An Improved Fingerprint Algorithm Based on Wireless Sensor Networks

Long Cheng, Ze Liu and Liang Feng

Department of Computer and Communication Engineering, North-eastern University, Qinhuangdao 066004, Hebei Province, China
Email: chenglong8501@gmail.com

Abstract. With the wide application of positioning technology in real life, people have particularly become concerned about the improvement of the accuracy of positioning. One of the common methods to deal with such problems is wireless sensor networks (WSN). Reducing the non-line-of-sight (NLOS) error and optimizing the positioning accuracy are the main technical problem. In this paper, we propose an improved fingerprint algorithm to enhance the accuracy of positioning. The traditional k-Nearest Neighbor (KNN) algorithm has the problem of sample imbalance, which leads to the individual data directly determining the decision result. Our proposed algorithm can effectively solve the problem of sample imbalance. Simulation results and experimental results illustrate that our algorithm is superior to KNN algorithm.

1. Introduction
Node technology is one of the main supporting technologies of wireless sensor networks. Node positioning is also a precondition for many applications of sensor networks. Due to the availability of the message of the location, there is a clipping development in WSN [1]. Global Positioning System (GPS) which is a mainstream positioning system in the world. However, the research shows the poor precision in the indoor or the forest environment of GPS. The WSN gradually is applied to various fields such as environmental monitoring, network protocols, security management, owing to the feature of agility, easy maintenance, strong fault tolerance, and strong adaptability. The primary positioning method of the WSN can be divided into: time of arrival (TOA) [2], time difference of arrival (TDOA) [3], angle of arrival (AOA) [4], received signal strength (RSS) [5]. Introducing a Statistical Learning Theory method aims to cut down the complexity of the propagation path and raises a positioning technology based on Support Vector Machines, which illustrates the application for the fingerprint problem [6]. The data analysis of RSS in indoor environment is carried out, and the influence of user positioning, stability and time dependence. RSS distribution are discussed based on location fingerprinting [7]. It explains the impact of a series of parameters such as radio propagation features for the fingerprint algorithm based indoor location system. This analysis provides a guide for fingerprint location [8]. A relatively close access points (APs) can reduce the unbiased estimation of Cramer-Rao Lower Bound (CRLB), and some similar features are used to select the boosting nodes in the data sequence for fingerprint algorithm. It solves the problem of node selection [9]. The traditional network based on radio frequency (RF) fingerprinting relies on data processing, and the output results also change with the choice of grid size. The Cluster-based RF Fingerprinting (CRFF) is proposed, which uses clustering fingerprint positioning without data processing and that decreases the cost of calculation [10]. However, the propagation between the beacon nodes and the mobile nodes are blocked by obstacles leading to reflection or diffusion, which is termed as NLOS error.
The main contribution of this paper is given as follows:
1) We construct a Euclidean distance fingerprint map to calculate the offset errors. Through the partial detection, each element is filtered out in a set of data to be affected by NLOS errors.
2) Through the overall detection, the individual elements in a set of data are filtered out to be affected by NLOS errors.
3) The obtained reference points are mutually detected to achieve the final positioning.

In this paper, an improved fingerprint algorithm is proposed. The architecture of this paper is as follows. The improved algorithm of this paper is proposed in the second part. The simulation results are obtained in the third part. The conclusion is given in the fourth part.

2. Proposed Algorithm and Model Establishment

2.1. Model Establishment

A measure model is employed in this section. The adopted scheme is given as follows. We randomly set $N$ beacon nodes in a region which are denoted as: $z_i = [x_i, y_i]^T$, $i \in [1, N]$. The location of the obstacles is unknown. The mobile nodes at the time $k$ are regarded as: $X = [x^m_k, y^m_k]^T$. The true distance between the $i$-th beacon node and the $i$-th mobile node at time $k$ is

$$
\begin{align*}
    d^i_k &= \sqrt{(x^u_k - x_i)^2 + (y^u_k - y_i)^2} \\
\end{align*}
$$

Under line-of-sight (LOS) conditions, this distance at time $k$ can be modelled as:

$$
\begin{align*}
    r^i_k &= d^i_k + n_i
\end{align*}
$$

where $n_i$ is the measurement noise with a mean of zero and a variance of $\delta^i_n$.

Under NLOS conditions, the path leading to signal transmission is not a straight line due to the influence of obstacles. The distance at time $k$ can be modelled as:

$$
\begin{align*}
    r^i_k &= d^i_k + n_i + b_{NLOS}
\end{align*}
$$

where $n_i$ is the measurement noise with a mean of zero and a variance of $\delta^i_n$. $b_{NLOS}$ is considered as NLOS error subjecting to a mean of $u$ and a variance of $\delta^2_{pNLOS}$. And, $b_{NLOS}$ subordinates to the of Gaussian distribution, Uniform distribution or Exponential distribution. The parameters are different under different indoor environment and measurement methods.

