Data Process for Indoor Positioning based on WiFi Fingerprint

Xuerong Cui[1*], Mengyan Wang[*], Juan Li[*], Meiqi Ji, Jianhang Liu, Tingpei Huang, Haihua Chen

cuixuerong@163.com, 2426326907@qq.com,lijuanlijuan@sina.com, 506353930@qq.com, liujianhang@upc.edu.cn, huangtingpei@upc.edu.cn, chenhaihua@upc.edu.cn
Department of Computer and Communication Engineering, China University of Petroleum (East China), Qingdao, China

Abstract. Currently, most of the existing location fingerprint indoor positioning algorithms are based on the original fingerprint database. The accuracy of the fingerprint database will directly affect the final positioning accuracy. A method based on skewness-kurtosis normality test and Kalman filter fusion is proposed in this paper. Experiments shows that the fusion algorithm can effectively remove the abrupt data and noise fluctuations for the RSSI (Received Signal Strength Indication) data, and achieve accurate and smooth output of the RSSI value.

Keywords: Normal Distribution, Kalman Filter, Received Signal Strength Indication, Fingerprint Database

1. INTRODUCTION

In recent years, with the rapid development of technology, and the continuous improvement of people's quality of life, the market for location services has developed rapidly, and the demand for location services has also shown a rapid development trend. Location-based services LBS has received widespread attention. For positioning in outdoor environments, the Global Navigation Satellite System (GNSS) [1]provides a very accurate positioning, is a mature positioning technology, and is widely used. However its indoor positioning is not ideal. Therefore, it can only rely on wireless technologies such as infrared, WiFi, ultrasound, RFID to achieve indoor positioning[2].

In the indoor environment where people live, the wifi signal is easy to receive, but the wifi signal is easily affected by the outside world during the process of propagating[3,4]. For example, walls, ground, human body, temperature, humidity etc, and they will cause the wifi signal to reflect the scatter and other effects during the process of propagating, so that the RSSI signal of receiving is highly time-varying at a fixed position in the room, and the data needs to be pre-processed[5].

2. ORIGINAL DATA PROCESSING

The greater the distribution density of the RSSI value at a certain location, the closer the measured RSSI value is to the true value. After statistical analysis of the received signal strength, it is not difficult to see that not all of samples are from a normal distribution. For this case, if it uses the normal distribution function to estimate the probability density of the population directly, it is not accurate. In this paper, firstly the received signal strength is checked by the skewness and kurtosis. If it is a normal distribution, the normal distribution function is used to estimate the probability density of the population. Otherwise, the kernel function is used to calculate the overall probability. Finally, the received signal strength value of the large probability density is left, and the others are filtered to retain the large probability data. However, the filtrating of data does not eliminate some fluctuations of data, it only filters out some of the data with relatively large errors. In order to smooth the RSSI data, kalman filtering algorithm is also used in this paper, After the Kalman filter is performed on the retained large probability data, and the mean value is the determined RSSI value.

2.1 Skewness - Kurtosis Normal Dictribution Check

* Corresponding author.

MOBIMEDIA 2019, June 29-30, Weihai, People's Republic of China

DOI 10.4108/eai.29-6-2019.2282793

There are many ways to check that whether a sample obeys a normal distribution. Here, the skewness-kurtness test[6] method that does not have strict requirements on sample size, and it is used flexible. Let $X_1, X_2, \ldots, X_n$ denote the sample from the population, $\overline{X}$ denote the mean of sample, and $m_i = \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X})^i$ denote the order center distance of sample. The skewness and kurtosis of the normal distribution are both 0, there in the skewness and kurtosis are respectively:

$$b_s = \frac{E(X-EX)^3}{[Var(X)^{3/2}]} \quad , \tag{1}$$

$$k = \frac{E(X-EX)^4}{[Var(X)^2]} - 3 \quad . \tag{2}$$

The condition for using the skewness-kurtosis normality test is that the population has a priori information that deviates from the normal state in both the skewness and the kurtosis direction[7]. When the random variable obeys the normal distribution, the skewne $b_s = 0$ and the kurtosis $k = 0$ [8].

