RF-MVO: Simultaneous 3D Object Localization and Camera Trajectory Recovery Using RFID Devices and a 2D Monocular Camera

Zhongqin Wang[1,2], Min Xu[2], Ning Ye[3], Ruchuan Wang[3], Haiping Huang[3]

[1]School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing, China
[2]School of Electrical and Data Engineering, University of Technology Sydney, Sydney, Australia
[3]School of Computer, Nanjing University of Posts and Telecommunications, Nanjing, China
Email: zhongqin.wang@student.uts.edu.au, min.xu@uts.edu.au, {yening,wangrc,hhp}@njupt.edu.cn

***Abstract*—Most of the existing RFID-based localization systems cannot well locate RFID-tagged objects in a 3D space. Limited robot-based RFID solutions require reader antennas to be carried by a robot moving along an already-known trajectory at a constant speed. As the first attempt, this paper presents RF-MVO, which fuses battery-free RFID and monocular visual odometry to locate stationary RFID tags in a 3D space and recover an unknown trajectory of reader antennas binding with a 2D monocular camera. The proposed hybrid system exhibits three unique features. Firstly, since the trajectory of a 2D monocular camera can only be recovered up to an unknown scale factor, RF-MVO combines the relative-scale camera trajectory with depth-enabled RF phase to estimate an absolute scale factor and spatially incident angles of an RFID tag. Secondly, we propose a joint optimization algorithm consisting of coarse-to-fine angular refinement, 3D tag localization and parameter nonlinear optimization, to improve real-time performance. Thirdly, RF-MVO can determine the effect of relative tag-antenna geometry on the estimation precision, providing optimal tag positions and absolute scale factors. Our experiments show that RF-MVO can achieve 6.23cm tag localization accuracy in a 3D space and 0.0158 absolute scale factor estimation accuracy for camera trajectory recovery.**

***Keywords*—RFID; Monocular Visual Odometry; Tag Localization**

I. Introduction

Radio Frequency IDentification (RFID) is a promising technique that uses backscatter signals to communicate with RFID tags for automatic object identification. Low-cost battery-free RFID tags make it easy to manage and track each RFID-tagged object, which is becoming the first choice of many industry solutions, especially in indoor object localization [1]. However, many state-of-the-art studies [2]–[4] for RFID localization can only work in a two-dimensional (2D) plane and require RFID tags or RFID reader antennas to move along a known trajectory. In this work, we mainly focus on how to locate stationary RFID tags in a 3D space when reader antennas move along an unpredictable trajectory. To achieve this goal, we introduce a 2D monocular camera into RFID-based systems. Many applications benefit from the RFID and Computer Vision (CV) fusion solutions. For example, a RFID-CV fusion robot can construct and update a map of an unknown environment using simultaneous localization and mapping (SLAM) technique [5]–[7] and further present the spatial positions of RFID-tagged objects in the map through RFID localization.

At present, two main problems challenge RFID localization accuracy. *1) Phase Wrapping*. In commercial RFID readers, the reported RF phase with 2π radians period repeats at distances between a reader antenna and an RFID tag separated by integer multiples of the one-half carrier wavelength. *2) Multipath Interference*. The reflected beams off surrounding objects of RFID tags can combine with primary backscatter signals at the tag end, thereby changing the reported phase. To address the above-mentioned problems, existing purely RFID-based methods for stationary object localization can be categorized into stationary-antenna [2], [4], [8] and moving-antenna [9], [10]: *Stationary-antenna*. The stationary-antenna approaches need to deploy more than three reader antennas at specified positions to remove the uncertainty of multiple tag positions caused by phase periodicity. However, the impact of multipath interference can not be effectively suppressed and these solutions fail to locate RFID-tagged objects in a 3D space. *Moving-antenna*. The moving-antenna approaches require to deploy reader antennas on a robot moving along a linear trajectory at a constant speed, because the trajectory can be easily and accurately estimated. In dynamic environments, however, the mobile robot might rely on its self-adaptive trajectory planning algorithm to avoid obstacles, which will inevitably produce an unpredictable trajectory. To our knowledge, capturing this trajectory for most of commonly-used low-cost robots (e.g., about 200 US dollars iRobot Create 2 [11]) is challenging.

Visual odometry (VO) enables the estimation of the locations and orientations of a camera by analyzing a sequence of captured images [12], [13]. Stereo and RGB-D VO can recover an actual camera trajectory, while monocular VO can only estimate the trajectory up to an unknown scale factor. However, in the case where the distance from a stereo camera to a working scenario is much higher than the distance between two camera lens, stereo VO will degenerate to the monocular

case. Similarly, a depth-enabled camera in most of commercial off-the-shelf RGB-D cameras (e.g. Kinect V2) also has a ranging limit of about 0∼3m measurement distance. Since a commonly 8dBi-gain antenna-equipped RFID system with the reading range of about 6∼10m is usually used in wide working areas like warehouse, it is not always reliable to obtain an exact trajectory using a stereo or RGB-D camera. Fortunately, RF phase is a tag-to-antenna distance function, which provides us with an opportunity to recover a 2D monocular camera trajectory up to an absolute scale factor. In general use, we mainly focus on monocular VO for fusion localization.

In this paper, we are the first to introduce RF-MVO, an RFID and CV hybrid system that can simultaneously locate stationary RFID-tagged objects in a 3D space and recover an unknown trajectory of reader antennas binding with a 2D monocular camera. An RFID reader, one or more reader antennas and a 2D monocular camera are deployed on a mobile utility cart. We move the cart in the region of interest to locate RFID tags using reported RF phase and a piece of camera trajectory up to an unknown scale factor. To achieve this hybrid system, we need to solve three key challenges:

Challenge 1: Determining a stationary RFID tag position in a 3D space given an already-known antenna trajectory like the previous work [14] or estimating camera/antenna displacements of a current 2D image relative to its last 2D image given an already-known RFID tag position is relatively easy. However, it is very challenging to simultaneously estimate the camera/antenna trajectory and 3D RFID tag position. Further, RF phase measurements are affected by multipath interference, and camera pose estimation suffers from accumulated errors over time. To deal with these problems, we utilize the mobility of reader antennas to emulate a sequence of overlapped antenna arrays. Then according to Direction of Arrival (DOA) estimation theory, we design a spatial power spectrum with strong tolerance to measurement noise, which can be used to calculate a pair of azimuth and elevation angles for tag localization as well as an absolute scale factor for trajectory recovery in each antenna array.

Challenge 2: To obtain the accurate incident angles and absolute scale factors, the smaller the searching granularities in the proposed spatial power spectrum, the higher the system computation jeopardizes the real-time performance. Instead, RF-MVO firstly uses a relatively small factor granularity and a high angular granularity to obtain the high-resolution absolute scale factor and low-resolution incident angles. On this basis, we propose a joint optimization problem to accelerate our task for tag localization and camera/antenna trajectory recovery.

Challenge 3: RF-MVO might output multiple tag positions of an RFID tag over antenna arrays. And in an antenna array, RF-MVO might simultaneously read more than one RFID tag, producing multiple absolute scale factors. Selecting an optimal tag position and absolute scale factor still remains challenging. To address this problem, a key intuition is that the tag localization error is sensitive to antenna-tag geometry. Antenna elements in an antenna array that are close to each other can not provide with good geometry as the antennas that are widely separated. Inspired by Global Positioning System (GPS), RF-MVO introduces horizontal dilution of precision (HDOP) to evaluate the effect of antenna-tag geometry on estimation accuracy.

To the best of our knowledge, RF-MVO is the first RFID and CV fusion system, which utilizes RF phase measurements and 2D images to simultaneously perform 3D RFID-tagged object localization and 2D monocular camera trajectory recovery. The main contributions are summarized as follows:

1) We propose a DOA-based spatial power spectrum with the suppression capability of multipath interference and accumulated errors to search incident angles (i.e., azimuth and elevation angles) and an absolute scale factor in each antenna array, which is a fundamental step for 3D tag localization and camera/antenna trajectory recovery.

2) To balance between estimation accuracy and real-time performance, we propose a joint optimization algorithm that iterates the steps of coarse-to-fine angular search, 3D tag localization and parameter nonlinear optimization. In our experiment, the algorithm can perform very quickly and achieve fine-grained estimation accuracy.

3) We exploit horizontal dilution of precision to determine the effect of tag-antenna geometry on estimation accuracy, which effectively helps our hybrid system obtain the optimal 3D tag position and absolute scale factor over each antenna array.

