An Improved Outliers Detection Algorithm for Received Signal Strength Measurements in Building Environment

Yan ZHANG¹,a, Jun LU¹,b, Xian-Sheng GUO²,c*

¹School of Architecture and Civil Engineering, Chengdu University, Chengdu, China
²School of Engineering, University of Electronic Science and Technology of China, Chengdu, China

a39624001@qq.com, bckchina@qq.com, cxsquo@uestc.edu.cn
*Corresponding author

Keywords: Outliers Detection, Received Signal Strength, Hample Filter.

Abstract. In this paper, we propose a new algorithm to detect outliers for RSS measurements in building environment. RSS measurements fluctuate severely in complex building environment because multi-path, obstacles, and moving people, can distort the propagation of radio wave, which leads to some inevitable large values in RSS measurement, known as outliers. Some classical outliers detection algorithm fail to detect them when outliers rate is higher. Based on this observation, we proposed a modified Hample filter algorithm to detect outliers. Simulation results show that our proposed algorithm can overcome the fluctuation of RSS in building localization environment.

Introduction

In smart home system, it is a key issue to localize mobile users by using a wireless networks, such as wireless sensor networks (WSNs) [1], wireless local area networks (WLAN) [2], ZigBee[3], etc. Unlike outdoor localization environment, indoor localization environment is full of multi-path and non-line-of-sight (NLOS); additionally, indoor environment always changes as time varies, for example, the close/open of door or windows, moving people are all the factors to distort the wave propagation in real practice. Due to the complex environment of indoor localization, some localization techniques which works better in outdoor environment cannot work in indoor localization environment. So how to improve the accuracy of indoor localization has drawn more attention in last decades.

The existing indoor localization techniques can be divided into two types: one is parametric method, and the other is so-called non-parametric method. The former is to estimate the position of target by measuring some parameters, such as received signal strength (RSS)[4], time-of-arrival (TOA)[5], time-difference-of-arrival (TDOA)[6], and angle-of-arrival (AOA)[7], or their combinations. By transforming these parameters into the distances, one can localize the interested targets by using the so-called trigonometry. The main drawback of these techniques is that the transforming between the measured parameters and distances has big bias due to the model parameters which is used to estimate distances are not precise. And it is always difficult to predict these parameters, for example, path loss factor. So the parametric method shows low accuracy in complex indoor localization environment.

The latter is also called as fingerprint based methods, which is to estimate the locations of targets by using pre-stored fingerprints. This technique does not estimate distances and can obtain higher accuracy than the former.

Most of existing methods of the latter are using RSS as the fingerprints because many mobile devices can measure RSS easily. However, there are some big fluctuations in RSS measurements due to multi-path, NLOS or environment changing. The values of RSS who have large error can be treated as outliers, which degrades the localization performance drastically. So how to detect these outliers in RSS and recover them is a key issue to obtain a higher accuracy localization estimator. In statistics, an outlier is an observation point that is distant from other observations. There is no rigid
mathematical definition of what constitutes an outlier; determining whether or not an observation is 
an outlier is ultimately a subjective exercise. There are various methods of outlier detection. The two 
main algorithms are Hampel filter [8] and 3σ algorithm [9]. In ZigBee multiple channels RSS 
measurements, because the correlation between adjacent channels, so the fingerprint matrix shows 
many same entries in adjacent channels, which degenerate the performance of these classical outliers 
detection algorithms.

In this paper, we propose an improved Hampel filter algorithm to detect the outliers in RSS 
measurements of ZigBee multiple channel measurements. The proposed algorithm can obtain a 
higher localization accuracy in complex indoor environment. Simulation results show the efficacy 
and feasible of our proposed algorithm.

