**RESEARCH**　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

# An improved k-NN algorithm for localization in multipath environments

Yang Zhao, Kaihua Liu, Yongtao Ma[*] and Zhuo Li

**Abstract**

To improve the localization accuracy in multipath environments, this paper presents an effective localization approach with the utilization of reference tags. In this approach, an improved k-nearest neighbor (k-NN) algorithm is proposed based on radio-frequency (RF) phases. The traditional k-NN algorithm only focuses on the weighting factors of the coordinates of the selected reference tags, while the improved k-NN algorithm aims at the estimation of direct distance from a reader antenna to a target tag. Then, the location is estimated by linear least squares with a new reference selection scheme. Simulation results show that our approach is superior to the traditional localization approaches under multipath environments. In addition, we conclude that phase has the superiority over strength in the selection of reference tags for range estimation, and range estimation is more accurate than coordinate estimation with k-NN algorithm for localization.

**Keywords:** Indoor environments; RF phase; Reference tags; Improved k-nearest neighbor algorithm; Linear least squares

## 1 Introduction

Indoor location awareness is a hot research subject in many wireless systems due to its capability to provide a wide range of location-based services. Due to the multipath propagation channel in indoor environments, it is hard to achieve a high accuracy in localization which is urgently needed for both commercial and industrial interests. Many technologies and systems have been developed for indoor localization, and one of the most popular technologies is the ultrahigh frequency (UHF) radio-frequency identification (RFID) technology. UHF RFID signals can be active or passive based on whether the RFID tag has an internal source of energy or not. In this paper, we focus on passive RFID tags for their less cost, long lifetime, and maintenance-free characteristics.

Existing range measurement techniques for RFID-based application include the time of arrival (TOA) technique [1-4], the time difference of arrival (TDOA) technique [5,6], the phase of arrival (POA) technique [7-13], and the received signal strength indicators (RSSI) technique [14,15]. The TOA technique measures the one-way propagation time, and the distance between the transmitter and

the receiver can be calculated using the propagation velocity of the signal. But, a precise synchronization of all transmitters and receivers is required to guarantee the correct localization estimation. The TDOA technique does not require the synchronization of the transmitter and receiver and uses the time difference of a transmitted signal at the receivers, so it requires that the reference receiver subtracted be precise. The POA technique can measure the complete propagation distance of the signal between a transmitter and receiver but has a problem of having whole-cycle phase ambiguities in the calculation. The RSSI technique uses the attenuation of the sent signal to estimate the distance between the transmitter and the receiver. Though this technique is less complex than other techniques, it suffers from poor accuracy because the signal strength can be easily affected by various environment factors, such as multipath. So, the impact of the wireless channel on the received signals [16] is significant. In a radar sensor network, information diversity has been utilized for the improvement of detection accuracy [17-19] and reliable communications [20]. Based on the above-mentioned range measurement techniques, the target location can be estimated by using the geometric properties of triangles [21].

Due to the inadequate accuracy of range measurement techniques, localization algorithms using scene analysis

* Correspondence: mayongtao@tju.edu.cn
School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China

is developed to improve the overall accuracy. As an effective scene analysis localization algorithm, the k-nearest neighbor (k-NN) [22] algorithm is proposed by utilizing reference tags. This algorithm estimates the target localization by using the coordinates of the reference tags which are closest to the target tag with RSSI Euclidean distances. Since severe multipath and interference effects can easily affect signal strength, a distant reference may be selected as the candidate tag, significantly lowering the localization accuracy [23]. Even so, strength is still commonly used in scene analysis algorithm. In recent years, another way to compute the location is proposed by estimating the distance from a reader antenna to the target tag with the help of selected reference tags just as range measurement techniques do [24-26]. Therefore, we consider that a similar measurement value means a similar distance between a pair of tag-and-reader antenna, not only the closest coordinate position in space.

In this paper, we propose an indoor localization approach following the concept of reference tags in the k-NN algorithm which is a further study on the localization approach proposed in [24]. We substitute phase for strength in the Euclidean distance calculation just as the approach proposed in [24]; however, we make a lot of improvement.

This paper is organized as follows. In Section 2, the radio-frequency (RF) phase is measured with a single-frequency subcarrier amplitude modulation. In Section 3, an improved k-NN algorithm is proposed to estimate the direct distance between a reader antenna and the target tag, and the localization is estimated by linear least squares algorithm with a new reference selection scheme. Simulative results are illustrated in Section 4, and the improvement in localization accuracy is shown to prove the effectiveness of the proposed approach. In the end, the conclusion is presented in Section 5.

## 2 RF Phase
Assuming that all reader antennas emit pure cosine signals, we can utilize the phase of arrival technique to measure distance which does not exceed half of the signal wavelength. Since the UHF RFID signal wavelength is tens of centimeters, the actual distance is a few meters or even dozens of meters between a reader and a tag. So, if we utilize the extracted phases of UHF RFID signals directly, there will be a whole-cycle carrier phase ambiguities problem in distance estimation. To avoid this problem, we should use a low-frequency signal with a longer wavelength to measure the phase accumulated during the signal travel from the reader to the tag and back.

### 2.1 Traditional RF-phase-based ranging algorithm
For the purpose of eliminating the whole-cycle carrier phase ambiguities, the frequency of the signal for phase extraction should be low. In other words, the round trip distance should be shorter than the wavelength of the signal. Meanwhile, we also need an extra energy signal to start up the passive tag. In summary, we should use a low-frequency signal to measure the phases and a high-frequency signal to support energy simultaneously in the whole communication process. With the concept of subcarrier amplitude modulation in [27], we utilize a single-frequency subcarrier in the communication process, where the low-frequency signal is taken as the subcarrier signal and the high-frequency signal is the carrier signal. The transmitted signal is expressed as Equation 1 and the received signal expressed as Equation 2.

$$s(n) = A[1 + m_a \cos(\omega_s t + \phi_s)] \cos(\omega_c t + \theta_s) \tag{1}$$

$$r(n) = A' m(t)[1 + m_a \cos(\omega_s t + \phi_r)] \\ \cos(\omega_c t + \theta_r) + n(t) \tag{2}$$

where $\omega_s$ stands for the angular frequency of the subcarrier; $\phi_s$ is the original phase of the subcarrier; $\omega_c$ stands for the angular frequency of the carrier; $\theta_s$ is the original phase of the carrier; $A$ and $A'$ are the amplitudes of the carrier in transmitting and receiving, respectively; $m_a$ denotes the modulation index; and $n(t)$ is the addictive noise of the wireless channel. $\phi_r$ and $\theta_r$ are the phases of the received subcarrier and the carrier, respectively. $m(t)$ is the baseband signal of the tag.