We build a prototype of RF-MVO using off-the-shelf RFID devices and a 2D monocular camera. Experimental results demonstrate that, when only deploying two reader antennas on the mobile utility cart, RF-MVO can locate stationary RFID tags with the average of 6.23 cm localization error in a 3D space and estimate absolute scale factors with the average of 0.0158 estimation error.

Paper outline: The rest of this paper is organized as follows. Section II describes the overview of our hybrid system. The main algorithms are described in III, IV and V, respectively. Experimental setup and results are introduced in Section VI and VII. Related work is reviewed in Section VIII. Finally, we conclude the work.

II. System Overview

The proposed RF-MVO can locate RFID-tagged objects in a 3D space and estimate absolute scale factors for camera/antenna trajectory recovery without additionally deploying any other sensor (e.g. wheel odometer) or measuring a reference object with a pre-known size. An RFID reader together with one or more directional antennas binding with a 2D monocular camera are carried by a utility cart moving along an unknown trajectory in the region of interest. RFID tags are affixed on stationary objects to be located in advance. The RFID reader collects RFID tag data, including Electronic Product Code (EPC), RF phase and reading timestamp. The 2D monocular camera captures a sequence of 2D images. Fig. 1 shows our RF-MVO system architecture, which contains four components: sampling synchronization, spatial power spectrum search, joint optimization and accuracy estimation.

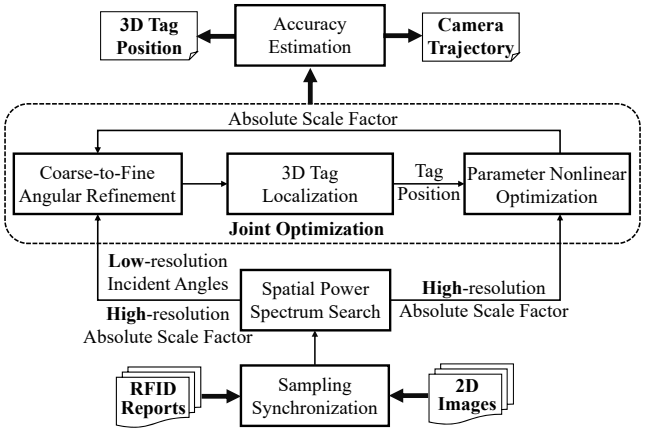


Fig. 1. RF-MVO system architecture.

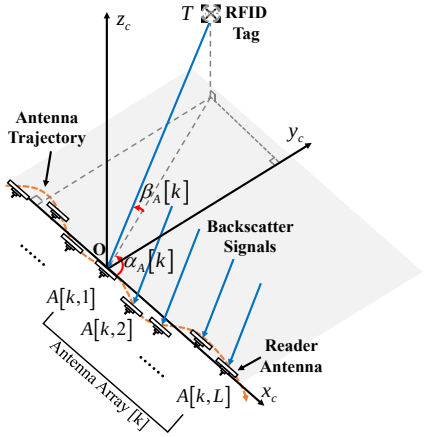


Fig. 2. DOA estimation model.

1) Sampling Synchronization. According to Gen2 standard [15], most of RFID readers rely on a slotted-aloha access scheme to read RFID tags randomly. The sampling time between consecutive inventories of the same tag is unpredictable, which is determined by reader settings (reader mode, search mode and session), tag population and environment interference. In our experiment, a 2D camera is set to capture video streams at 30 frames per second (FPS), while an RFID reader can read the same RFID tag more than 10 times per second. We synchronize the RFID reader clock with an Internet time server, and then match RFID reports to 2D images by minimizing sampling time difference [16].

2) Spatial Power Spectrum Search. In Section III, we introduce how to build an antenna array and a DOA-based spatial power spectrum. A pair of low-resolution incident angles of an RFID tag in a 3D space and a high-resolution absolute scale factor are estimated by finding the highest peak in the spectrum.

3) Joint Optimization. In Section IV, we give details on the joint optimization solution. Firstly, the step of coarse-to-fine angular refinement is to refine the low-resolution incident angles given the high-resolution absolute scale factor. Secondly, the step of 3D tag localization is to locate the tag position given the refined incident angles and the absolute scale factor. Thirdly, the step of parameter nonlinear optimization is to simultaneously refine the 3D tag position and the absolute scale factor because they are interrelated with each other. The algorithm takes iterations before it converges.

4) Accuracy Estimation. In Section V, we describe how to calculate Horizontal dilution of precision to evaluate the effect of tag-antenna geometry on estimation accuracy and select the optimal 3D tag position and absolute scale factor.

III. Searching Spatial Power Spectrum and Locating RFID Tags

In this section, a spatial power spectrum is built to search a pair of incident angles and an absolute scale factor in each antenna array. Then we locate RFID tags in a 3D space with coarse localization accuracy.

A. Estimation of Incident Angles and Absolute Scale Factor

Direction of Arrival Estimation Model. In Fig. 2, a geometric model in a 3D scenario is to intuitively introduce our basic idea. As a reader antenna moves over time, it reads a stationary RFID tag at different positions, which is regarded to deploy multiple antennas at each of these positions. An antenna array of a reader antenna A contains the length of L phase measurements, and the step size between two adjacent antenna arrays is one phase sample. Suppose that in the k-th antenna array, phase values are reported at the antenna positions $X_A[k,1], ..., X_A[k,L]$, respectively. The azimuth and elevation angles of RFID backscatter signals arriving at the k-th antenna array are denoted as $\alpha_A[k]$ and $\beta_A[k]$. The displacement of two adjacent reader antennas in x-axis, y-axis and z-axis is $\Delta \boldsymbol{X}_A[k,i] = (\Delta x_A[k,i], \Delta y_A[k,i], \Delta z_A[k,i])$, which is the same as the 2D monocular camera displacement at each time. Here we assume that $\Delta \boldsymbol{X}_A[k,i] = \boldsymbol{X}_A[k,i+1] - \boldsymbol{X}_A[k,i]$. In this case, the tag-to-antenna distance difference between two consecutive i-th and $(i+1)$-th elements in the k-th antenna array can be approximated as follows:

$$\begin{aligned}\Delta h_A[k,i] =& \Delta x_A[k,i] \times cos\alpha_A[k]\, cos\beta_A[k] + \\ & \Delta y_A[k,i] \times sin\alpha_A[k]\, cos\beta_A[k] + \\ & \Delta z_A[k,i] \times sin\beta_A[k]\end{aligned} \tag{1}$$

In our system, a 2D monocular camera poses (i.e., positions and orientations) are estimated by analyzing a sequence of 2D images based on monocular visual odometry (MVO) technique [17]. Without additional information, however, the estimated camera positions of the current view relative to the previous view can be only recovered up to an unknown scale factor. We exploit Perspective-n-point (PnP) algorithm [18] for camera pose estimation to make the scale factors be same over an antenna array. In this way, the estimated displacement between adjacent camera positions up to an relative-scale factor is $\Delta \boldsymbol{X}_C[k,i] = \boldsymbol{X}_C[k,i+1] - \boldsymbol{X}_C[k,i]$ in the k-th antenna array. Suppose that the ground-truth scale factor is denoted as $\gamma[k]$, called absolute scale factor in our work, then $\Delta \boldsymbol{X}_A[k,i] = \gamma[k] \times \Delta \boldsymbol{X}_C[k,i]$. Note that we also use windowed bundle adjustment [19] in MVO to further optimize

the camera poses corresponding to the last L image views. In this case, however, the absolute scale factors in different antenna arrays will be different from each other.