Propose assumption: $H0$:obbeying normal distribution, $H1$:Disobeying normal distribution. $H0$ is the original hypothesis, and its condition: $H_0 : b_s = 0, k = 0$,and H_1 is the alternative hypothesis,and its condition: $H_1 : b_s \neq 0, k \neq 0$ .Firstly calculate the statistics

$$T = \frac{\hat{b}_s^{\,2}(n+1)(n+3)}{6(n+2)} + \frac{(\hat{k} + \frac{6}{n+1})^2 (n+1)^2 (n+3)(n+5)}{24n(n-2)(n-3)} \quad . \tag{3}$$

Then according to the limit distribution freedom of statistic is χ^2 distribution of 2, So the rejection domain whose level is α test is $\{T > \chi_{1-a}{}^2(2)\}$, therein, $x_{1-a}{}^2(2)$ is the $1-\alpha$ quantile of the χ^2 distribution with a degree of freedom of 2.

This paper statistically analyzes the overall distribution of RSSI samples in indoor environment. A total of forty RSSI samples are selected to analyze the skewness and kurtosis values. It can be clearly found that most of the values do not satisfy the normal distribution, as shown in Fig.1,Fig.2 shows:

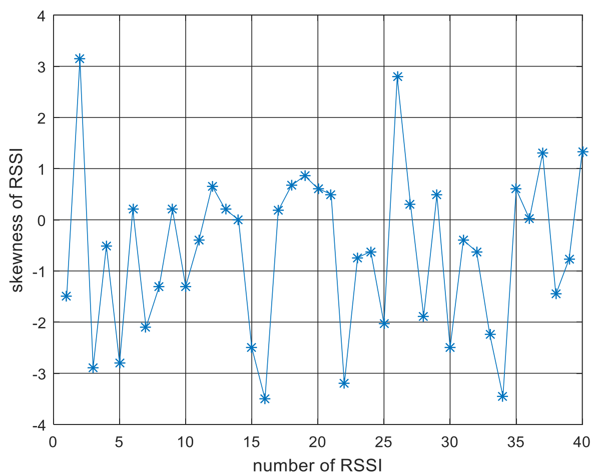

Fig. 1. The skewness of RSSI

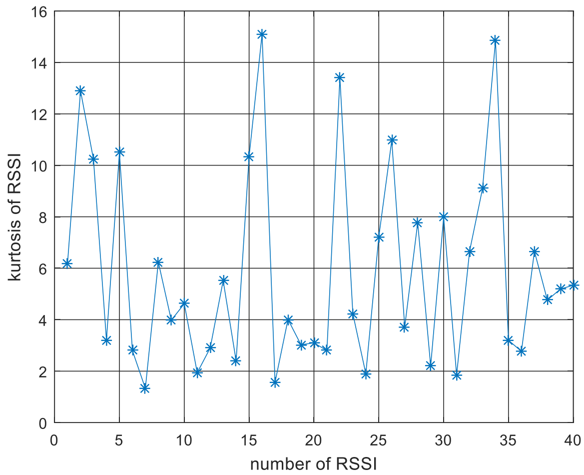

Fig. 2. The kurtosis of RSSI

2.2 The model of normal distribution

The RSSI sample collected in the offline phase is tested as described above to determine whether the rejection domain of H_0 is satisfied. If it is not satisfied, H_0 is accepted. It is considered that the population of the sample obeys a normal distribution when the significant level is $\alpha(0 < \alpha < 1)$. At this time, the probability density of the sample is approximately:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)}{(2\sigma^2)}} , \tag{4}$$

where $\mu = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[X_i - \mu]^2}$ are the parameter of normally distributed. X_i represents the i^{th} RSSI value in the sample RSSI and n is the capacity of RSSI sample.

2.3 Kernel Density Function

If the total of RSSI signal sample collected in the offline phase rejects the hypothesis of H_0, it is considered that the overall distribution of the sample is significantly different from the normal distribution when the significant level is α, it means that the sample population does not conform to the normal distribution. In this paper, the method of kernel density function is used to estimate the overall distribution of these samples. The kernel density estimation is a method for estimating the overall probability density in the case where the overall distribution is unknown[9].