The measurement noise of $n_i$ with the probability distribution function (pdf) is assumed as:

$$
\begin{align*}
    f(n_i) &= \frac{1}{\sqrt{2 \cdot \pi \delta^2_n}} \exp \left( -\frac{n_i^2}{2\delta^2_n} \right) \\
\end{align*}
$$

The NLOS error of pdf with the Gaussian distribution is given by

$$
\begin{align*}
    f(b_{NLOS}) &= \frac{1}{\sqrt{2 \cdot \pi \delta^2_p}} \exp \left( -\frac{(b_{NLOS} - u)^2}{2\delta^2_p} \right) \\
\end{align*}
$$

The NLOS error of pdf with the Uniform distribution is given by

$$
\begin{align*}
    f(b_{NLOS}) &= \begin{cases} 
        \frac{1}{u_{max} - u_{min}} & \text{if } u_{min} \leq b_{NLOS} \leq u_{max} \\
        0 & \text{else} 
    \end{cases}
\end{align*}
$$

The NLOS error of pdf with the Exponential distribution is given by
\[ f(b_{\text{NLOS}}) = \begin{cases} a^{-b_{\text{NLOS}}/b} & b_{\text{NLOS}} \geq 0 \\ 0 & \text{else} \end{cases} \]  

\[ A = \sqrt{(x_i - p_x^a)^2 + (y_i - p_y^a)^2} \]  

**Figure 1.** Flow chart of the algorithm

2.2. Building A Fingerprint Map

The specific flow chart is as shown in **Figure 1.** We randomly place reference points in the field. The state vector can be defined as \( T_{a,b} = [p_x^a, p_y^b] \), \( a \in (1,101), b \in (1,101) \). We mainly calculate the Euclidean distance between the beacon nodes and reference points to construct a Euclidean distance fingerprint map. The number of the Euclidean distance fingerprint map is prodigious, which is enough to judge the location information of mobile nodes. The Euclidean distance fingerprint map can be modelled as:

\[ E_i = \left| \sqrt{A} - A \right| \]  

2.3. Detection

- Partial detection: We default the threshold to 3, if the deviation of \( E_i \leq 3 \) for each node. The deviation which is considered to be acceptable, and can be used as a reference value. At the same time, if the number \( p \) of the deviation of each node in the group is \( p \geq 3 \), the data of this group is reserved. The above conditions are satisfied, the process proceeds to step 2.
- Overall detection method 1: In the Euclidean distance map, there are \( N \) offset error values in a set of data. The sum of these offset error values can be calculated as:

\[ \text{sum} = \sum_{i=1}^{N} E_i \]
Since we assume that the threshold of the offset error value is 3, the overall threshold of the \( N \) values is \( (3 \cdot N) \) in a group. Determining the sum and the total threshold, if sum is less, the process goes to step 4. However, the influence of the NLOS error results in the occurrence of individual outliers, which leads to the final abandonment. Because other values in the group are useful in reality, they can be used as reference information to judge mobile nodes. So, we perform the second screening.

- Overall detection method 2: For the second screening, the summation of the offset errors can be regarded as:

\[
ALL = \sum_{i=1}^{N} \sum_{j=1}^{10201} E_{i,j}
\]  

(11)

The mean value of overall offset errors can be given as \( \frac{ALL}{\text{len}} \), \( \text{len} \in [1,10201] \). \( \text{len} \) is the number of groups selected for screening. If it is greater than the sum, mobile node is too far away from the beacon node, which will bring in a great deviation for our final positioning. Collecting filtered data, and the process proceeds to step 4.

- Reference node mutual detection: The data collected in the above steps is counted and summarized. The distance between each node is measured, the measurement results are stored, and all the statistical measurements are summed to obtain a final value \( W \). The final value \( W \) divide by \( \text{len} \), and get the average of all the distances and the mean is recorded as \( M \). Then we compare the distances between each node with \( M \). If the points are smaller than the mean, the estimated value of the nodes are obtained after mutual detection, and the process proceeds to step 5.

- Positioning: Recording the position of the collected nodes in step 4, and the data obtained is used as an estimate of our final location. Because the more reference nodes and beacon nodes, the richer the known information we obtain, and the higher the positioning accuracy. Therefore, the method has a particularly high requirement on the amount of data, which results in a particularly large computational strength, this is also a place where the method needs to be improved in the subsequent research process.