The Range of Absolute Scale Factor. If the speed at which the camera captures photos is set to 30 FPS and $\|\Delta \boldsymbol{X}_A[k,i]\| < \lambda/2$ (λ is the signal wavelength, about 32 cm), the maximum speed of the utility cart can be as high as 4.8 m/s, which could be meet the requirement for most of application scenarios. However, previous work [20] indicates that when an RFID tag moves at a high speed, most of RFID systems would miss a lot reported packets from the tag and suffer from serious doppler frequency shift. In practice, our experience suggests to set the maximum speed to about 100 cm/s. Suppose that the Euclidean distance from the reader antenna at $\boldsymbol{X}_A[k,i]$ to an RFID tag is denoted by $d_A[k,i]$. According to the relationship between the distance and phase [21] as well as the triangle rule that the length difference of two sides is smaller than the length of the third side, the absolute value of the distance difference between $d_A[k,i]$ and $d_A[k,i+1]$ meets the following equation:

$$\begin{aligned}|\Delta d_A[k,i]| &= |d_A[k,i+1] - d_A[k,i]| \\ &= \frac{\lambda}{4\pi}\left|\Delta\varphi_A[k,i] + \Delta N_A[k,i]\right| < \|\Delta \boldsymbol{X}_A[k,i]\|\end{aligned} \tag{2}$$

where

$$\begin{cases} \|\Delta \boldsymbol{X}_A[k,i]\| = \gamma[k] \times \|\Delta X_C[k,i]\| \\ \Delta\varphi_A[k,i] = \varphi_A[k,i+1] - \varphi_A[k,i] \\ \Delta N_A[k,i] = N_A[k,i+1] - N_A[k,i] \end{cases}$$

and the phase φ_A is reported by an RFID reader, ranging within $[0, 2\pi]$. The unknown parameter N_A, called phase ambiguity, is an integral multiple of 2π to make φ_A fall within $[0, 2\pi]$.

To determine the range of an absolute scale factor, we firstly assume all of the antenna displacements is within $[0, \lambda/4]$. Due to $\Delta\varphi_A[k,i] \in [-2\pi, 2\pi]$, the difference of the phase ambiguity $\Delta N_A[k,i]$ can be determined as follows:

$$\Delta N_A[k,i] = \begin{cases} 2\pi, & -2\pi \le \Delta\varphi_A[k,i] < -\pi \\ 0, & |\Delta\varphi_A[k,i]| \le \pi \\ -2\pi, & \pi < \Delta\varphi_A[k,i] \le 2\pi \end{cases} \tag{3}$$

The minimum absolute scale factor is calculated by

$$\gamma_{min}[k] = \max_{i\in[1,L-1]}\left\{\frac{\Delta d_A[k,i]}{\|\Delta \boldsymbol{X}_C[k,i]\|}\right\} \tag{4}$$

And due to $\|\Delta X_A[k,i]\| \le 100/10$ (the maximum speed of the cart is 100 cm/s and the tag is read 10 times/s at least in the experiment), the maximum absolute scale factor is

$$\gamma_{max}[k] = \min_{i\in[1,L-1]}\left\{\frac{10}{\|\Delta \boldsymbol{X}_C[k,i]\|}\right\} \tag{5}$$

Spatial Power Spectrum. Given an absolute scale factor $\gamma[k] \in (\gamma_{min}[k], \gamma_{max}[k])$ in the k-th antenna array, the spatial power spectrum of backscatter signals along an azimuth angle $\alpha_A \in [0°, 180°]$ and an elevation angle $\beta_A \in [0°, 180°]$ is a 2D intensity graph in which each pixel represents the likelihood of each pair of spatially incident angles:

$$\boldsymbol{P}_k(\alpha_A, \beta_A, \gamma) = \frac{1}{L-1}\sum_{i=1}^{L-1}\cos\Delta\vartheta_A[k,i] \tag{6}$$

where $\Delta\vartheta_A[k,i] = \Delta\varphi_A[k,i] + \frac{4\pi}{\lambda}\Delta h_A[k,i]$. The closer the parameters $\alpha_A[k]$, $\beta_A[k]$ and $\gamma[k]$ to the ground truth, the larger the peak of the proposed spatial power spectrum is.

Analysis: RFID-related hardware such as RFID tags circuit, RFID reader antennas feed cable and RFID readers components may all introduce additional phase shift φ_h. And RF signals transmitted along multiple paths (reflected by the ceiling, wall or even human bodies) in the actual circumstances may also produce additional phase shift φ_e. In practice, the reported phase can be further formulated as

$$\varphi_A = \left(\frac{4\pi d_A}{\lambda} + \varphi_h + \varphi_e\right) \mod 2\pi \tag{7}$$

where the function mod represents the modulo operation, which can be removed by the cosine function. For consecutive phase $\varphi_A[k,i]$ and $\varphi_A[k,i+1]$, additional phase shifts φ_h and φ_e will be very close to each other. Hence, consecutive phase differencing can effectively reduce the effect of unexpected phase shifts. In addition, as the estimates of camera poses can not exactly match the ground truth despite using windowed bundle adjustment, the pose errors inevitably accumulate over time, which can be suppressed by differencing the consecutive estimates of camera/antenna positions.

Search Absolute Scale Factor and Incident Angles. Given M physical reader antennas on the utility cart, we obtain the optimal absolute scale factor and incident angles (i.e., azimuth and elevation angles) in the k-th antenna array by searching the highest peak in the proposed spatial power spectrum:

$$\begin{cases} \boldsymbol{S}_k(\alpha_{A_i}, \beta_{A_i}, \gamma) = \max\limits_{\alpha_{A_i},\beta_{A_i}\in[0,180°]} \boldsymbol{P}_k(\alpha_{A_i}, \beta_{A_i}, \gamma) \\ \gamma^*[k] = \arg\max\limits_{\gamma[k]\in(\gamma_{min}[k],\gamma_{max}[k])} \dfrac{1}{M}\sum\limits_{i=1}^{M}\boldsymbol{S}_k(\alpha_{A_i}, \beta_{A_i}, \gamma) \\ \left(\alpha^*_{A_i}[k], \beta^*_{A_i}[k]\right) = \arg\max\limits_{\alpha_{A_i},\beta_{A_i}\in[0,180°]} \boldsymbol{P}_k(\alpha_{A_i}, \beta_{A_i}, \gamma^*) \end{cases} \tag{8}$$

B. RFID Tag Localization in a 3D Space

Average Incident Angles. The proposed system takes an incremental process to generate an antenna array. Once receiving the length of L RFID-CV data, RF-MVO produces an absolute scale factor and a pair of azimuth and elevation angles. Hence, the length of $2(L-1)$ data should be provided for our hybrid system initialization at the beginning. In Fig. 3, since an antenna element may be included in multiple antenna arrays, we firstly take an average of the estimated incident angles in these arrays. For the k-th reader antenna, the averaged azimuth and elevation angles $\overline{\alpha}_A[k]$ and $\overline{\beta}_A[k]$ are calculated as follows:

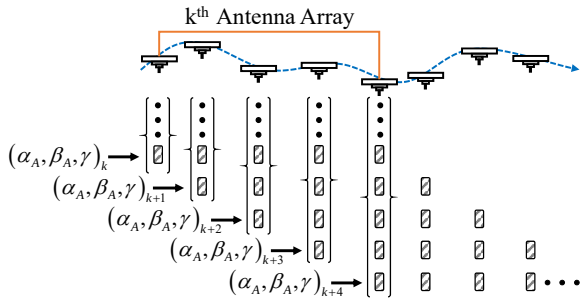


Fig. 3. Incident angle averaging.

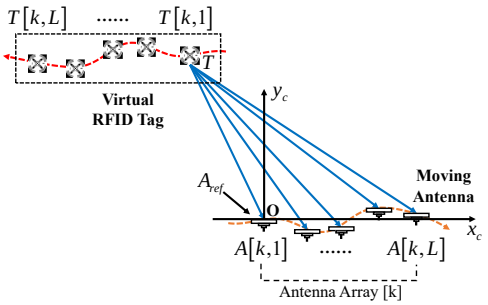


Fig. 4. RFID tag localization.

