Scattering Clustering Method for Terminal Fingerprint Positioning Based on MIMO Base Station

Fei TANG, Yijie REN, Xiaojun WANG [1], Weiguang SUN and Xiaoshu CHEN

School of Information Science and Engineering, National Mobile Communications Research Laboratory, Southeast University, China

Abstract. With the rapid development of wireless communication technology, location-based services are playing an increasingly important role in people's lives. However, as the living environment becomes more and more complex, the existence of obstructions and various scatterers makes the accuracy of traditional positioning algorithms decrease, thus, fingerprint positioning has gradually become a research hotspot in the field of positioning.

This paper researches the 5th Generation (5G) fingerprint location method based on machine learning. A massive multiple-in multiple-out (MIMO) channel is constructed on the MATLAB simulation platform, from which the fingerprint information is extracted to establish a fingerprint database. Considering the huge amount of data in the fingerprint database, and under the multipath effect, the channel characteristics are mainly affected by the scatterers near the point to be located. This paper proposes a scattering-based clustering method that combines the particularity of multipath propagation for clustering. Research shows that this method has excellent clustering effects, which can effectively improve algorithm efficiency and reduce data storage pressure on the base station side.[2]

Keywords. Massive MIMO, Fingerprint extraction, Clustering

1. Introduction

In recent years, with the development of wireless communications, location-based services have attracted widespread attention in various industries. In outdoor scenes, especially in open areas, traditional positioning algorithms such as Global Positioning System (GPS) [1] have sufficient positioning accuracy. However, in some urban areas with tall buildings, or inside buildings, due to the influence of occlusion and non-line-of-sight transmission, the positioning effect of the traditional positioning algorithm is very limited [2]. In addition, in modern society, most of people's activities are concentrated indoors. New user needs make traditional positioning algorithms face new challenges, and fingerprint positioning technology emerges at the historic moment.

Different from the traditional positioning algorithms, the fingerprint positioning algorithm uses some corresponding characteristics generated during the signal

---

transmission to achieve positioning. In the process of wireless signal transmission, due to the influence of multipath effects, the channel characteristics at different physical locations will reflect some unique characteristics, which called the fingerprints of the signal. When there is a positioning requirement, the fingerprint information of the point to be located is compared with the fingerprint information in the database, and the information of the point is estimated according to a certain similarity criterion. Through the above-mentioned series of processing, the fingerprint positioning algorithm transforms a positioning problem into a pattern recognition problem. Literature [3-4] use channel state information to achieve reliable fingerprint positioning, and literature [5-6] combine the 5th Generation (5G) massive multiple-in multiple-out (MIMO) scenarios to explore the application prospects of fingerprint positioning in the new context. The flow chart of the fingerprint location algorithm is shown in Figure.1 [7-8].
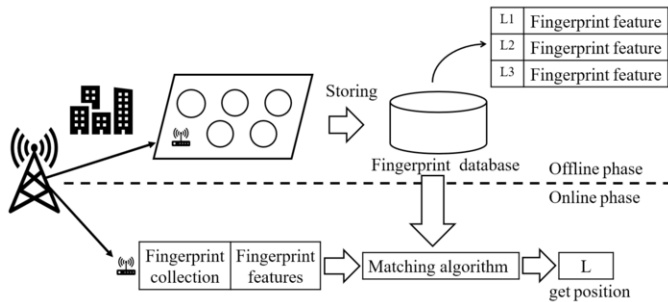

Figure 1. The flow chart of the fingerprint location algorithm.

This paper builds a massive multiple-in multiple-out-Orthogonal Frequency Division Multiplexing (MIMO-OFDM) system based on a simulation platform, and uses a scattering-based clustering method to complete fingerprint location. The main work is as follows:

- STEP 1: Build a massive MIMO-OFDM system, extract information from the channel and perform further processing to obtain fingerprints. The fingerprints make full use of the multipath effect and are easy to extract;

- STEP 2: Propose a clustering method based on scattering, which can not only improve the efficiency of the algorithm, but also ensure the positioning accuracy of the algorithm;

- STEP 3: Based on the similarity criterion of the Joint Angle Delay Similarity Coefficient (JADSC), the Weighted K Nearest Neighbor (WKNN) algorithm is used to improve the positioning accuracy of the algorithm.