The zero-IF scheme is a way to realize the down conversion of the received signal. This scheme needs neither modification of the current structure of the RFID readers nor extra hardware, which reduces the cost to a great extent than the sub-Nyquist sampling used in [24]. In this scheme, the received signal $r(n)$ will be converted in an in-phase/quadrature (I/Q) demodulator, which splits the received signal into two parts by mixing it with the local oscillator signal $\cos(\omega_c t + \varphi)$ and the local oscillator signal shift by 90°, respectively, as in Equations 3 and 4.

$$r_I(t) = r(t) \cos(\omega_c t + \varphi) \tag{3}$$

$$r_Q(t) = r(t) \sin(\omega_c t + \varphi) \tag{4}$$

The two $I$ and $Q$ component signals are filtered to remove the high-frequency signal components, leaving behind the subcarrier signal containing the phase of signal as in Equations 5 and 6. Then, we use the sampling frequency $f_s$ to sample the signals and extract the subcarrier phase information in the digital domain with the all-phase FFT [28].

$$I(t) = \frac{A'}{2} m(t) m_a \cos(\omega_s t + \varphi_r) \cos(\theta_r - \phi) \\ + N(t) \cos(\omega_c t + \varphi) \tag{5}$$

$$Q(t) = -\frac{A'}{2} m(t) m_a \cos(\omega_s t + \phi_r) \sin(\theta_r - \phi) + N(t) \sin(\omega_c t + \phi) \tag{6}$$

So, the subcarrier phase information extracted can be expressed as

$$\phi = \phi_r - \phi_s \tag{7}$$

In the end, the calculated propagation distance can be written as

$$\overset{\wedge}{d} = \frac{c\phi}{2\omega_s} \tag{8}$$

where $c$ is the speed of the electromagnetic wave propagation.

## 2.2 RF Phase analysis

As we know, the phase of the received signal in real propagation environments can be expressed as

$$\phi = \phi_{prop} + \phi_m + \phi_o + \phi_b \tag{9}$$

where $\phi_{prop}$ stands for the phase accumulated during the round direct trip between a reader antenna and a tag antenna; $\phi_m$ is the phase offset of the multipath effect and the reflection from the obstacles; $\phi_o$ represents phase shifts of the reader and antenna components, as well as cables; and $\phi_b$ is the backscattered phase of the tag modulation [29].

For accurate localization estimation, we should only extract the $\phi_{prop}$ component for distance calculation in Equation 10. As the other phase error components cannot be eliminated completely, the exact phase for measuring propagation distance is a tough task.

$$d = \frac{c\phi_{prop}}{2\omega_s} \tag{10}$$

In our paper, we introduce the reference tags just like [22,26] but substitute phase for strength in the reference tag selection. The positions of reference tags and reader antennas are known beforehand which means that the direct distance between them can be calculated directly. As the reference tags are involved in the same environment as the tags to be located, the direct distances between them can be used to improve localization accuracy.

## 3 Localization approach

### 3.1 Traditional k-NN algorithm

The LocAtioN iDentification based on dynaMic Active Rfid Calibration (LANDMARC) [22] system uses the k-NN algorithm with reference tags to estimate the locations of target tags. Let $N$ be the number of reader antennas and $M$ be the number of reference tags. The

signal strength vector of the $i$th target tag is defined as $TS_i = (ts_{(i,1)}, ts_{(i,2)}, ..., ts_{(i,N)})$ where $ts_{(i,j)}$ denotes the signal strength of the $i$th target tag perceived by the $j$th reader antenna, where $j \in (1, N)$. For the $m$th reference tag, the corresponding strength vector is $RS_m = (rs_{(m,1)}, rs_{(m,2)}, ..., rs_{(m,N)})$. The Euclidean distance in signal strength between the $i$th target tag and $m$th reference tag is defined as

$$e_{(i,m)} = \sqrt{\sum_{j=1}^{N} \left( ts_{(i,j)} - rs_{(m,j)} \right)^2} \tag{11}$$

So the $i$th target tag has its strength Euclidean distance vector $e_i$ as

$$e_i = \left( e_{(i,1)}, e_{(i,2)}, ..., e_{(i,M)} \right) \tag{12}$$

After sorting $e_i$ in ascending order, the vector can be defined as

$$E_i = \left( E_{(i,1)}, E_{(i,2)}, ..., E_{(i,M)} \right) \tag{13}$$

Then, according to the k-NN algorithm, the first $K$ values can make a new Euclidean distance vector $E_i = (E_{(i,1)}, E_{(i,2)}, ..., E_{(i,K)})$. The weighting factor for each selected reference tag can be calculated by

$$\omega_{(i,k)} = \frac{1/E_{(i,k)}^2}{\sum_{k=1}^{K} 1/E_{(i,k)}^2} \tag{14}$$

In the last, the estimated location of the $i$th target tag is given by

$$(\overset{\wedge}{x}_i, \overset{\wedge}{y}_i, \overset{\wedge}{z}_i) = \sum_{k=1}^{K} \omega_{(i,k)} (x_k, y_k, z_k) \tag{15}$$

where $(\overset{\wedge}{x}_i, \overset{\wedge}{y}_i, \overset{\wedge}{z}_i)$ is the estimated coordinate of the $i$th target tag, and $(x_k, y_k, z_k)$ denotes the coordinate of the $k$th selected reference tag.

### 3.2 Phase-based improved k-NN algorithm

In this paper, we propose a phase-based improved k-NN algorithm (Phase (improved k-NN)). The phase vector of the $i$th target tag is defined as $TP = (tp_{(i,1)}, tp_{(i,2)}, ..., tp_{(i,N)})$, where $tp_{(i,j)}$ denotes the phase of the $i$th target tag perceived by the $j$th reader antenna, where $j \in (1, N)$. And, we define the phase of the reference tags perceived by the $j$th reader antenna as $RP_j = (rp_{(1,j)}, rp_{(2,j)}, ..., rp_{(M,j)})$, where $rp_{(m,j)}$ denotes the phase of the $m$th reference tag perceived by the $j$th reader antenna, where $m \in (1, M)$. The phase Euclidean distance between the $i$th target tag and $m$th reference tag about the $j$th reader antenna is defined as

$$\mathrm{ep}_{(m,j)} = \mathrm{abs}\left(\mathrm{tp}_{(i,j)} - \mathrm{rp}_{(m,j)}\right) \qquad (16)$$