1) If $1 \leq k < L$, we have

$$\begin{cases} \overline{\alpha}_A[k] = \dfrac{1}{k}\displaystyle\sum_{i=1}^{k} \alpha_A[i] \\ \overline{\beta}_A[k] = \dfrac{1}{k}\displaystyle\sum_{i=1}^{k} \beta_A[i] \end{cases} \tag{9}$$

2) If $k \geq L$, we have

$$\begin{cases} \overline{\alpha}_A[k] = \dfrac{1}{L}\displaystyle\sum_{i=k-L+1}^{k} \alpha_A[i] \\ \overline{\beta}_A[k] = \dfrac{1}{L}\displaystyle\sum_{i=k-L+1}^{k} \beta_A[i] \end{cases} \tag{10}$$

Locate RFID Tag in a 3D Space. To reduce the effect of accumulated errors in camera trajectory estimation, we pinpoint the relative location of an RFID tag in each antenna array. Instead of viewing that reader antennas binding with a 2D monocular camera move along an unknown trajectory, the fixed RFID tag T can be considered to virtually move in the opposite direction, as illustrated in Fig. 4. Suppose that the first element of a physical reader antenna A_{ref} in the k-th antenna array is at the coordinate origin O, then the i-th ($i \geq 2$) element position is denoted as

$$\boldsymbol{X}_{A_{ref}}[k,i] = \gamma[k] \times \sum_{j=1}^{i-1} (\boldsymbol{X}_C[k,j+1] - \boldsymbol{X}_C[k,j]) \tag{11}$$

Since the displacement $\Delta\boldsymbol{X}_{A_{ref}}$ of another antenna A relative to A_{ref} can be measured in advance, the corresponding position of A is

$$\boldsymbol{X}_A[k,i] = \boldsymbol{X}_{A_{ref}}[k,i] + \Delta\boldsymbol{X}_{A_{ref}} \tag{12}$$

Since a spatial line can be represented by the intersection of two planes in a 3D space, the equation of a line $\ell_A[k,i]$ passing through the RFID tag and the i-th antenna element at 3D coordinates $\boldsymbol{X}_T[k]$ and $\boldsymbol{X}_A[k,i]$ are given by

$$\frac{x_T[k]-x_A[k,i]}{u_A[k,i]} = \frac{y_T[k]-y_A[k,i]}{v_A[k,i]} = \frac{z_T[k]-z_A[k,i]}{w_A[k,i]} \tag{13}$$

where

$$\begin{cases} \boldsymbol{X}_T[k] = (x_T[k], y_T[k], z_T[k]) \\ \boldsymbol{X}_A[k,i] = (x_A[k,i], y_A[k,i], z_A[k,i]) \end{cases} \tag{14}$$

and the direction vector $(u_A[k,i], v_A[k,i], w_A[k,i])$ along the line $\ell_A[k,i]$ is calculated as follows:

$$\begin{cases} u_A[k,i] = \cos\overline{\alpha}_A[k,i]\cos\overline{\beta}_A[k,i] \\ v_A[k,i] = \sin\overline{\alpha}_A[k,i]\cos\overline{\beta}_A[k,i] \\ w_A[k,i] = sin\overline{\beta}_A[k,i] \end{cases} \tag{15}$$

Here $\overline{\alpha}_A[k,i] = \overline{\alpha}_A[k+i-1]$, $\overline{\beta}_A[k,i] = \overline{\beta}_A[k+i-1]$.

Given a total of $M \times L$ lines, there might exist multiple non-intersecting lines and even parallel lines in a 3D space. We regard the nearest point to these lines as the spatial coordinate $\boldsymbol{X}_T[k]$ of the RFID tag T, which is solved using Singular Value Decomposition (SVD) technique [22].

IV. Jointly Optimizing Tag Position and Absolute Scale Factor

In the above discussion, exhaustively searching incident angles from $0°$ to $180°$ and an absolute scale factor from γ_{min} to γ_{max} is a time-consuming process. The time complexity is $O\left(\frac{180}{\Delta\alpha} \times \frac{180}{\Delta\beta} \times \frac{\gamma_{max}-\gamma_{min}}{\Delta\gamma}\right)$, where $\Delta\alpha$, $\Delta\beta$ and $\Delta\gamma$ represent the corresponding searching granularities. In this section, a joint optimization algorithm is proposed to balance the estimation accuracy and real-time performance:

1) **Initialization**. Instead of searching incident angles and an absolute scale factor with fine granularities simultaneously, we firstly obtain a high-resolution absolute scale factor $\gamma^0[k]$ and low-resolution incident angles $\alpha_A^0[k]$ and $\beta_A^0[k]$ given a small value of $\Delta\gamma$ and relatively large values of $\Delta\alpha$ and $\Delta\beta$, which can ensure low computational load.

2) **Refine Incident Angles $\alpha_A[k]$ and $\beta_A[k]$**. A coarse-to-fine angular refinement is proposed to refine the incident angles using small searching spacings of $\Delta\alpha_{opt}$ and $\Delta\beta_{opt}$. The angular searching ranges are $\alpha_A^l[k] \pm \mu \in [0°, 180°]$ and $\beta_A^l[k] \pm \mu \in [0°, 180°]$, where μ is the predefined searching threshold.

In the l-th iteration, the azimuth and elevation angles $\alpha_A^{l+1}[k]$ and $\beta_A^{l+1}[k]$ of the antenna A are updated given $\gamma^l[k]$:

$$(\alpha_A^{l+1}[k], \beta_A^{l+1}[k]) = \underset{\substack{\alpha_A \in [\alpha_A^l[k]-\mu, \alpha_A^l[k]+\mu] \\ \beta_A \in [\beta_A^l[k]-\mu, \beta_A^l[k]+\mu]}}{\arg\max} \boldsymbol{P}_k(\alpha_A^l, \beta_A^l, \gamma^l) \tag{16}$$

Then we update the averaged incident angles according to Eq.(9) and (10), so we have $\overline{\alpha}_A^{l+1}[k]$ and $\overline{\beta}_A^{l+1}[k]$.

3) **Pinpoint Tag Position $\boldsymbol{X}_T[k]$.** In the l-th iteration, the 3D tag position $\boldsymbol{X}_T^l[k]$ is calculated given $\gamma^l[k]$ and a set of averaged incident angles $\left\{\overline{\alpha}_{A_j}^{l+1}[k], \overline{\beta}_{A_j}^{l+1}[k]\right\}(j=1,...,M)$ corresponding to M reader antennas (refer to Section III.B):

$$\boldsymbol{X}_T^l[k] = \underset{\substack{i\in[1,L]\\ j\in[1,M]}}{SVD}\left\{\ell\left(\overline{\alpha}_{A_j}^{l+1}[k,i], \overline{\beta}_{A_j}^{l+1}[k,i], \gamma^l[k]\right)\right\} \quad (17)$$

4) **Optimize Tag Position $\boldsymbol{X}_T[k]$ and Absolute Scale Factor $\gamma[k]$.** In the l-th iteration, the tag position $\boldsymbol{X}_T^{l+1}[k]$ and the absolute scale factor $\gamma^{l+1}[k]$ are both updated by minimizing the distance error given $\boldsymbol{X}_T{}^l[k]$ and $\gamma^l[k]$ as follows:

$$\left(\boldsymbol{X}_T^{l+1}[k], \gamma^{l+1}[k]\right) = \underset{\boldsymbol{X}_T^l[k],\gamma^l[k]}{\arg\min} \sum_{i=1}^{L-1}\sum_{j=1}^{M}\left\|\Delta d_{A_j}^l[k,i] - \Delta d_{C_j}^l[k,i]\right\| \quad (18)$$

where

$$\begin{cases} \Delta d_{A_j}^l[k,i] = \dfrac{\lambda}{4\pi}\left(\Delta\varphi_{A_j}[k,i] + \Delta N_{A_j}^l[k,i]\right) \\ \Delta d_{C_j}^l[k,i] = \left\|\boldsymbol{X}_{A_j}^l[k,i+1] - \boldsymbol{X}_T^l[k]\right\| - \left\|\boldsymbol{X}_{A_j}^l[k,i] - \boldsymbol{X}_T^l[k]\right\| \\ \boldsymbol{X}_{A_j}^l[k,i] = \Delta\boldsymbol{X}_{A_{ref}}[j] + \gamma^l[k]\times(\boldsymbol{X}_C[k,i] - \boldsymbol{X}_C[k,1]) \end{cases} \quad (19)$$

Since the spacing of consecutive antenna positions can be estimated by

$$\begin{aligned}\Delta\boldsymbol{X}_A^l[k,i] &= \boldsymbol{X}_A^l[k,i+1] - \boldsymbol{X}_A^l[k,i] \\ &= \gamma^l[k]\times(\boldsymbol{X}_C[k,i+1] - \boldsymbol{X}_C[k,i])\end{aligned} \quad (20)$$

and $\left\|\Delta\boldsymbol{X}_A^l[k,i]\right\| \le \frac{\lambda}{2}$, the phase ambiguity difference $\Delta N_A^l[k,i]$ is determined as:

a) If $\left\|\Delta\overline{\boldsymbol{X}}_A^l[k,i]\right\| \in \left[0, \frac{\lambda}{4}\right]$, we have

$$\Delta N_A^l[k,i] = \begin{cases} 0, & \left|\Delta\varphi_A^l[k,i]\right| \le \pi \\ -2\pi, & \pi < \Delta\varphi_A^l[k,i] \le 2\pi \\ 2\pi, & -2\pi \le \Delta\varphi_A^l[k,i] < -\pi \end{cases} \quad (21)$$

b) If $\left\|\Delta\overline{\boldsymbol{X}}_A^l[k,i]\right\| \in \left(\frac{\lambda}{4}, \frac{\lambda}{2}\right]$, we have

$$\Delta N_A^l[k,i] = \begin{cases} 0, & \pi \le \left|\Delta\varphi_A^l[k,i]\right| \le 2\pi \\ -2\pi, & 0 < \Delta\varphi_A^l[k,i] < \pi \\ 2\pi, & -\pi < \Delta\varphi_A^l[k,i] < 0 \end{cases} \quad (22)$$

The process for the tag position and absolute scale factor refinement is interrelated with each other, so they need to be optimized simultaneously. The nonlinear optimization problem is solved by Levenberg-Marquardt algorithm [23]. Once the estimates of tag position and absolute scale factor are updated, we repeat the joint optimization from step 2 until it converges.

