterly*, vol. 3, pp. 253–258, 1956.
[33] D. Gale and L. S. Shapley, "College admissions and the stability of marriage," *American Mathematical Monthly*, vol. 69, pp. 9–14, 1962.

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 60 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, IEEE INFOCOM, IEEE ICNP, etc.

Chuyu Wang received his B.S. degree in Software Engineering from Dalian University of Technology, China in 2012. He is currently a Ph.D. candidate in the Department of Computer Science and Technology at Nanjing University. His research interests include RFID Systems and Indoor Localization.

Yanling Bu received her B.S. degree in Geographic Information System from Nanjing Normal University, China in 2015, and received her M.S. degree in Department of Computer Science and Technology from Nanjing University, China in 2018. She is currently a Ph.D. candidate in the Department of Computer Science and Technology at Nanjing University. Her research interests include RFID Systems and Localization.

Jianqiang Sun received his B.S. degree in information security from Nanjing Normal University, China in 2011 and received his M.S. degree in the Department of Computer Science and Technology at Nanjing University in 2016. His research interests include RFID Systems, and Internet of Things.

Qingliang Cai received his B.S. degree in information security from Suzhou University, China in 2012 and received his M.S. degree in the Department of Computer Science and Technology at Nanjing University in 2017. His research interests include RFID Systems, and Internet of Things.

Jie Wu is the chair and a Laura H. Carnell professor in the Department of Computer and Information Sciences at Temple University. He is also an Intellectual Ventures endowed visiting chair professor at the National Laboratory for Information Science and Technology, Tsinghua University. His current research interests include mobile computing and wireless networks, routing protocols, cloud and green computing, network trust and security, and social network applications. Dr. Wu regularly publishes in scholarly journals, conference proceedings, and books. He serves on several editorial boards, including IEEE Transactions on Service Computing and the Journal of Parallel and Distributed Computing. He was an IEEE Computer Society Distinguished Visitor, ACM Distinguished Speaker, and chair for the IEEE Technical Committee on Distributed Processing (TCDP). Dr. Wu is a CCF Distinguished Speaker and a Fellow of the IEEE. He is the recipient of the 2011 China Computer Federation (CCF) Overseas Outstanding Achievement Award.

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.

1536-1233 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.