An important difference in our algorithm is that we select the $K$ nearest reference tags by calculating the difference between the target tag and the reference tags about each one of the reader antennas as Equation 16, not the sum of all those reader antennas as Equation 11. So, the phase Euclidean distance vector $\mathrm{ep}_j$ about the $j$th reader antenna is defined as $\mathrm{ep}_j = (\mathrm{ep}_{(1,j)}, \mathrm{ep}_{(2,j)},\ldots, \mathrm{ep}_{(M,j)})$. We also sort $\mathrm{ep}_j$ in ascending order and select the first $K$ values, i.e., $\mathrm{EP}_j = (\mathrm{EP}_{(1,j)}, \mathrm{EP}_{(2,j)},\ldots, \mathrm{EP}_{(K,j)})$. As we know, the noise is random, we repeat the above calculation for $Q$ times and record the selected reference tag numbers and its corresponding $\mathrm{EP}_{(k,j)}$ each time. Then, we get a list of the number of the selected reference tag and calculate the frequency of selection for each reference tag number, then select the first $K$ reference tags with the highest frequency as the final selected reference tags, i.e., $f_1, f_2,\ldots, f_K$. Meanwhile, we compute the averaged phase Euclidean distance of these selected reference tags, i.e., $\mathrm{ave\_EP}_{(k,j)}$. The weighting factor for each neighboring reference tag is calculated by

$$\omega_{(k,j)} = \frac{f_k \mathrm{ave\_EP}_{(K,j)}}{\displaystyle\sum_{k=1}^{K} f_k \mathrm{ave\_EP}_{(K,j)}} \qquad (17)$$

When we get the weighting factor for each selected reference tag, we use these weighting factors to estimate the direct distance $\hat{d}_j$ between $i$th target tag and the $j$th reader antenna.

$$\hat{d}_j = \sum_{k=1}^{K} \omega_{(k,j)} \mathrm{Rd}_{(k,j)} \qquad (18)$$

where $\mathrm{Rd}_{(k,j)}$ is the direct distance between $k$th selected reference tag and the $j$th reader antenna, and as the reference tags are fixed, $\mathrm{Rd}_{(k,j)}$ is known beforehand. And, the ranging error is given by

$$e_{\mathrm{d}} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\hat{d}_j - d_j\right)^2} \quad , j = 1, 2, ..., N \qquad (19)$$

where $d_j$ is the real distance of the target tag.

The reason that we prefer the direct distance between the reader antenna and reference tags to the coordinate of the reference tags can be explained in Figure 1. We utilize one reader antenna as an example; the transmitted signal is sent via the downlink channel (marked as a solid line) to the tag and returned over the uplink channel (marked as a dotted line). The square stands for the target tag and the triangles (R1, R2, R3, and R4) stand for the reference tags. As the R1, R2, and R3 reference tags are surrounding the target tag, they are likely to

have some similarities in the measurement informations. Reference R4 has the same direct distance as the target tag from the reader antenna. In multipath propagation environments, the measurement of reference R4 may also be similar with the target tag. If we set the $k$ value in the k-NN algorithm as 4, the selected reference tags will be R1, R2, R3, and R4. If we still use the coordinates of the selected tags, the localization will be estimated with considerable error. On the contrary, their direct distances are approximately equal. So, we think range estimation is more appropriate for localization than coordinate estimation.

### 3.3 Localization algorithm

The distance $d_j$ between $i$th target tag and the $j$th reader antenna defines a spherical surface around the $j$th reader antenna, and the localization equation is nonlinear as Equation 19.

$$\begin{aligned}
(x_i - x_1)^2 + (y_i - y_1)^2 &= \hat{d}_1^2 \\
(x_i - x_2)^2 + (y_i - y_2)^2 &= \hat{d}_2^2 \\
&\cdots \\
(x_i - x_j)^2 + \left(y_i - y_j\right)^2 &= \hat{d}_j^2 \\
&\cdots \\
(x_i - x_N)^2 + (y_i - y_N)^2 &= \hat{d}_N^2
\end{aligned} \qquad (20)$$

where $(x_i, y_i)$ is the $i$th target tag localization and $(x_j, y_j)$ is the $j$th reader antenna localization. We can make Equation 20 linear by fixing the expression of the $r$th reader antenna, i.e., the reference reader antenna, subtracting it from the rest of the equations. As for the selection of the $r$th reader antenna, we propose a reference reader antenna selection scheme. Firstly, we calculate the distance $\mathrm{rd}_{(m,j)}$ between the $m$th reference tag and $j$th reader antenna by Equation 8 with the $Q$-averaged reference tag phases. Next, we compute the estimation error of each reader antenna as

$$\mathrm{rd}_j = \sum_{m=1}^{M} \left(\mathrm{rd}_{(m,j)} - \mathrm{Rd}_{(m,j)}\right)^2, \quad j = 1, 2, ..., N \qquad (21)$$

The one with the least estimation error is marked as the $r$th reader antenna. Then, we can get the final localization of the $i$th target tag by utilizing the linear least squares (LLS) algorithm. In [30], the reference selection is given by comparing the distances from the target tag to reader antennas, so we name it LLS-RS-TT.