Analysis: For a non-linear optimization problem, a good initial guess closer to the ground truth can achieve faster convergence, so the tag position refinement can speed up the optimization process. Given $\Delta\alpha_{opt} = 1°$, $\Delta\beta_{opt} = 1°$ and $\mu = 20°$ in our experiment, the number of updating the absolute scale factor and tag position is 4 times on average, taking the average of 16 milliseconds in our platform illustrated in Section VI. In addition, the proposed algorithm just provides a local optimum in each antenna array and can not achieve a global optimum over all antenna arrays. Due to adjacent antenna arrays with one sample step, our algorithm performs very well in practice shown in Section VII.

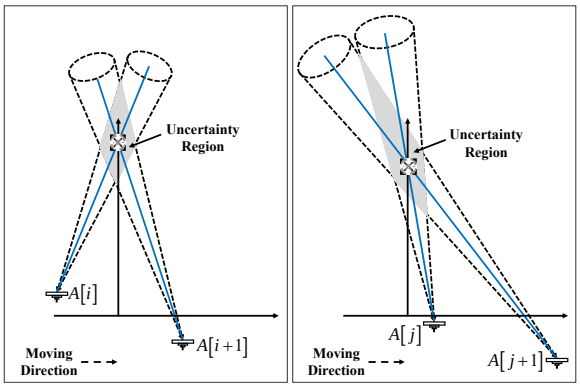


Fig. 5. Uncertainty of RFID Tag Localization.

V. Estimating Localization Accuracy And Recovering 2D Camera Trajectory

As the cart moves near a target RFID tag, RF-MVO produces a series of antenna arrays and then outputs the corresponding tag position estimates. In Fig. 5, the shaded region in each case illustrates the size of uncertainty region. Ignoring the measurement error, the smaller the uncertainty region, the higher the localization accuracy is. The RFID tag positions are less precisely located as the lines through the antenna coordinates become more parallel, which means that in an antenna array the average distance from the RFID tag to each antenna element becomes larger and/or the average spacing between consecutive antenna elements becomes smaller. In [3], the authors also indicate that reader antennas deployed close to each other and fay away from RFID tags cannot provide stable and accurate localization results. Since our cart moves along the horizontal coordinates (X-Y dimension), we mainly focus on horizontal position errors. Inspired by global positioning system, we introduce Horizontal Dilution of Precision (HDOP) to evaluate RFID tag localization accuracy.

According to the definition of HDOP, we linearize the distance function $d_A[k,i]$ from the ground-truth position $\boldsymbol{X}_{GT}[k]$ of the RFID tag T to the i-th element in the k-th antenna array by expanding a Taylor series at the optimized tag coordinate $\boldsymbol{X}_T^*[k]$, and then ignore second and higher order terms,

$$\begin{aligned} d_A[k,i] \approx &\frac{\partial d_A[k,i]}{\partial x}\Delta x_T[k] + \frac{\partial d_A[k,i]}{\partial y}\Delta y_T[k] + \\ &\frac{\partial d_A[k,i]}{\partial z}\Delta z_T[k] + \widetilde{d}_A[k,i] \end{aligned} \quad (23)$$

where $\frac{\partial d_A[k,i]}{\partial X}$ represents the first partial derivative of each distance function $d_A[k,i]$, $\widetilde{d}_A[k,i] = \|\boldsymbol{X}_A[k,i] - \boldsymbol{X}_T^*[k]\|$ and $\boldsymbol{X}_{GT}[k] = \boldsymbol{X}_T^*[k] + \Delta\boldsymbol{X}_T[k]$. $\Delta\boldsymbol{X}_T[k]$ represents the tag position's measurement error.

Consider M reader antennas on the cart, we group $M \times L$ equations in the k-th antenna array together and represent them in matrix form,

$$\boldsymbol{D}[k] = \boldsymbol{G}[k] \Delta \boldsymbol{X}_T[k] + \widetilde{\boldsymbol{D}}[k] \tag{24}$$

where $\boldsymbol{G}[k]$ is a matrix of the partial derivatives in X-Y dimension with $M \times L$ rows and 2 columns,

$$\boldsymbol{G}[k] = \left[\begin{array}{ccc} \boldsymbol{G}_{A_1}[k] & \cdots & \boldsymbol{G}_{A_M}[k] \end{array} \right]^T \tag{25}$$

and the matrix corresponding to the j-th reader antenna can be represented by

$$\boldsymbol{G}_{A_j}[k] = \left[\begin{array}{cc} \frac{\partial d_{A_j}[k,1]}{\partial x} & \frac{\partial d_{A_j}[k,1]}{\partial y} \\ \vdots & \vdots \\ \frac{\partial d_{A_j}[k,L]}{\partial x} & \frac{\partial d_{A_j}[k,L]}{\partial y} \end{array} \right] \tag{26}$$

We construct the covariance matrix $Q[k]$ for localization error analysis,

$$\boldsymbol{Q}[k] = \left[(\boldsymbol{G}[k])^T \boldsymbol{G}[k] \right]^{-1} = \left[\begin{array}{cc} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{array} \right] \tag{27}$$

Since σ_x^2 and σ_y^2 are the variances of X-axis and Y-axis components of the tag position estimate, HDOP is given by

$$HDOP = \sqrt{\sigma_x^2 + \sigma_y^2} \tag{28}$$

Low HDOP values represent a better tag positional accuracy due to strong tag-antenna geometry. In this case, when the errors from camera pose estimation and RF phase measurement are at the same level in each antenna array, the lower the value of HDOP, the higher the positioning accuracy will be. Additionally, there might be many RFID tags to be located, so our system will output multiple candidates of the absolute scale factor in an antenna array. We can find out the optimal absolute scale factor $\gamma_T^*[k]$ corresponding to the tag T with the minimum HDOP value. Suppose that the initial position of the 2D camera is $\widehat{\boldsymbol{X}}_C[1] = \boldsymbol{X}_C[1] = 0$ at the beginning, then the estimated camera trajectory recovered up to absolute scale factors can be expressed as

$$\widehat{\boldsymbol{X}}_C[k+1] = \widehat{\boldsymbol{X}}_C[k] + \gamma_T^*[k] \times (\boldsymbol{X}_C[k+1] - \boldsymbol{X}_C[k]) \tag{29}$$

VI. Implementation

Hardware: We employ an Impinj R420 RFID reader without any hardware or firmware modification. The reader works in the Australia operating frequency band of 920∼926 MHz with 500 kHz channel spacing. The reader is directly connected to a laptop via a standard Ethernet cable. Two 8dBi circular polarization antennas with about 6∼10m reading range connect to the RFID reader. Impinj H47 battery-free RFID tags with the size of 4.4cm×4.4cm are attached on double-sided checkerboards. Besides, a Microsoft Kinect V2 is used to capture 2D images and estimate the ground truth of camera trajectory for performance test. The frame rate is configured as 30 FPS. Note that we only input 2D images into the proposed system. These devices are all deployed on a utility cart.