2. Massive MIMO Channel Model

In the massive MIMO-OFDM single station system, the channel is constructed based on the ray tracing model, as shown in Figure.2.
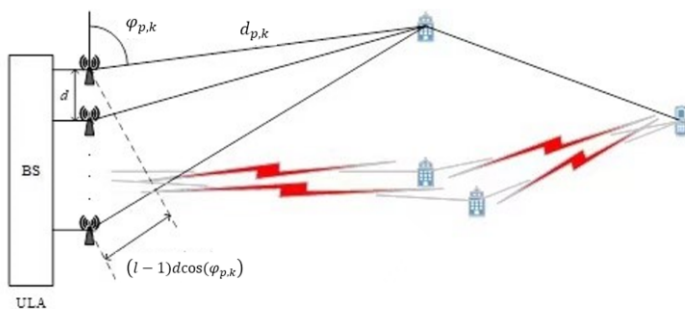
Figure 2. Massive MIMO-OFDM single station system.

The base station is equipped with N_t antennas in the form of Uniform Linear Array (ULA), the distance between each antenna element is d, the number of users is K, and its distribution law obeys random distribution, which satisfies $N_t \gg K$. Due to the influence of the scatterer, the signal will propagate in the channel in the form of multipath, and the base station estimates the information from the user k to the base station through the uplink channel in the communication link. In summary, the Channel Impulse Response (CIR) of user k on the p-th path can be expressed as:

$$\boldsymbol{q}_{p,k} = a_{p,k} \boldsymbol{e}(\varphi_{p,k}) \exp\left(-j2\pi \frac{d_{p,k}}{\lambda_c}\right) \tag{1}$$

Notation: $a_{p,k} \sim \mathcal{CN}(0, \delta_{p,k})$ represents the complex channel gain of the k-th user on the p-th path, $\varphi_{p,k} \in (0, \pi)$ represents the angle of arrival of the signal, $d_{p,k}$ represents the physical distance from the antenna element of the transmitting signal on the p-th path to the first receiving antenna element of the base station, λ_c represents the carrier wavelength, $\boldsymbol{e}(\varphi_{p,k})$ represents the corresponding array response vector, as shown in the following formula:

$$\boldsymbol{e}(\varphi_{p,k}) = \left[1, \exp\left(-j2\pi \frac{d\cos(\varphi_{p,k})}{\lambda_c}\right), \ldots, \exp\left(-j2\pi \frac{(N_t-1)d\cos(\varphi_{p,k})}{\lambda_c}\right)\right]^{\mathrm{T}} \tag{2}$$

Define a discrete Fourier phase shift matrix $V \in \mathbb{C}^{N_t \times N_t}$, which satisfies the following formula:

$$[\boldsymbol{V}_{i,j}] \triangleq \frac{1}{\sqrt{N_t}} \exp\left(-j2\pi \frac{i(j - N_t/2)}{N_t}\right) \tag{3}$$

The phase shift matrix can map the CIR from the antenna domain to the corresponding angle domain. Considering that the Time of Arrival (TOA) of each path in multipath transmission is different, the total CIR of user k can be expressed as the sum of the CIR of all paths:

$$q_k(\tau) = \sum_{p=1}^{P} q_{p,k} \cdot \delta(\tau - \tau_{p,k}) \tag{4}$$

In the formula, $\tau_{p,k} = d_{p,k}/v$ represents the TOA of each path in the multipath, and v represents the speed of light. Sampling the antenna array output based on the sampling interval T_s, then the symbol interval of OFDM can be expressed as $T_c = N_c T_s$.

Through the modulation of the OFDM system, the frequency selective fading channel caused by the multipath effect can be transformed into a frequency flat channel, and then time domain sampling is performed to obtain the TOA of each path. According to the mathematical relationship between channel responses, the Channel Frequency Response (CFR) of the l-th subcarrier can be expressed as the Fourier transform of its CIR:

$$\begin{aligned}
h_{k,l} &= \sum_{p=1}^{P} \sum_{n=0}^{N_g-1} a_{p,k} e(\varphi_{p,k}) \exp(-j2\pi f_l \tau_{p,k}) \cdot \delta(nT_s - \tau_{p,k}) \\
&= \sum_{p=1}^{P} a_{p,k} e(\varphi_{p,k}) \exp\left(-j2\pi \frac{l \cdot n_{p,k}}{N_c}\right)
\end{aligned} \tag{5}$$

In the formula, $n_{p,k}$ represents the propagation delay of the p-th path in multipath propagation, and $\lfloor x \rceil$ represents the integer closest to x. The overall CFR matrix of user k is composed of the corresponding CFRs of all its subcarriers, as shown in the following formula:

$$H_k = \left[h_{k,0}, h_{k,1}, \ldots, h_{k,L-1} \right] \tag{6}$$

3. Clustering Algorithm and Terminal Positioning

In fingerprint positioning, because the fingerprint database is generally large, traversing the entire fingerprint database will consume a lot of resources. Therefore, before online matching, clustering algorithms are used to classify reference points. Considering that fingerprint features are mainly affected by adjacent scatterers, this chapter proposes a scattering-based clustering method to improve the clustering effect, and finally uses WKNN in the terminal positioning part to improve the accuracy of the algorithm.