This article chooses the kernel function as

$$k(y) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) . \tag{5}$$

The kernel function is used to estimate the probability density function of the population as

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n} \exp\left[-(X_i - x)^2 / 2h^2\right]. \tag{6}$$

2.4 Kalman Filter

The basic idea of Kalman filtering is to describe the linear system by using state equations and measurement equations. The optimal estimate of the state vector is obtained by processing the measured data[10], The Kalman filter algorithm is divided into two stages: prediction and correction.

Forecast stage:

$$x(k) = \Phi x(k-1) + w(k-1) , \tag{7}$$

$$\begin{aligned} P(k \mid k-1) = \Phi(k \mid k-1) P(k-1) \cdot \Phi^T(k \mid k-1) + \\ \Gamma(k \mid K-1) Q(k-1) \Gamma^T(k \mid k-1) \end{aligned} . \tag{8}$$

Revision stage:

$$\hat{x}(k) = \Phi(k \mid k-1)\hat{x}(k-1) + K_k[z(k) - H(k)\Phi(k \mid k-1)\hat{x}(k-1)], \quad (9)$$

$$P(k) = [I - K_k H(k)]P(k \mid k-1), \quad (10)$$

$$K_k = P(k \mid k-1)H^T(k)[H(k)P(k \mid k-1)H^T(k) + R(k)]^{-1}, \quad (11)$$

where $x(k)$ is the state value; $z(k)$ is the measured value; Φ is the state transition matrix; $w(k-1)$ is the state noise; $\Gamma(k \mid K-1)$ is the system noise matrix; I is the unit matrix; K_k is the Kalman gain matrix; $R(k)$ is the system observation noise variance matrix.

The effect of the signal strength value processed by Kalman filter is shown in Fig. 3.

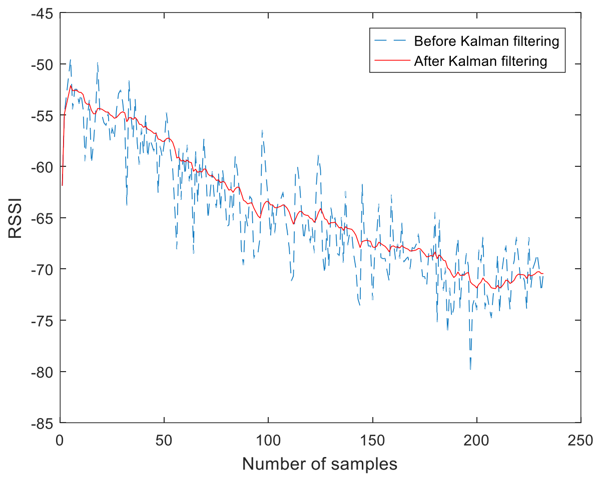

Fig.3. Comparison chart after using Kalman filter

The dotted line in the figure represents the data without Kalman filtering. It can be seen that the signal drift is more serious, which affects the positioning accuracy seriously; the solid line is the effect of Kalman filter processing. It can be seen that the Kalman filter can well process the signal strength values obtained by scanning, and ensure that the RSSI value is close to the actual value.

3. EXPERIMENTAL RESULTS

The training set data used in the experiment is a data set published by Paolo Barsocchi, Antonino Crivello, Davide La Rosa, Filippo Palumbo et al.

During the acquisition, the smartphone was kept at the chest level with the screen facing up. Every time the user was on a predefined location, the device recorded the following additional data concerning the detected Wi-Fi access points (APs)[11]:

• WiFi network name;

• AP MAC address;

• AP Received Signal Strength Indication (RSSI) expressed in dB

The flowchart of the fusion algorithm is shown in Fig.4.