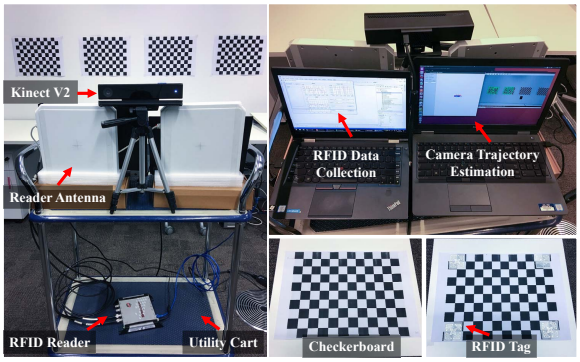


Fig. 6. Experiment setup.

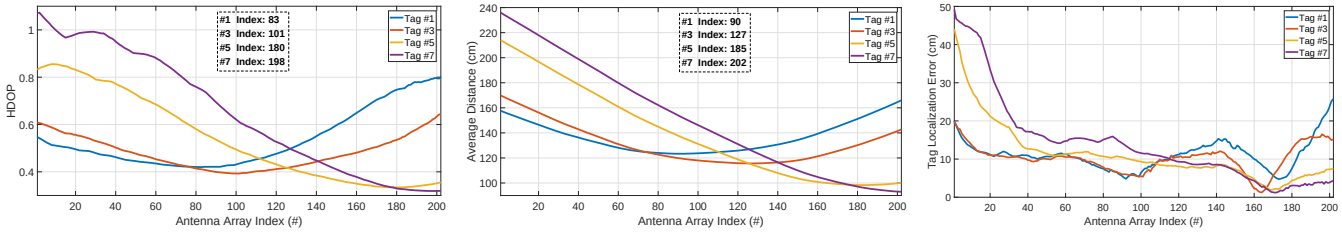
Software: According to Impinj LLRP Toolkit (LTK) [24], we program an RFID data collection application in C#. Each RFID report contains Electronic Product Code (EPC), RF phase, antenna port number, operating frequency and reading timestamp. We adopt phase calibration mechanism [25] to eliminate the effect of frequency hopping on phase measurements. ORB-SLAM2 [26] is a real-time SLAM system to estimate camera trajectory and 3D reconstruction for monocular, stereo and RGB-D cameras. It can achieve loop closure detection and camera re-localization, which is of essential importance in visual SLAM systems to reduce accumulated errors over time. We run the RGB-D component on Ubuntu and save the camera trajectory as ground truth. However, ORB-SLAM2 can only save keyframe camera poses for monocular camera rather than all frames. Hence, we record a video stream to a file by running ROS tool [27] and then calculate the 2D monocular poses (i.e., position and orientation) in Matlab [28] by analyzing these 2D images. Finally, the system performance of tag localization and trajectory recovery are all evaluated in Matlab, running on our laptop with 2.3 GHz CPU (Intel Core i5-6200U) and 4 G memory.

VII. Evaluation

In this section, we evaluate the 3D tag localization and camera trajectory recovery performance of RF-MVO. We first introduce the experiment setup and metrics, followed by the detailed experiment results.

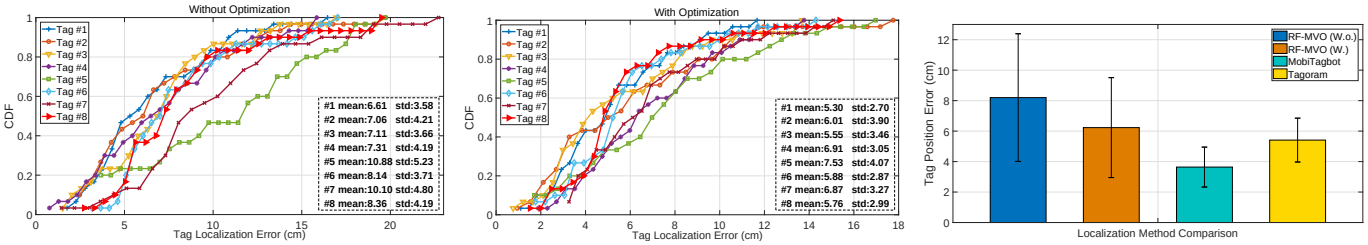
A. Experiment setup

Methodology: The experiment setup is shown in Fig. 6. We print the same checkerboards for camera calibration on both sides of a paper, where the size of each square is 29 mm. Four double-sided checkerboards are affixed on the flat wall, two of which has 4 RFID tags to be located at specified corners on its back. The 3D coordinate system is referred to the RGB camera of Kinect V2 at the initial time. The ground truth of these RFID tags in the 3D space can be measured by mapping points in the image coordinates to points in the world coordinates based on camera calibration [29]. In our experiment, the default parameters are set as follows: the number of antenna elements is $L = 60$; the granularities for initial incident angle and absolute scale factor searching are $\Delta\alpha = 5°$, $\Delta\beta = 5°$ and $\Delta\gamma = 0.1$; the granularities



(a) HDOP value. (b) Average distance. (c) Localization error with optimization.

Fig. 7. HDOP Performance.



(a) Without optimization. (b) With optimization. (c) Localization error comparison.

Fig. 8. Tag localization performance.

for coarse-to-fine angular refinement are $\Delta\alpha_{opt} = 1^\circ$ and $\Delta\beta_{opt} = 1^\circ$; the searching threshold is $\mu = 20^\circ$; the cart moves with the maximum speed of 100 cm/s.

Metrics: To verify RF-MVO performance, we focus on the deviations of the estimates from the ground truth. Suppose that $\mathbb{X}_T[1]$ is the 3D coordinate of an RFID tag T at initial time and $\mathbb{X}_C[i]$ is the ground-truth camera position corresponding to the i-th 2D image, then the tag position error is denoted as

$$T_{err}[i] = \|\boldsymbol{X}_T^*[i] - \mathbb{X}_T[1] + \mathbb{X}_C[i]\| \tag{30}$$

and the absolute scale factor error is

$$F_{err}[i] = \left|\gamma^*[i] - \left|\frac{\mathbb{X}_C[i+1] - \mathbb{X}_C[i]}{\boldsymbol{X}_C[i+1] - \boldsymbol{X}_C[i]}\right|\right| \tag{31}$$

B. HDOP Performance

To validate whether HDOP can effectively indicate the localization error, we refer to the average distances of each RFID tag to antenna elements in each antenna array. We move the cart at the almost same speed through RFID tags so that the displacement between consecutive antennas is close to each other. When the antenna moves close to an RFID tag, the wider angular separations between the elements can reduce the tag position uncertainty. In this case, the tag-antenna geometry is strong and the corresponding HDOP value is low. In contrast, when the antenna becomes far away from the tag, the geometry is weak and HDOP is high. Fig. 7 (a) and (b) show that for each RFID tag, the trend of HDOP and average distance curves is basically the same, and their indexes with the minimum value are also closer to each other, which demonstrates the proposed HDOP is an effective indicator for tag localization error. In Fig. 7(c), however, the lowest HDOP value does not mean the highest localization accuracy because the error level coming from camera pose estimation and RF phase measurement determines the final localization accuracy.

C. Tag localization performance

We move the utility cart carrying two RFID antennas and the camera near the RFID tags with different speeds and trajectories each time. For each RFID tag, the 3D tag position with minimum HDOP value over antenna arrays is selected as the optimal one. We repeat the experiment 30 times. Fig. 8 (a) and (b) plot the CDF of tag position error without and with the proposed joint optimization algorithm, and indicate the average and standard deviation of localization error for 8 RFID tags. We can see that localization errors are substantially reduced after using optimization algorithm. The overall average error drops from 8.20cm to 6.23cm.

In Fig. 8 (c), we also compare our fusion system with other purely RFID-based localization systems, i.e., Tagoram [1] and MobiTagbot [14]. In Tagoram, to eliminate the position ambiguity coming from phase periodicity, more than one fixed reader antennas are deployed far apart from each other. An RFID tag to be located is required to move along a given trajectory at a constant speed. To match our experiment, we can view that the tag is fixed at the initial position and the antennas move in the opposite direction. Its enhanced version MobiTagbot is equipped with stronger capability of multipath suppression than Tagoram by exploiting frequency hopping technique. The ground-truth trajectory is used in the experiment to evaluate their localization performance. Note that since our RFID-CV fusion system is designed for simultaneous tag localization and camera/antenna trajectory recovery, the tag localization performance in RF-MVO is subject to the estimation accuracy of absolute scale factor. The localization results show that Tagoram and MobiTagbot have the average localization errors of 5.41cm and 3.64cm, outperforming our method without and with optimization by more than 2.25 times and 1.71 times, respectively. However, both Tagoram and MobiTagbot need to provide an already-known tag/antenna trajectory. And they also need to specify the surveillance region of interest

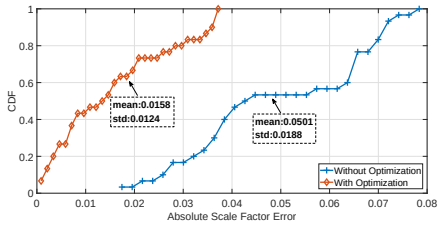


Fig. 9. The CDF of absolute scale factor error.