$$r = \arg \min_j \left\{\hat{d}_j\right\}, \quad j = 1, 2, ..., N \qquad (22)$$

where $\hat{d}_j$ is the result of Equation 18. While the reference selection scheme proposed in our paper is processed by comparing the distances from the reference tags to reader antennas, we name it LLS-RS-RT.
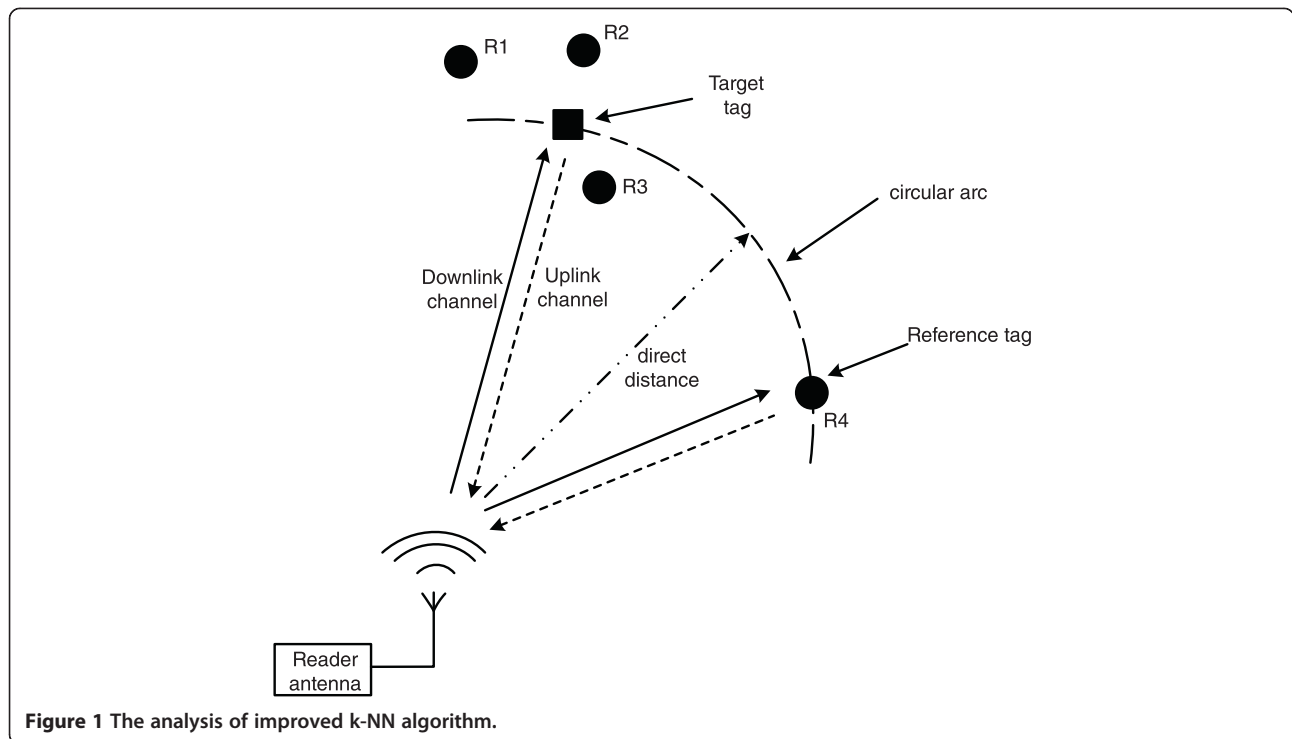
**Figure 1 The analysis of improved k-NN algorithm.**

In the end, the location estimation error $e_i$ is given by

$$e_i = \sqrt{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2 + (\hat{z}_i - z_i)^2} \qquad (23)$$

where $(\hat{x}_i, \hat{y}_i, \hat{z}_i)$ and $(x_i, y_i, z_i)$ denote the estimated coordinate and the real coordinate of the $i$th target tag, respectively.

### 3.4 Comparison with other localization approaches

To investigate the performance of the approach Phase (improved k-NN) + LLS-RS-RT proposed in our paper, we simulate some other localization approaches in the same scenario of this paper. These comparison approaches are as follows:

1) Strength (k-NN): It is processed just as the LANDMARC [17] system is processed. And for comparison, the localization results of this approach are averaged for $Q$ times, as we can get a result per simulation time during the $Q$ times.
2) Phase (k-NN): It has the same procedure of Strength (k-NN), but it utilizes the phase instead of the strength of the signals to compute the Euclidean distance. The final results are averaged for $Q$ times.
3) Strength (improved k-NN) + LLS-RS-TT: It calculates the strength Euclidean distance by Equation 16 like calculating the phase Euclidean distance and estimates the direct distances from reader antennas and the

target tag by Equations 17 and 18. Location is estimated by LLS-RS-TT.
4) Phase (1-NN) + LLS-RS-TT: In [24], we use the 1-nearest neighbor (1-NN) to estimate the direct distance from the target tag to the reader antenna and utilize LLS-RS-TT to calculate the location. For comparison, we average the localization results of each target tag for $Q$ times.
5) Multi-frequency + LLS-RS-TT: Here we use the method FM2 with multi-frequency in [31] to calculate the tag distance and estimate the location by LLS-RS-TT.
6) Phase (improved k-NN) + LLS-RS-TT: It has the same ranging procedure of Phase (improved k-NN) + LLS-RS-RT except the location is estimated by LLS-RS-TT.

## 4 Simulation

We evaluate the performance of Phase (improved k-NN) + LLS-RS-RT and compare it with the other five approaches explained in Section 3.4. It indicates that our approach can improve the accuracy significantly under multipath environments and is superior to the other approaches.

### 4.1 Simulation setup

These simulations are implemented in MATLAB language and tested on a PC with an Intel Core i3-3220 CPU of 3.30 GHz and DDR3 SDRAM of 4 GB. Here, we

use eight reader antennas placed around the eight corners of an indoor scene which is modeled in a space of $4 \times 4 \times 4$ m$^3$. As for the reference tags, we choose 13 reference tags which are numbered and arranged as Figure 2, with a height of 2 m. The carrier frequency is 910 MHz, and the subcarrier frequency is 2 MHz. The $Q$ value in the improved k-NN algorithm is 100, and the $K$ value is set to 4 as a matter of experience. We set 1,000 target tags randomly in the scenario and each with a height of 2 m for space average. Because the $Q$ factor used in the improved k-NN algorithm is like some kind of time average, we should also simulate the comparison approaches without using the improved k-NN algorithm for $Q$ times and average the results for comparison. In this paper, the root mean square error (RMSE) evaluation factor is used as the rule of localization accuracy. And, we express it as

$$\text{RMSE} = \sqrt{E\left[(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2 + (\hat{z}_i - z_i)^2\right]} \quad (24)$$

where $E$ denotes the averaging operator for the 1,000 target tags, $(\hat{x}_i, \hat{y}_i, \hat{z}_i)$ denotes the averaged estimated position of the $i$th tag for $Q$ iterations, and $(x_i, y_i, z)$ denotes the real position of the $i$th tag. So, in our simulation, RMSE is the average value in both space and time. The smaller the value of RMSE is, the higher the localization accuracy is.

The RFID channel contains a downlink channel and an uplink channel at one time. In our simulation, each of these two channels is modeled as a linear time variant filter channel. The corresponding channel pulse response is
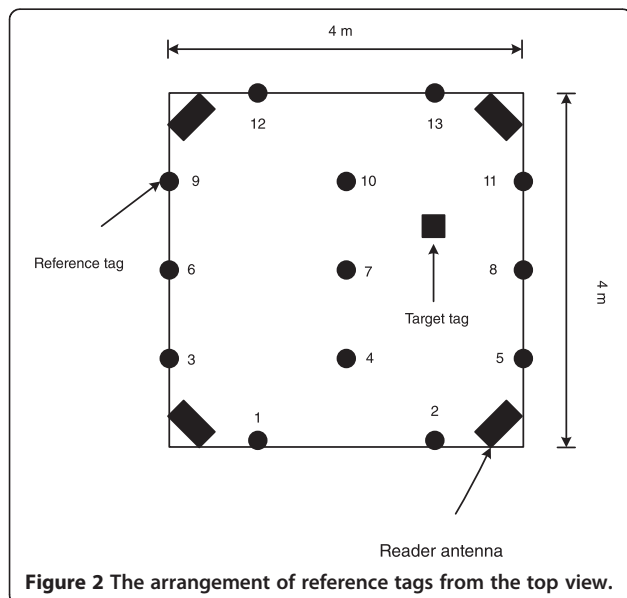