3.1 Clustering method based on scattering

As one of the classic machine learning, clustering algorithm mainly includes K-means and fuzzy C-means method, among which the fuzzy C-means method belongs to the fuzzy clustering algorithm. It can be seen from the above that the fingerprint feature at a certain location is mainly determined by the nearest scatterers around it. Taking into account the complexity of the multipath scene, the number of scatterers that affect the reference point can be correlated with its physical location, and the number of scatterers affected by the signal at a certain location is used as a standard to classify the fingerprint data.

The **scattering-based clustering algorithm** refers to the fuzzy clustering algorithm. The same reference point can belong to different categories. The membership degree in the fuzzy clustering algorithm is the number of scatterers that will affect the signal here,

which is called m_c. Record the membership degree of each reference point, establish the mapping between the membership degree and the physical location of the reference point, and set the cluster center according to a certain step:

$$O_i = O_0 + i * s_{CBS} \tag{7}$$

Notation: O_i is the i-th cluster center, O_0 is the initial cluster center, which satisfies $O_0 = \min(m_c) + s_{CBS}$, and s_{CBS} is the step size of the cluster center.

According to the recorded membership degree, set a certain numerical diffusion for the membership degree center to perform clustering:

$$m_C = [m_c - \Delta m, m_c + \Delta m] \tag{8}$$

In the formula, Δm is the numerical diffusion of the degree of membership. The fingerprint matrix of Clustering Based on Scattering (CBS) is as follows:

$$\boldsymbol{F}_{CBS} \triangleq \mathbb{E}\left\{M_{CIR} \odot M_{CIR}^*\right\} \in \mathbb{R}^{N_t \times N_g} \tag{9}$$

Notation: M_{CIR} is the total CIR matrix of the user, and $\boldsymbol{F}_{CBS}$ is the final fingerprint matrix in the CBS.

3.2 Terminal positioning

3.2.1 Fingerprint Similarity Criteria.

For the calculation of the distance between fingerprints, due to the needs of the positioning itself and the consideration of the practicality of the algorithm, the overall distance function between the fingerprints should show a monotonous increasing or monotonous decreasing trend. In this section, the Joint Angle Delay Similarity Coefficient **(JADSC)** is introduced to measure the similarity between fingerprints [5].

For the joint angle delay similarity coefficient at position i and position j, it is defined as:

$$J_{ij} = \max_{n \in (-L+1, L-1)} \left(\frac{1}{N_t^2} [C_i]_t^T [C_j]_{t-n}\right) \tag{10}$$

Notation: Matrix C_i represents the fingerprint matrix at position I, $[C_i]_t$ represents the t-th column of the fingerprint matrix, n represents the delay compensation between two fingerprints, and L represents the maximum delay compensation.

From the above, in the massive MIMO-OFDM system, the base station is usually deployed in a high physical location. When the number of antennas increases, the JADSC is as follows:

$$J_{ij} = \sum_{p_1=1}^{P_i} \sum_{p_2=1}^{P_j} \sigma_{p_1,i} \sigma_{p_2,j} \delta(\varphi_{p_1,i} - \varphi_{p_2,j}) \delta(n_{p_1,i} - n_{p_2,j}) \tag{11}$$

In the formula, P represents the nearest P scatterers that affect the fingerprint characteristics, and σ represents the channel energy of the corresponding scatterers. It can be further inferred from the above formula that the JADSC is mainly determined by the overlap of the scatterers of the fingerprints at the two positions. The simulation diagram is shown in Figure.3.
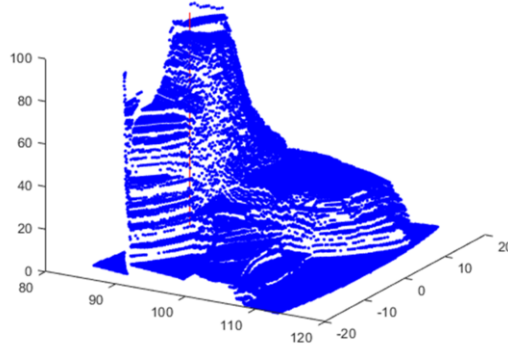

Figure 3. Schematic diagram of JADSC.