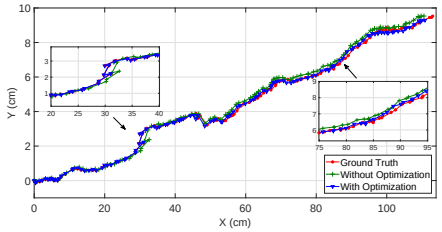
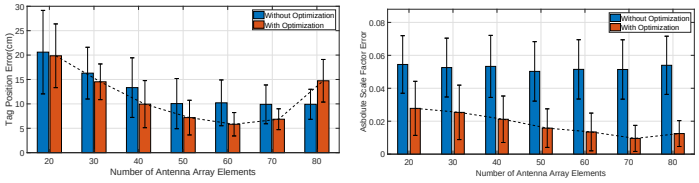
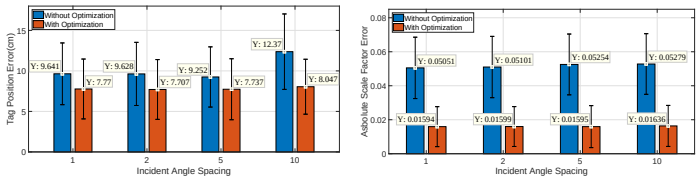


Fig. 10. Camera Trajectory Recovery.



(a) Tag Position error. (b) Absolute Scale factor error.

Fig. 11. Impact of array element size.



(a) Tag Position error. (b) Absolute Scale factor error.

Fig. 12. Impact of incident angle spacing.

where target RFID tags exist in advance. The region in the 3D space is partitioned into cuboids with the centimeter-level size. Each of them is determined the likelihood of containing a target RFID tag. 60×60×25 cuboids with the width of 1cm around the ground-truth tag positions are configured in the Tagoram and MobiTagbot. As the region of interest increases, huge searching computations will seriously affect the real-time performance. The proposed system of RF-MVO exploits the DOA-based method for tag localization, which is definitely different from them.

D. Trajectory recovery performance

According to above experiment data, we further verify the estimation accuracy of absolute scale factor. Since eight RFID tags may produce multiple candidates, an absolute scale factor with the minimum HDOP value in each antenna array is selected as the optimal one. Fig. 9 plots the CDF of absolute scale factor error. Without optimization, the mean estimation error is 0.0501 with the standard deviation of 0.0188. After applying the joint optimization, the average is reduced by 3.17 times, down to 0.0158 with the standard deviation of 0.0124. To better show the trajectory recovery performance, we choose an experiment data and plot the corresponding camera trajectory in Fig. 10. The estimated trajectory with optimization can better match the ground truth. However, estimating the camera poses and absolute scale factors will inevitably generate accumulating errors over time, making the trajectory increasingly deviate from the ground truth. We will take further study to reduce the drift in our future work.

E. Impact of array element size

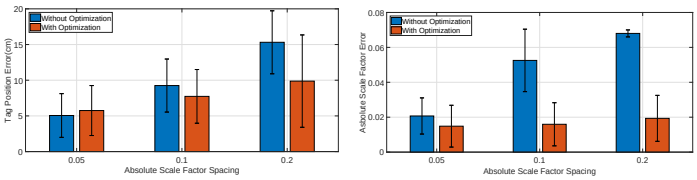
Given a cart trajectory, we vary the number of elements in an antenna array from 20 to 80 with a spacing of 10 elements. The estimation error is calculated by averaging the estimates of tag position and absolute scale factor in all antenna arrays. Fig. 11 (a) and (b) show that the more antenna elements can effectively reduce the estimation errors when the element size increases from 20 to 60 (70 in Fig. 11 b). The capability of the proposed spatial power spectrum to distinguish the absolute scale factor and incident angles in each antenna array can be reinforced around the ground truth. And when locating RFID tags, we take an average of azimuth and elevation angles over antenna arrays. As an element is involved in more antenna arrays, the estimates of incident angles also become more accurate for each element. However, when the element size increases from 60 (70 in Fig. 11 b) to 80, increasingly accumulated errors in camera pose estimation will reduce our system accuracy. In our experiment, the size of about 60∼70 elements in an antenna array can balance between estimation accuracy and accumulated error in monocular VO.

F. Impact of scale-factor and angular granularities

We vary the angular granularity with 1°, 2°, 5° and 10° while fixing the scale-factor granularity at 0.1. Fig. 12 (a) and (b) show its effect on estimation accuracy. Without optimization, tag localization error obviously increases when the angular granularity increases to 10°, while it has very little impact on absolute scale factor error. We consider that the proposed spatial power spectrum can find the optimal estimate of absolute scale factor if we could provide enough samples for our system. And after applying optimization algorithm, the estimation accuracy of tag position over angle spacings is very similar to each other. This is mainly attributed to coarse-to-fine incident angle refinement in the joint optimization. However, the larger angular granularity requires to set a larger search threshold μ in our joint optimization algorithm, which will inevitably increase computational load.

Then we vary the scale-factor granularity with 0.05, 0.1, 0.2 while the angular granularity is set to 5°. In Fig. 13 (a) and (b), the smaller scale-factor granularity is helpful for improving estimation accuracy. However, it will inevitably incur much more computations. For the granularity of 0.05, the spatial power spectrum takes the average of 14.24s runtime on our experiment platform for search.

There is a trade-off between real-time performance and estimation accuracy. Here we suggest scale-factor and angular granularities are set to 0.1 and 5°, respectively. In this case, RF-MVO will take on average 0.52s in each estimation of absolute scale factor and incident angles. We can potentially



(a) Tag Position error. (b) Absolute Scale factor error.

Fig. 13. Impact of absolute scale factor spacing.

improve real-time performance by reducing the searching range of absolute scale factor.

VIII. Related Work

Pure RFID Localization. BackPos [2] and RF-IDraw [4] locate RFID tags in a 2D plane using more than three reader antennas at specified positions. However, the localization accuracy will drop seriously with the increase of the distance between the RFID tag and reader antennas. Tagoram [1] can track moving RFID tags under known trajectories at cm-level tracking accuracy. PolarDraw [30] and Pantomime [31] can only achieve relative position tracking for moving RFID tags by exploiting two linearly reader antennas polarization and building a tag array with multiple RFID tags, respectively. STPP [9], PinIt [10], 3DLoc [32] and MobiTagbot [14] use moving reader antennas equipped on a mobile robot for localization. STPP [9] can only distinguish the left/right/up/down ordering of RFID tags. PinIt [10] exploits multipath propagation to locate stationary RFID tags while it needs to pre-deploy many reference RFID tags at specified positions. MobiTagbot [14] leverages the antenna movement and frequency hopping to emulate a set of virtual reader antennas and build a holography to locate RFID tags. However, as the surveillance region of interest increases, huge computations will jeopardize the real-time localization performance. 3DLoc [32] performs 3D localization by attaching three RFID tag arrays on different orthogonal surfaces of the cuboid object. Not all objects allow people to attach RFID tags like this. In this work, our hybrid system needs not to pre-deploy any reference RFID tag in the surveillance region, but also can deal with the unpredictable movement of reader antennas.

Camera-based Recognition. Convolutional Neural Networks (CNNs) have been widely employed to detect target objects from images, such as Faster R-CNN [33],YOLO [34] and SSD [35]. They need to pre-train a CNN-based detector for target object detection. When inputting an image into the detector, the systems enable to determine whether a target object exists in the image and if so where it occurs in the image. The estimated results are enclosed by a set of bounding boxes. In purely camera-based recognition systems, however, when multiple objects with same appearance exist or the target object is occluded, they may produce some unwanted errors. RFID technique can provide accurate object identification and even work in non-line-of-sight (NLOS) scenarios.