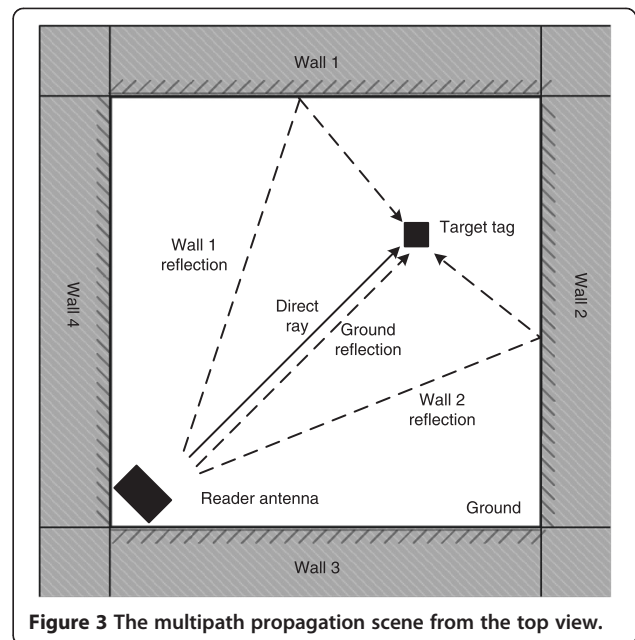**Figure 2 The arrangement of reference tags from the top view.**



**Figure 3 The multipath propagation scene from the top view.**

$$h(t) = a_{\text{los}} \cdot \delta(t - \tau_{\text{los}}) + \sum_{i=1}^{P} a_{\text{nlos},i} \cdot \delta(t - \tau_{\text{nlos},i}) \quad (25)$$

contains a direct line-of-sight (LOS) path with a time delay of $\tau_{\text{los}}$ and path gain of $a_{\text{los}}$ and a sum of $P$ non-line-of-sight (NLOS) paths of delay $\tau_{\text{nlos},i}$ and path gain of $a_{\text{nlos},i}$. The path gain is modeled by path-loss and log-normal shadowing, and the expressions of reflection coefficients in [32] are also considered for the calculation of path gain of each NLOS path. Then, the complex path gain model of the deterministic multipath channel between the reader and the tag is given by [8]. Here, we suppose there are four propagation paths under the multipath environment, including the direct ray, the two wall reflection paths, and the ground reflection path, as shown in Figure 3. For simplicity, we just consider the case that the signal propagates back just as the way it sends. Thus, the received RF signal can be expressed

$$y(t) = x(t) \otimes h_d(t) \otimes h_u(t) + n(t) \quad (26)$$

where $x(t)$ denotes the sending signals and $h_d$ and $h_u$ are channel pulse responses of downlink and uplink channels, respectively, and are equal in our simulation. $n(t)$ is the additive Gaussian noise.

## 4.2 Effect of the factor $Q$

In our approach, the location of each target tag is estimated by $Q$ iterations. So, $Q$ is a key factor for the improvement of performance in localization. But, the number of iterations will directly influence the running time of calculation. So, we should make a compromise between the accuracy and the time complexity. Figure 4

**Figure 4 Effect of the factor Q.**
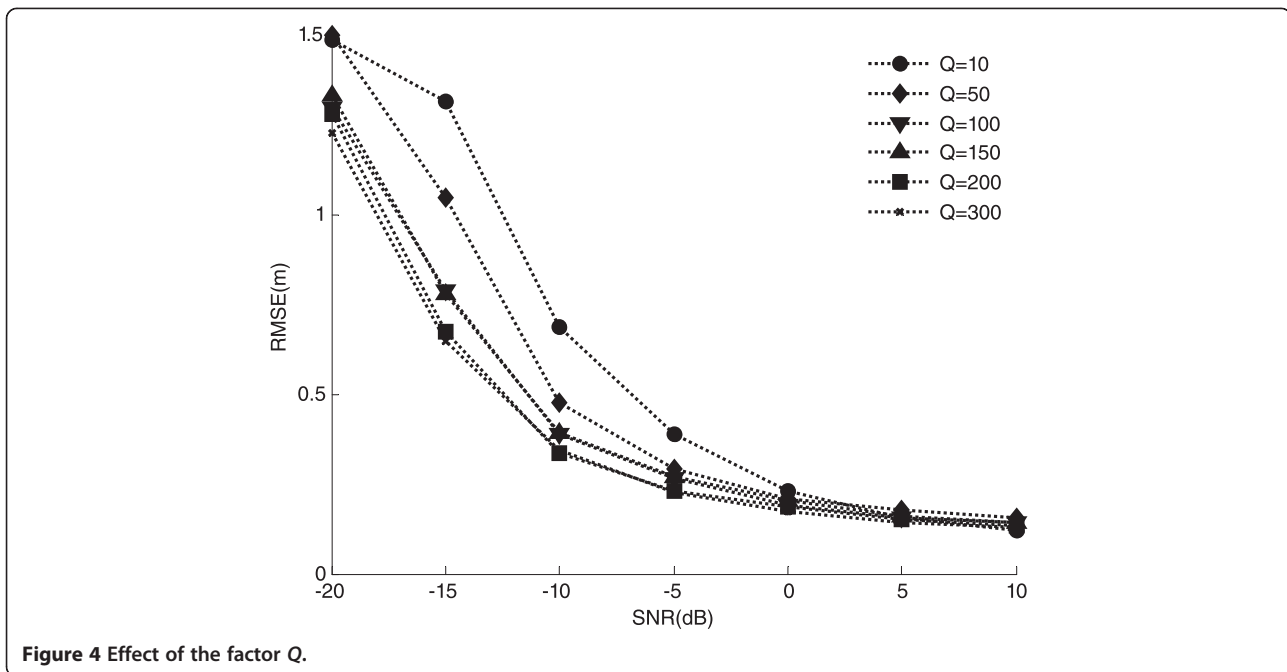
shows the accuracy comparison between different $Q$ values. From the simulation results, we find that the location estimation error will decrease with the increase of the $Q$ value. But when the $Q$ value is larger than 100, there is no obvious increase in accuracy. So, 100 is an appropriate value for factor $Q$ in our following simulations.

### 4.3 Ranging accuracy

In this section, the averaged ranging error is calculated by averaging the errors of all target tags. Figure 5 shows the ranging performance comparison between different approaches. From the figure, we can see that our algorithm is superior to the other three approaches. When the signal-to-noise ratio (SNR) value is −20 dB, the error of Phase (improved k-NN) is almost the same as the result of Strength (improved k-NN), but our algorithm is much better under higher SNR values. The results indicate that phase is an excellent measurement indicator for range estimation in k-NN algorithm.