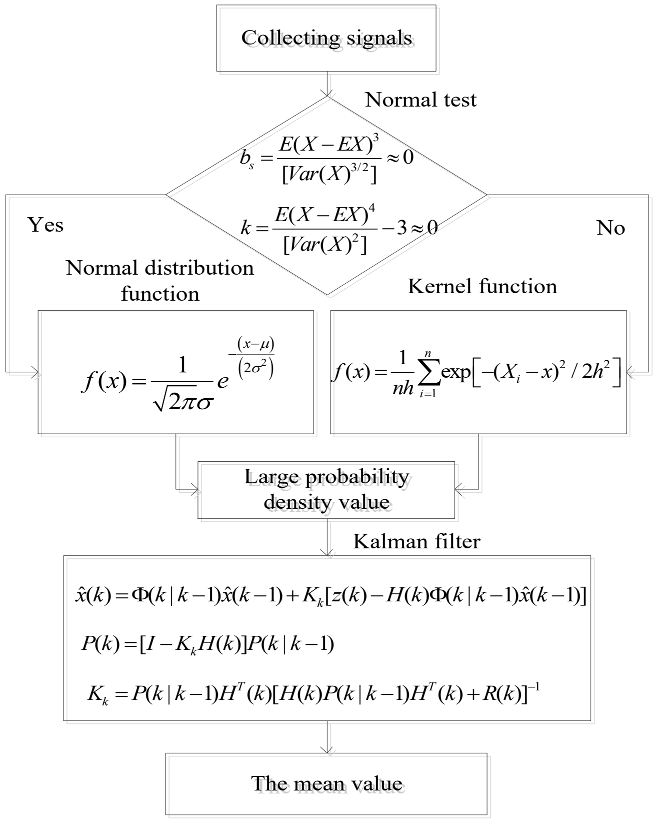

Fig.4. The flowchart of the fusion algorithm

Using the above data set to verify the effectiveness of the preprocessing mechanism based on skewness-kurtosis normality test is fusion of Kalman filter. The sampling nodes are randomly selected from the AP transmitting nodes from 1m to 30m, and 30 RSSI measurements are randomly selected. The fusion algorithm, normality test and Kalman filter are used to process the collected RSSI signals. The curve of the signal sample is shown in Fig.5.

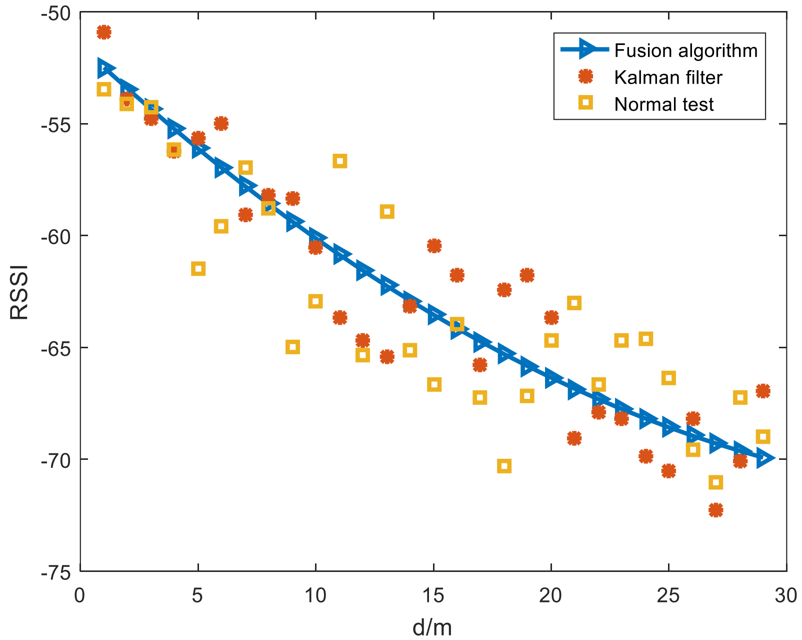

Fig.5 Distance and signal sample curve

The experimental results show that the fusion algorithm based on the skewness-kurtosis normality test and the Kalman filter is the closest to the ideal value. The fusion algorithm can not only eliminate the value of high error, but also can smooth the noise smoothly. The smoothed RSSI value has little delay.