**Problem Formulation**

The indoor path-loss model of \( N \) ZigBee reference nodes can be expressed as

\[
P_i = P_0 - 10\beta \log_{10} \frac{d_i}{d_0} + n_i, 1 \leq i \leq N,
\]

(1)

Where \( P_i, P_0 \) are the received signal power from the distance from \( d_i \) and \( d_0 \), respectively, \( n_i \) is 
model noise, and \( \beta \) is path-loss factor.

Assume that we have \( M \) grids in an indoor scenario, \( C \) ZigBee channels are used to transmit 
signals and \( K \) samples are collected. The RSS fingerprints of the \( i \) th grid can be denoted as

\[
P_i = [p_{i,1}, p_{i,2}, \ldots, p_{i,N}],
\]

(2)

Where

\[
P_{i,j} = \begin{pmatrix}
R_{i,j}^{1,1} & \cdots & R_{i,j}^{1,C} \\
\vdots & \ddots & \vdots \\
R_{i,j}^{K,1} & \cdots & R_{i,j}^{K,C}
\end{pmatrix}
\]

(3)

In which \( R_{m,n}^{i,j} (1 \leq m \leq K, 1 \leq n \leq C) \) is the \( m \) th measurement of the \( n \) th channel received signals at 
the \( i \) th grid from the \( j \) th reference node. All the fingerprints can be denoted as \([P_1, P_2, \ldots, P_M]\). In 
ZigBee multiple channels RSS fingerprints, because the correlation between adjacent channels, so 
the fingerprint matrix shows many same entries in adjacent channels, which degenerate the 
performance of some classical outliers detection algorithms, such as Hampel filter [8] and 
3σ algorithm [9].

**Proposed Algorithm**

As compared with 3σ algorithm, Hampel filters are known to perform better for outlier detection 
than median filters. This is because, other than median filters, Hampel filters also take the distribution 
of data into account. While median filters often work with a fixed threshold for outlier detection, 
Hampel filters use a scaling factor which takes the distribution of the data into account. In other 
words: If all recorded values are really close to the median, the Hampel filter adjusts to find the values 
that are not so close to the median. When later in the process recorded values are generally farther 
away from the median, the Hampel filter automatically adjusts to this and identifies outliers as points
that are even farther away from the median. With usual Median filters with a fixed threshold this behavior is difficult to achieve.

However, the biggest drawback of the Hampel filter is that the algorithm will take all elements of data to be outliers if over 50% entries are the same. In order to improve the performance of Hampel filter, we propose an improved Hampel filter algorithm, which combines kernel density efficiently as follows.

Given data set \( G = \{ g_1, g_2, \cdots, g_N \} \), its probability density function (pdf) can be estimated by

\[
\hat{f}(g) = \frac{1}{|G|} \sum_{g \in G} k(g - g_i)
\]

Where \( |G| \) is the element number of \( G \), kernel function has the following form

\[
k(g) = \begin{cases} 
\frac{3}{4B} \left( 1 - \left( \frac{g}{B} \right)^2 \right), & \left| \frac{g}{B} \right| < 1 \\
0, & \text{other}
\end{cases}
\]

Where \( B \) is the bandwidth of kernel function \( k(g) \). We can define a median value of absolute bias level as

\[
m_i = \frac{|g_i - \text{median}(G)|}{R}, (1 \leq i \leq N)
\]

Where \( \text{median} (G) \) is the median value of \( G \), and \( R = 1.4826 \times \text{median} \left( \sum_{g \in G} |g_i - \text{median}(G)| \right) \) is the absolute median bias. For each entries in \( G \), we can compute the confidence level by using the following equation

\[
c_i = \frac{\text{Prob}(g_i)}{m_i}, (1 \leq i \leq N)
\]

where \( \text{Prob}(g_i) \) can be calculated by using (4). From the above observation, we find that the \( i \) th entries in \( G \) can be regarded as a credible one for the bigger \( c_i \) and smaller \( m_i \). In contrast, we take it as an incredible one for the smaller \( c_i \) and bigger \( m_i \). So we can select a threshold to test if it is an outliers. We now summarize our proposed improved outliers detection algorithm in Table 1.