### 4.4 Localization accuracy

In this section, averaged location estimation error is shown as the RMSE changing with the SNR values. As shown in Figure 6, the performance of the Multi-frequency + LLS-RS-TT is the worst than other approaches under low SNR values. Phase (1-NN) + LLS-RS-TT does not improve the accuracy effectively when the SNR value increases, so the ranging procedure needs to be revised. Phase (k-NN) and Strength (k-NN) have the same performance when the SNR value is low, but Strength (k-NN) works much better than Phase (k-NN) in higher SNR

values. So, we can conclude that strength is more suitable for non-ranging algorithms than phase. Strength (improved k-NN) + LLS-RS-TT is better than the above four approaches but it is still worse than Phase (improved k-NN) + LLS-RS-RT and Phase (improved k-NN) + LLS-RS-TT. It means that phase has the superiority over strength in the selection of the reference tags for range measurement, and range estimation is more accurate than coordinate estimation with k-NN algorithm for localization. The accuracy of Phase (improved k-NN) + LLS-RS-RT is superior to Phase (improved k-NN) + LLS-RS-TT under low SNR environments and almost equal when the SNR is higher. So, we think LLS-RS-RT is a better reference selection scheme in LLS algorithm with reference tags, and LLS-RS-TT is also an effective selection criterion in approaches without reference tags. In addition, the localization accuracy of our approach does not change so obviously among the range of the SNR values. So, our approach is robust to environment interferences.

Figure 7 illustrates the cumulative distribution function (CDF) of the localization accuracy for different approaches under two different SNR values. Figure 7a shows the CDF results when the SNR is −20 dB. Based on the statistics, it can be seen that the 80 percentile has a localization estimation error under 0.78 m and the 90 percentile is under 0.90 m in Phase (improved k-NN) + LLS-RS-RT which is the approach proposed in our paper. The performance of Phase (improved k-NN) + LLS-RS-TT is 0.93 and 1.04 m, respectively. Then, the results of the other approaches are much more than 1 m corresponding to the two percentiles. Figure 7b illustrates
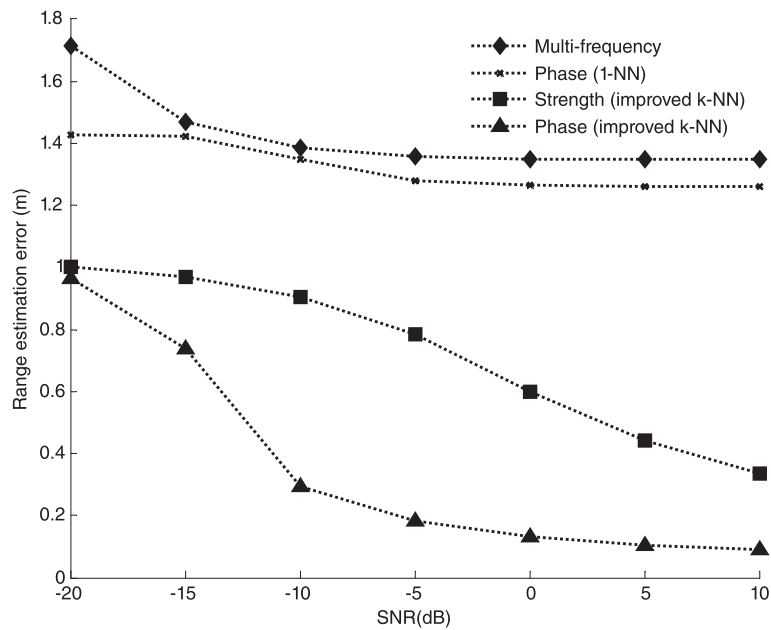
**Figure 5 Ranging error.**

the corresponding results when the SNR is 0 dB. The curves of the CDF for Phase (improved k-NN) + LLS-RS-RT and Phase (improved k-NN) + LLS-RS-TT overlap almost completely. With the improvement of the SNR value, the difference between the other approaches is more obvious. The error range of Phase (k-NN) is between 0.44 and 2.87 m. We think phase is more appropriate for ranging algorithms.