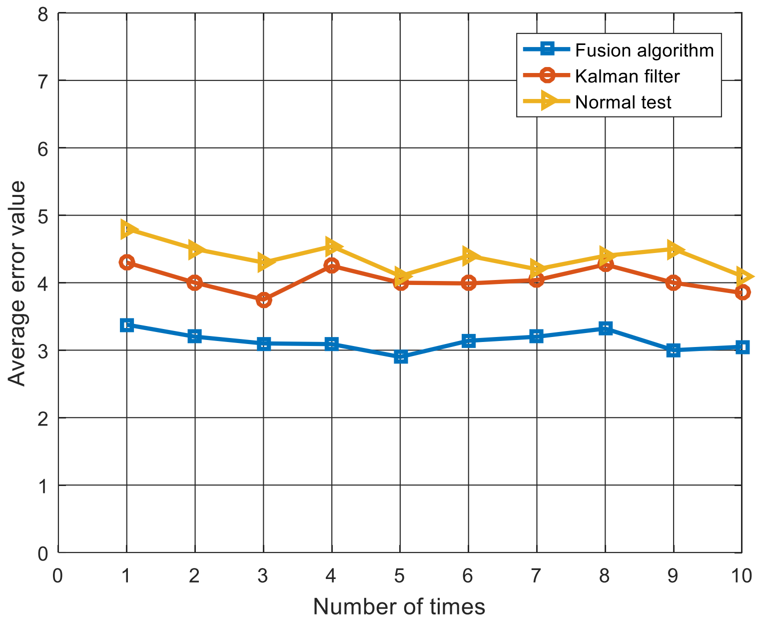

Fig.6 Three kinds of filtering algorithm error comparison chart

Figure 6 is a comparison of average error curves after positioning 59 coordinate positions by three algorithms. As can be seen from the figure, the fusion filtering algorithm proposed in this paper is the smallest error. At the same time, the sum of the positioning errors of 59 coordinate positions in the proposed algorithm is also the smallest of the three algorithms. Experimental results show that the fusion method has higher precision than that.

4. CONCLUSION

The complexity of the indoor environment makes it impossible to estimate the distribution of the measured RSSI samples by a function. Therefore, one of the keys to effectively improve the positioning accuracy is to accurately estimate the distribution of the RSSI sample population. The fusion algorithm based on skewness-kurtosis normality test and Kalman filter can effectively eliminate the singular values in the acquired RSSI signal and eliminate the fluctuation of the data, thus making the position estimation more accurate. The experimental results show that the proposed algorithm is closer to the real value than the traditional algorithm.

Acknowledgments. The authors would like to thank National Natural Science Foundation of China (No. 61671482, 61872385 and 61601519) for support in this work. Also, the authors would express the gratitude to Fundamental Research Funds for the Central Universities (No. 17CX02042A and 18CX02136A) in this report.

References

[1] HAN G, XU H, DUONG T Q, et al. Localization algorithms of Wireless Sensor Networks: a survey [J]. Telecommunication Systems, 2013, 52(4): 2419-36.

[2] CHEN S, QIAO X, MAO J, et al. Indoor secondary positioning method based on received signal strength [J]. Journal of Sensing Technology, 2015, 28(04): 572-7.

[3] FANG S, LIN T. A dynamic system approach for radio location fingerprinting in wireless local area networks [J]. IEEE Transactions on Communications, 2010, 58(4): 1020-5.

[4] PEI L, LIU J, CHEN Y, et al. Evaluation of fingerprinting-based WiFi indoor localization coexisted with Bluetooth [J]. The Journal of Global Positioning Systems, 2017, 15(1): 3.

[5] CHEN X, CHEN X, ZOU S. Indoor location algorithm based on normality test [J]. 2017, 38(03): 41-5.

[6] WANG Y, WEN L, CHEN M. Mathematical dictionary [M]. Science Press. 2010.

[7] MEI C, FAN J. Data analysis method [M]. Beijing: Higher education press, 2006.

[8] H K P, B V. Nonparametric Statistics with Applications to Science and Engineering [M]. 2007.

[9] Q L Y, W Z L, Q W, et al. Nonparametric Statistics Methods [M]. Chendu:Southwest Jiaotong University Press. 2010.

[10] HUANG Y, QI B, CHEN S. Modification of advanced boundary discriminative noise detection algorithm; proceedings of the 2013 10th IEEE International Conference on Control and Automation (ICCA), F 12-14 June 2013, 2013.

[11] BARSOCCHI P, CRIVELLO A, ROSA D L, et al. A multisource and multivariate dataset for indoor localization methods based on WLAN and geo-magnetic field fingerprinting; proceedings of the 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), F 4-7 Oct. 2016, 2016.