3. Simulation Results

In this section, we compare the proposed algorithm with the classical KNN algorithm. We analyse the simulation results, detailed default parameters are visible in the Table 1. The manifestation of the proposed algorithm is estimated by the average localization error (ALE):

\[
ALE = \frac{1}{M \cdot t_n} \sum_{i=1}^{M} \sum_{k=1}^{t_n} \sqrt{(x_k^i - x_{i(k)})^2 + (y_k^i - y_{i(k)})^2}
\]  

(12)

where \( M=100 \) for the number of Monte Carlo runs, \( t_n=1000 \). The mobile node is set at time \( k \) is \( z_k^i = [x_k^i, y_k^i] \). The estimated value for the \( i \)-th test, and the time interval \( k \) is given as \( X_{i(k)} = (x_{i(k)}, y_{i(k)}) \).

| Parameters | Symbol | Default Values |
|------------|--------|----------------|
| The number of beacon nodes | \( N \) | 6 |
| The standard deviation of measurement noise | \( \delta_t \) | 1 |
| The NLOS error | \( U(u_{min}, u_{max}) \) | \( U(2,9) \) |
| The standard deviation of particle movement | \( \delta_p^2 \) | 6 |
| The number of sample points | \( T \) | 100 |
| The number of cycle times | \( t_n \) | 100 |
Figure 2. The localization results

Figure 2 shows the trajectory of KNN algorithm and our proposed algorithm. In this figure, we can see that our proposed algorithm is better than the KNN algorithm. Figure 3 shows the effect of different number of nodes on the average localization error. When the number of nodes is 6, the proposed algorithm improves the positioning accuracy by 0.4 meters. As the number of beacon nodes increases, the ALE of our proposed algorithm is lower than that of the KNN algorithm.

Figure 3. The number of beacon nodes versus ALE

Figure 4. The $u_{\text{max}}$ versus ALE

In Figure 4, when $u_{\text{max}} = 13$, the ALE of the two positioning algorithms reaches the maximum. The proposed algorithm has better robustness than the KNN algorithm.

4. Conclusion

In this paper, we propose an improved fingerprint algorithm, which can effectively prevent the sample imbalance comparing with the traditional KNN algorithm. We construct the Euclidean distance map as a distance database of the beacon nodes and reference points. Through the detection, filtering out node information that is greatly affected by NLOS error. The obtained reference points are mutually detected to achieve the final positioning. The simulation results show that the proposed algorithm is more accurate than the KNN algorithm and have better robustness.

5. References

[1] Zheng, J., Sun, Y. E., Huang, Y., Wang, Y., & Xiao, Y. (2012). Error analysis of range-based localisation algorithms in wireless sensor networks. International Journal of Sensor Networks, 12(2), 78-88.

[2] Guvenc, I., & Chong, C. C. (2009). A survey on toa based wireless localization and nlos mitigation techniques. IEEE Communications Surveys & Tutorials, 11(3), 107-124.

[3] Cheng, L., Wu, C. D., & Zhang, Y. Z. (2011). Indoor robot localization based on wireless sensor networks. IEEE Transactions on Consumer Electronics, 57(3), 1099-1104.

[4] Yang, S. L., Park, J. W., & Barolli, L. (2012). A localization algorithm based on aoa for ad-hoc sensor networks. Mobile Information Systems, 8(1), 61-72.
[5] I. Amundson, J. Sallai, X. Koutsoukos and A. Ledeczi, “Mobile Sensor Waypoint Navigation via RF-Based Angle of Arrival Localization,” International Journal of Distributed Sensor Networks, vol. 2012, pp. 1-15, 2012.

[6] Brunato, M., & Battiti, R. (2005). Statistical learning theory for location fingerprinting in wireless lans. Computer Networks, 47(6), 825-845.

[7] Kaemarungsi, K., & Krishnamurthy, P. (2004). Properties of Indoor Received Signal Strength for WLAN Location Fingerprinting. International Conference on Mobile & Ubiquitous Systems: Networking & Services.

[8] Kaemarungsi, K., & Krishnamurthy, P. (2004). Modeling of indoor positioning systems based on location fingerprinting. Joint Conference of the IEEE Computer & Communications Societies.

[9] Hossain, A. K. M. M., & Soh, W. S. (2010). Cramer-Rao Bound Analysis of Localization Using Signal Strength Difference as Location Fingerprint. Conference on Information Communications. IEEE Press.

[10] Mondal, R. U., Ristaniemi, T., & Turkka, J. (2017). Cluster-based rf fingerprint positioning using lte and wlan signal strengths. International Journal of Wireless Information Networks, 24(4), 413-423.