### Simulation Results

The RSS data is generated by using the path loss model (1) with Gaussian noise variance 1, and after generating data, we normalized our generated RSS so that the elements of RSS matrix follow \( \mathbb{N}(0,5) \). The dimension of our generated RSS matrix is 50X20. We random generate outliers in RSS matrix with outliers probability from 0.01 to 0.5 with interval 0.01. Under each outliers probability, we generate outliers which is uniformly distributed form 50 to 100 and add them to RSS matrix. Two parameters, namely, missed alarm probability (MAP) and false alarm probability (FAP), are adopted to evaluate the performance of our proposed algorithm. They are defined as
Table 1. An Improved Outliers Detection Algorithm based on Hample Filter.

<table>
<thead>
<tr>
<th>Input: Data set $G = {g_1, g_2, \ldots, g_N}$, $\text{index} = \emptyset$, threshold $\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Calculate the median value $\mu^* = \text{median}(G)$ in $G$;</td>
</tr>
<tr>
<td>for $i=1:N$</td>
</tr>
<tr>
<td>2) Calculate the absolute median bias $R = 1.4826 \times \text{median}_{g_i \in G} \left{</td>
</tr>
<tr>
<td>3) Calculate the median value of absolute bias level $m_i = \frac{</td>
</tr>
<tr>
<td>4) Construct the probability density function $\hat{f}(g) = \frac{1}{</td>
</tr>
<tr>
<td>5) Get the PDF of each entry by using $\text{Prob}(g_i) = \hat{f}(g_i)$;</td>
</tr>
<tr>
<td>6) Compute the confidence level $c_i = \frac{\text{Prob}(g_i)}{m_i}$;</td>
</tr>
<tr>
<td>if $c_i &lt; \epsilon$, $g_i$ is an outlier, $\text{index} = \text{index} \cup i$;</td>
</tr>
<tr>
<td>else $\text{index} = \text{index} \cup \emptyset$;</td>
</tr>
<tr>
<td>end for</td>
</tr>
</tbody>
</table>

Output: $\text{index}$

Figure 1. The MAP Curves versus Different Outliers Rate.

\[
\begin{align*}
\text{MAP} & = 1 - \frac{n_2}{n_1} \\
\text{FAP} & = -\text{MAP}
\end{align*}
\]

(8)

Where, $n_1$ is the total number of outliers in RSS matrix, and $n_2$ is the detected number of outliers.

Figure 2. The FAP Curves versus Different Outliers Rate.
From Fig. 1, we find that our proposed algorithm has the lowest MAP as the outliers rate increase. As compared with our proposed algorithm, Hampel filter and $3\sigma$ algorithms fail to detect outliers when outliers rate equals to one. Our algorithm shows more robustness when outliers rate is higher.

Fig. 2 is the FAP curves versus different outliers rate. From this figure, it is seen that the FAP tendency of our proposed algorithm decreases as the outliers rate increases. As compared with our proposed algorithm, the rest two algorithm, namely, $3\sigma$ and Hampel filter algorithms, both fail to detect outliers entirely, so the FAP is nearly zeros. Hence, our proposed algorithm can work well in RSS fingerprint outliers detection.

**Conclusion**

This paper proposes an improved outliers detection algorithm, which is used to detect outliers in indoor localization environment. Because the complex environment of indoor localization, outliers are always generated due to the existence of multi-path, obstacles, and moving people, etc. Some classical outliers detection algorithms fail to detect them when outliers rate is higher. Based on this observation, we proposed a modified Hampel filter algorithm to detect outliers. Simulation results show that our proposed algorithm can overcome the fluctuation of RSS in building localization environment. Simulation results show the efficacy and feasibility of our proposed algorithm.

**Acknowledgement**

This research was financially supported by National Natural Science Foundation of China (No. 61201277).

**References**