RFID and CV Fusion. The systems "Tell me what I see" [36], Stereo-RSSI [37], RF-ISee [38] and ID-Match [39] can recognize relative positions of multiple RFID-tagged objects by associating with the depth features of the stereo camera and RFID reports. TagVision [16] uses a 2D camera to capture trajectories of multiple mobile objects and then differentiate from them according to the correlations between RF phase and the distance from the camera to motion blobs in each trajectory. Existing work only addresses the matching problem between mobile RFID-tagged objects and CV-captured trajectories while our hybrid system can calculate the absolute position of the stationary target object carrying an RFID tag.

IX. Conclusion

In this work, we present RF-MVO, which relies on off-the-shelf RFID devices and a 2D monocular camera for 3D object localization and camera trajectory recovery. By exploiting the antenna mobility to build a series of antenna arrays, RF-MVO combines RF phase difference with the camera poses to estimate spatially incident angles and absolute scale factor in each antenna array. A joint optimization algorithm is proposed to improve real-time performance of RF-MVO. The concept of HDOP is introduced to determine the precision of estimated tag positions and scale factors. Experimental results show that RF-MVO can achieve fine-grained 3D tag localization and camera trajectory recovery. The proposed fusion system not only can locate RFID tags in a 3D space, but also can help existing purely RFID-based systems like [14], [40] work with an unknown antenna trajectory.

Acknowledgment

This research is partially supported by College Graduate Research Innovation Program of Jiangsu Province, China (No.KYZZ16_0260), National Natural Science Foundation of China (No.61572260, No.61572261, No.61373017 and No.61672297), Key Research and Development Program of Jiangsu Province, China (Social Development Program, No. BE2015702 and No.BE2017742) and Natural Science Foundation for Excellent Young Scholar of Jiangsu Province, China (BK20160089).

References

[1] L. Yang, Y. Chen, X. Li, C. Xiao, M. Li and Y. Liu, "Tagoram: Realtime tracking of mobile RFID tags to high precision using COTS devices," in *Proc. ACM MobiCom*, 2014.

[2] T. Liu, Y. Liu, L. Yang, Y. Guo and C. Wang, "BackPos: High Accuracy Backscatter Positioning System," *IEEE Trans. on Mobile Computing*, vol. 15, no. 3, pp. 586-598, Apr. 2016.

[3] F. Xiao, Z. Wang, N. Ye, R. Wang and X. Li, "One More Tag Enables Fine-Grained RFID Localization and Tracking," *IEEE/ACM Trans. on Networking*, vol. PP, no. 99, pp.1-14, Nov. 2017.

[4] J. Wang, D. Vasisht and D. Katabi, "RF-IDraw: Virtual touch screen in the air using RF signals," in Proc. ACM SIGCOMM, 2014.

[5] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): Part II," *IEEE Robotics & Automation Magazine,* vol. 13, no.3, pp. 108-117, 2006.

[6] R. Mur-Artal, J. M. M. Montiel and J. D. Tardos, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147-1163, Oct. 2015.

[7] R. Mur-Artal and J. D. Tards, "ORB-SLAM2: An Open-source SLAM System for Monocular, Stereo and RGB-D Cameras," *IEEE Transactions on Robotics*, vol. 33, no.5, pp. 1255-1262, 2017.

[8] P. V. Nikitin, R. Martinez, S. Ramamurthy, H. Leland, G. Spiess and K. Rao, "Phase Based Spatial Identification of UHF RFID Tags," in *Proc. IEEE RFID*, 2010.

[9] L. Shangguan, Z. Yang, A. X. Liu, Z. Zhou and Y. Liu, "Relative Localization of RFID Tags Using Spatial-Temporal Phase Profiling," in *Proc. USENIX NSDI*, 2015.

[10] J. Wang and D. Katabi, "Dude, wheres my card? RFID Positioning That Works with Multipath and Non-line of Sight", in *Proc. of ACM SIGCOMM*, 2013.

[11] iRobot Create 2, http://www.irobot.com/About-iRobot/STEM/Create-2.aspx.

[12] S. Davide and F. Friedrich, "Visual Odometry: Part I: The First 30 Years and Fundamentals," *IEEE Robotics & Automation Magazine*, vol. 18, no. 4, Dec. 2011.

[13] F. Fraundorfer and D. Scaramuzza, "Visual Odometry: Part II - Matching, Robustness, and Applications," *IEEE Robotics and Automation Magazine*, vol. 19, no.2, Feb. 2012.

[14] L. Shangguan and S. K. Jamieson, "The Design and Implementation of a Mobile RFID Tag Sorting Robot," in *Proc. of ACM MobiSys*, 2016.

[15] EPC Gen2, EPCglobal. www.gs1.org/epcglobal.

[16] C. Duan, X. Rao and L. Yang, "Fusing RFID and Computer Vision for Fine-grained Object Tracking," in *Proc. IEEE INFOCOM*, 2017.

[17] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision (Second Edition)," *Cambridge University Press*, 2000.

[18] X. S. Gao, X. R. Hou, J. Tang and H. F. Cheng, "Complete solution classification for the perspective-three-point problem," *IEEE TRANS on TPAMI*, vol. 25, no. 8, pp. 930-943, Aug. 2003.

[19] B. Triggs, P. F. McLauchlan, R. I. Hartley and A. W. Fitzgibbon, "Bundle adjustment-a modern synthesis," in *Proc. of International Workshop on Vision Algorithm*, 1999.

[20] P. Y. Zhang, J. Gummeson and D. Ganesan, "Blink: A high throughput link layer for backscatter communication," in *Proc. of ACM MobiSys*, 2012.

[21] Speedway Revolution Reader-Low Level User Data Support, https://support.impinj.com/hc/en-us/articles/202755318-Application-Note-Low-Level-User-Data-Support.

[22] L. Han and J. C. Bancroft, "Nearest approaches to multiple lines in n-dimensional space," *CREWES Research Report*, vol. 22, pp. 1-17, 2010.

[23] K. Madsen, H. B. Nielsen and O. Tingleff, "Methods for Non-Linear Least Squares Problems (Second Edition)", *Informatics and Mathematical Modelling*, Technical University of Denmark, 2004.

[24] LTK SDK. https://support.impinj.com.

[25] T. Wei and X. Zhang, "Gyro in the air: tracking 3D orientation of batteryless internet-of-things," in *Proc. of ACM MobiCom*, 2016.

[26] ORB_SLAM2. https://github.com/raulmur/ORB_SLAM2.

[27] ROS Tool. http://wiki.ros.org/image_view.

[28] Monocular Visual Odometry. https://au.mathworks.com/help/vision/examples/monocular-visual-odometry.html.

[29] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Trans. on TPAMI*, vol. 22, no. 11, pp. 1330-1334, Nov. 2000.

[30] L. Shangguan and K. Jamieson, "Leveraging Electromagnetic Polarization in a Two-Antenna Whiteboard in the Air," in *Proc. of ACM CoNEXT*, 2016.

[31] L. Shangguan, Z. Zhou and K. Jamieson, "Enabling Gesture-based Interactions with Objects," in *Proc. of ACM MobiSys*, 2017.

[32] Y. Zhang, L. Xie, Y. Bu, Y. Wang, J. Wu and S. Lu, "3-Dimensional Localization via RFID Tag Array," in *Proc. of IEEE MASS*, 2017.

[33] S. Ren, K. He, R. Girshick and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Proc. of NIPS*, 2015.

[34] J. Redmon, S. Divvala, R. Girshick, et al., "You only look once: Unified, real-time object detection," in *Proc. of IEEE CVPR*, 2016.

[35] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, et al., "SSD: Single shot multibox detector," in *Proc. of ECCV*, 2016.

[36] 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.

[37] D. F. Llorca, R. Quintero and I. Parra "Fusing Directional Passive UHF RFID and Stereo Vision for Tag Association in Outdoor Scenarios," in *Proc. of IEEE ITSC*, 2016.

[38] F. Cafaro, A. Panella, L. Lyons, J. Roberts and J. Radinsky, "I see you there! Developing Identity-preserving Embodied Interaction for Museum Exhibits," in *Proc. of ACM CHI*, 2013.

[39] H. Li, P. Zhang, S.A. 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 ACM CHI*, 2016.

[40] J. Wang, J. Xiong, X. Chen, H. Jiang, R. K. Balan and D. Fang, "TagScan: Simultaneous Target Imaging and Material Identification with Commodity RFID Devices," In *Proc. ACM MobiCom*, 2017.