In the figure, each point corresponds to a reference point, the X-axis and Y-axis coordinates represent the physical position of the reference point, and the Z-axis value represents JADSC. The straight line perpendicular to the XOY plane is a randomly selected point, that is, the actual point to be located.

The simulation test proves that the reference points with larger JADSC are all near the points to be located, and the overall distribution of JADSC shows a monotonous decreasing trend, thus the JADSC meet the conditions of similarity standards between fingerprints.

3.2.2 Terminal positioning algorithm.

Nearest Neighbor (NN) is a basic classification and regression method in machine learning, mainly including K-Nearest Neighbor (KNN) and weighted K-Nearest Neighbor (WKNN). The KNN method contains two basic elements: distance measurement and the choice of K value. This article uses JADSC as the distance measurement standard. In the selection of the K value, if the value is small, it may lead to over-fitting; if the value is too large, it will cause too many points to be substituted into the calculation, resulting in a decrease in accuracy and causing under-fitting. The calculation formula for positioning by KNN method is as follows:

$$\hat{\boldsymbol{p}}_{x,y} = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{p}_{x,y}^{k} \tag{12}$$

In the formula, $\hat{\boldsymbol{p}}_{x,y}$ represents the position coordinates of the point to be located, and $\boldsymbol{p}_{x,y}^{k}$ represents the coordinates of the k-th reference point.

The operation of the KNN algorithm is simple and easy to implement. When a more reasonable K value is selected, the positioning accuracy is also guaranteed to a certain extent. However, the selected K reference points have the same weight, which makes the KNN algorithm susceptible to some outliers.

Therefore, this paper proposes a better algorithm: WKNN algorithm. The steps of the WKNN algorithm are roughly the same as KNN, but when substituting the reference point position coordinates, a weight is set for each reference point. The more similar the reference point to the point to be located, the greater the impact on the positioning result, which ensures the stability of the algorithm. The calculation formula of WKNN is:

$$\hat{\boldsymbol{p}}_{x,y} = \sum_{k=1}^{K} w_k \cdot \boldsymbol{p}_{x,y}^k, \quad w_k = \frac{\dfrac{1}{d_k}}{\sum_{k=1}^{K} \dfrac{1}{d_k}} \tag{13}$$

In the formula, w_k represents the weight corresponding to the k-th reference point, and its value satisfies $\sum w_k = 1$, and d_k is the JADSC of the k-th reference point and the point to be located.

The nearest neighbor method and related improved algorithms do not need to train the model in advance, and can be used directly during matching, but it will consume a lot of time to traverse the fingerprint database during online matching. In specific use, it can be combined with the clustering algorithm to greatly improve the efficiency of the algorithm.

4. Simulation and Analysis

4.1 Simulation parameter setting

The simulation is based on MATLAB to build a massive MIMO-OFDM single station system. The cell radius is R=500m, the radius of the area to be located is R_{loc}=300m, the width is W_{loc}=100m, and the angle range is A_{loc}=9/π. The base station is located in the center of the cell. The position of the scatterer is shown in Figure.4.
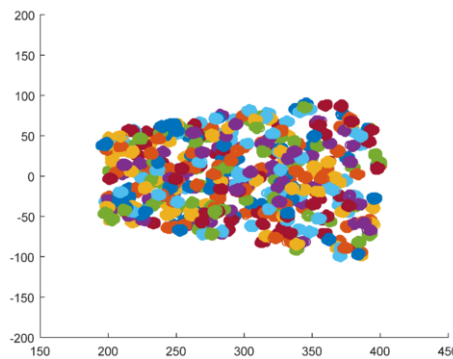

Figure 4. Scatterer model.

The reference points are divided at an interval of 2m, evenly distributed in the area to be located. When collecting fingerprint information, it is necessary to traverse all reference points. Then compress the fingerprint information, and finally store it in the

fingerprint database together with the corresponding position information. The reference point distribution model is shown in Figure.5.
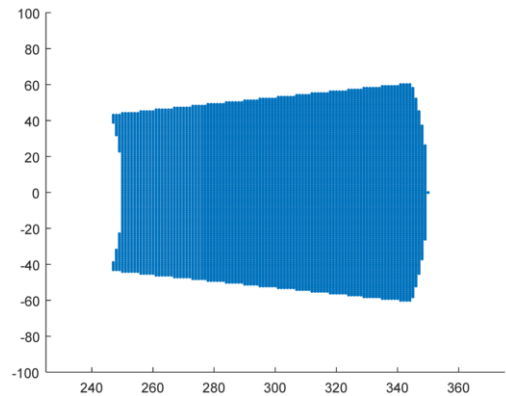

Figure 5. The reference point distribution model.