## 4.5 Cost analysis

We analyze the cost of different approaches compared in our paper. From the description of each approach, we know that Phase (improved k-NN) + LLS-RS-RT, Phase (improved k-NN) + LLS-RS-TT, and Strength (improved k-NN) + LLS-RS-TT all finish the location estimation of each target tag after $Q$ iterations, while the other approaches can get the final results at one
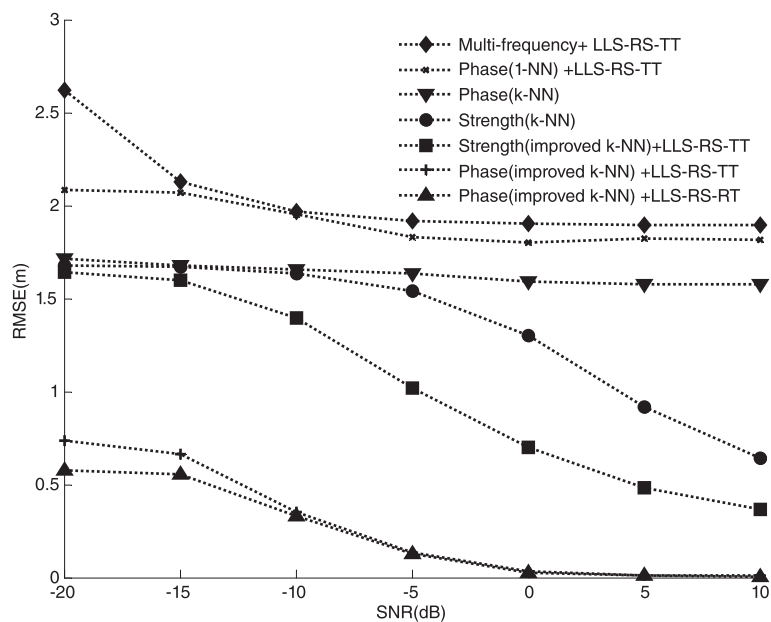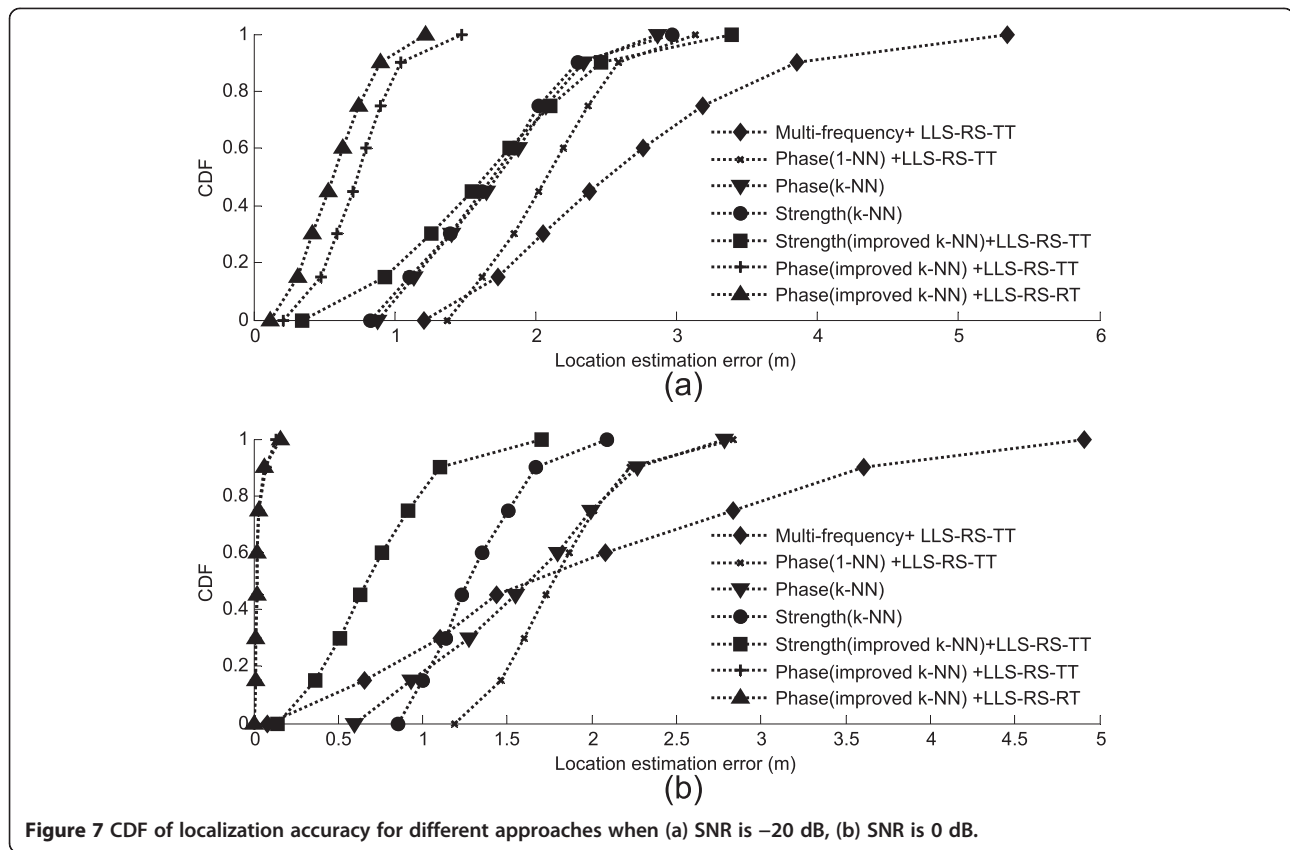


**Figure 6 RMSE versus SNR values for the different approaches.**

**Figure 7 CDF of localization accuracy for different approaches when (a) SNR is −20 dB, (b) SNR is 0 dB.**

time. Considering that the *Q*-times averaged results should be utilized to compare the performance of different approaches, we can find that the time complexity of each approach is almost the same. And, we also record the running time of each approach in Table 1. From the results in Table 1, we think the time complexity of each approach is of the same order.

In reference to space complexity, the size of Phase (improved k-NN) + LLS-RS-RT approximately equals to QNM × 64 bits ≈ 0.66 Mbits. So, the cost of Phase (improved k-NN) + LLS-RS-RT is far lower than 50 Mbits which is the standard maximum on-chip memory size of SRAM; we can consider memory costs are the same among these approaches.

As the averaged accuracy performance of Phase (improved k-NN) + LLS-RS-RT is significantly better than the other approaches, considering the current hardware technology, performance of our approach is superior to the other approaches.

## 5 Conclusions

In this paper, we use the RF phase to evaluate the similarity of the reference tags in an improved k-NN algorithm for range estimation. And the location is estimated by linear least squares with a new reference selection scheme. To validate the performance of the approach proposed in our paper, we compare it with six different approaches. Simulation results show that our approach is superior to the traditional localization approaches under multipath environments. We also find that phase can provide a more effective selection of the reference tags than strength in ranging estimation. And, ranging estimation is more accurate than coordinate estimation with k-NN algorithm for localization. Considering the current hardware technology, the performance of our approach is superior to the other approaches. Our future work is to study the effects of the placement of the reference tags and the tag interaction on the phase measurement accuracy and develop more effective localization algorithms in the NLOS environments.

**Table 1 Running time**

| Cost analysis | Strength (k-NN) | Phase (k-NN) | Strength (improved k-NN) + LLS-RS-RT | Phase (1-NN) + LLS-RS-TT | Multi-frequency + LLS-RS-TT | Phase (improved k-NN) + LLS-RS-RT |
|---|---|---|---|---|---|---|
| Running time (ms) | 9 | 9 | 13 | 16 | 4.5 | 13 |

## Abbreviations

UHF: ultrahigh frequency; RFID: radio-frequency identification; k-NN: k-nearest neighbor; LLS: linear least squares; TDOA: time difference of arrival; RSSI: received signal strength indicators; I/Q: in-phase/quadrature; LOS: line-of-sight; NLOS: non-line-of-sight; CDF: cumulative distribution function; SNR: signal-to-noise ratio.

## Competing interests

The authors declare that they have no competing interests.

## References

1. NA Alsindi, B Alavi, K Pahlavan, Measurement and modeling of ultrawideband TOA-based ranging in indoor multipath environments. Vehicul. Technol. IEEE Trans. **58**(3), 1046–1058 (2009). doi:10.1109/tvt.2008.926071
2. Y Geng, J He, K Pahlavan, Modeling the effect of human body on TOA based indoor human tracking. Int. J. Wireless Inf. Netw. **20**(4), 306–317 (2013). doi:10.1007/s10776-013-0227-3
3. KW Cheung, HC So, WK Ma, YT Chan, Least squares algorithms for time-of-arrival-based mobile location. Signal Process IEEE Trans. **52**(4), 1121–1130 (2004). doi:10.1109/tsp.2004.823465
4. A Mallat, J Louveaux, L Vandendorpe, M Di Dio, M Luise, Discrete Fourier transform-based TOA estimation in UWB systems. EURASIP J. Wireless Commun. Netw. **3** (2012). doi:10.1186/1687-1499-2012-3
5. C Steffes, S Rau, Multipath detection in TDOA localization scenarios, in *Sensor Data Fusion: Trends, Solutions, Applications (SDF), 2012 Workshop on, Bonn*, 2012
6. H Yanchuan, PV Brennan, A Seeds, Active RFID location system based on time-difference measurement using a linear FM chirp tag signal, in *Personal, Indoor and Mobile Radio Communications, IEEE 19th International Symposium on* (Cannes, 2008)
7. C Steiner, A Wittneben, Efficient training phase for ultrawideband-based location fingerprinting systems. Signal Process IEEE Trans. **59**(12), 6021–6032 (2011). doi:10.1109/tsp.2011.2166390
8. PV Nikitin, R Martinez, S Ramamurthy, H Leland, G Spiess, KVS Rao, Phase based spatial identification of UHF RFID tags, in *RFID, 2010 IEEE International Conference on* (Orlando, FL, 2010)
9. A Wille, M Broll, S Winter, Phase difference based RFID navigation for medical applications, in *RFID, 2011 IEEE International Conference on* (Orlando, FL, 2011)
10. C Hekimian-Williams, B Grant, L Xiuwen, Z Zhenghao, P Kuma, Accurate localization of RFID tags using phase difference, in *RFID, 2010 IEEE International Conference on* (Orlando, FL, 2010)
11. Z Chenming, JD Griffin, Accurate phase-based ranging measurements for backscatter RFID tags. Ant. Wireless Pro. Lett. IEEE **11**, 152–155 (2012). doi:10.1109/lawp.2012.2186110
12. E DiGiampaolo, F Martinelli, Mobile robot localization using the phase of passive UHF RFID signals. Indust. Electron. IEEE Trans. **61**(1), 365–376 (2014). doi:10.1109/tie.2013.2248333
13. M Yongtao, Z Liuji, L Kaihua, W Jinlong, Iterative phase reconstruction and weighted localization algorithm for indoor RFID-based localization in NLOS environment. Sens. J. IEEE **14**(2), 597–611 (2014). doi:10.1109/jsen.2013.2286220
14. C-H Ko, RFID 3D location sensing algorithms. Autom. Constr. **19**(5), 588–595 (2010). doi:10.1016/j.autcon.2010.02.003
15. MB Zeytinci, V Sari, FK Harmanci, E Anarim, M Akar, Location estimation using RSS measurements with unknown path loss exponents. EURASIP J. Wireless Commun. Netw. **2013**(1), 178 (2013). doi:10.1186/1687-1499-2013-178
16. Q Liang, Radar sensor wireless channel modeling in foliage environment: UWB versus narrowband. IEEE Sens. J. **11**(6), 1448–1457 (2011). doi:10.1109/JSEN.2010.2097586
17. D Liu, K Liu, Y Ma, J Yu, Joint TOA and DOA localization in indoor environment using virtual stations. Commun. Lett. IEEE **18**(8), 1423–1426 (2014). doi:10.1109/LCOMM.2014.2333006
18. Q Liang, Automatic target recognition using waveform diversity in radar sensor networks. Pattern Recognit. Lett. **29**(3), 377–381 (2008)
19. J Liang, Q Liang, Design and analysis of distributed radar sensor networks. Parallel Distribut. Syst. IEEE Trans. **22**(11), 1926–1933 (2011). doi:10.1109/TPDS.2011.45
20. Q Ren, Q Liang, Throughput and energy-efficiency-aware protocol for ultrawideband communication in wireless sensor networks: a cross-layer approach. IEEE Trans. Mobile Comput. **7**(6), 805–816 (2008). doi:10.1109/TMC.2007.70765
21. L Hui, H Darabi, P Banerjee, L Jing, Survey of wireless indoor positioning techniques and systems. Syst. Man. Cybern. Part C **37**(6), 1067–1080 (2007). doi:10.1109/tsmcc.2007.905750
22. LM Ni, L Yunhao, L Yiu Cho, AP Patil, LANDMARC: indoor location sensing using active RFID. Wireless Netw. **10**(6), 701–10 (2004). doi:10.1023/B:WINE.0000044029.06344.dd
23. LM Ni, Z Dian, MR Souryal, RFID-based localization and tracking technologies. Wireless Commun. IEEE **18**(2), 45–51 (2011). doi:10.1109/mwc.2011.5751295
24. Y Zhao, K Liu, Y Ma, L Zhou, J Wang, An effective phase-based localization approach under multipath environments, in *Proceedings of the second international conference on communications, signal processing, and systems* (Tianjin, China, 2013)
25. W Chong, H Wu, T Nian-Feng, RFID-based 3-D positioning schemes, in *INFOCOM 2007 26th IEEE International Conference on Computer Communications* (IEEE, Anchorage, AK, 2007)
26. Z Zhi, L Zhonghai, V Saakian, Q Xing, C Qiang, Z Li-Rong, Item-level indoor localization with passive UHF RFID based on tag interaction analysis. Indust. Electron IEEE Trans. **61**(4), 2122–2135 (2014). doi:10.1109/tie.2013.2264785
27. W Shi, K Liu, J Fang, P Luo, J Yu, X Huang, UHF RFID location algorithm based on dual frequency subcarriers amplitude modulation. J. Harbin Inst. Technol. **44**(3), 81–86 (2012)
28. Z Wang, X Huang, W Yang, The measuring phase method of all-phase FFT. World Sci. Tech. R D **29**(4), 28–32 (2007). doi:10.3969/j.issn.1006-6055.2007.04.006
29. D Arnitz, U Muehlmann, K Witrisal, Tag-based sensing and positioning in passive UHF RFID: tag reflection, in *3rd Int EURASIP workshop on RFID Technology* (Cartagena, Spain, 2010)
30. I Guvenc, S Gezici, F Watanabe, H Inamura, *Enhancements to linear least squares localization through reference selection and ML estimation* (Wireless Communications and Networking Conference, WCNC IEEE, Las Vegas, NV, USA, 2008). doi:10.1109/WCNC.2008.55
31. X Li, Y Zhang, MG Amin, *Multifrequency-based range estimation of RFID tags. RFID, 2009 IEEE International Conference on* (IEEE, Orlando, FL, USA, 2009). doi:10.1109/RFID.2009.4911199
32. H Xia, H Bertoni, L Maciel, A Lindsay-Stewart, R Rowe, Radio propagation characteristics for line-of-sight microcellular and personal communications. IEEE Trans. Ant. Pro. **41**(10), 1439–1447 (1993). doi:10.1109/8.247785