Other relevant parameters are shown in Table 1.

Table1. System Simulation Parameters

OFDM parameters	Transmission bandwidth B_W	5-80MHz
	Subcarrier spacing Δf	15kHz
	Pilot interval T_g	4.7 μs
	Sampling interval T_s	1/(2*BW)
Antenna parameters	Number of antennas N_t	128
	The location of the base station BS_{pos}	[0;0]
Other parameters	Scatterer radius R_{Scat}	20m
	Scatterer density	$0.02/m^2$
	Subpath N_{Sub}	40
	Test points **N**	1000

4.2 Simulation results

Perform a simulation test based on the WKNN algorithm to select the best K value for subsequent comparison and analysis. The simulation result based on different K values in the WKNN algorithm is shown in Figure.6.
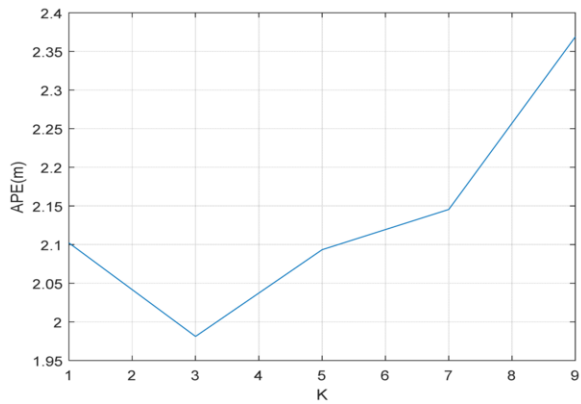
Figure 6. Average positioning accuracy based on different K values in WKNN.

Simulation analysis: In the WKNN algorithm, when the K value is small, the number of reference points is small, which brings greater uncertainty to the result, resulting in loss of some positioning accuracy. When the K value is large, there are too many reference points, and some reference points that are not particularly correlated are also included in the calculation of WKNN, resulting in a decrease in the accuracy of the algorithm. It can be seen from the figure above that the optimal K value should be 3.

The simulation test is carried out around the clustering algorithm to verify the performance of the clustering algorithm proposed in this paper. The positioning result of the two clustering algorithms is shown in Figure.7.
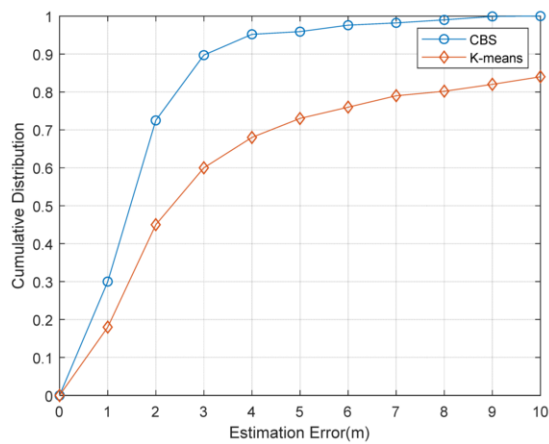

Figure 7. Simulation result of the clustering algorithm.

Simulation analysis: As can be seen from the figure above, the scattering-based clustering algorithm proposed in this paper has more superior performance. The K-means algorithm is a hard-clustering algorithm. When calculating JADSC, due to the large number of reference points, the JADSC value is approximately continuous, and the decrease in certain distances is not obvious, which leads to the low similarity of the searched similar points. In addition, the K-means algorithm uses the reference point selected by the cluster as the standard. The clustering effect of points near the reference point is acceptable, but for other reference points, there may be some errors in the

clustering division. In the simulation data, there are many high error points, and the error value may be close to 10m.

5. Conclusion

This paper constructs a massive MIMO-OFDM system model, extracts fingerprints based on CIR information, and associates location information with fingerprint information to establish a fingerprint database. Due to the huge amount of data in the fingerprint database, after introducing the classic clustering algorithm and combining with specific scenarios, this paper proposes a scattering-based clustering algorithm to improve the performance of the algorithm. In the terminal positioning part, the proposed clustering algorithm is verified by simulation test based on WKNN algorithm, which proves its superiority.

References

[1] Angjelicinoski M, Atanasovski V, Gavrilovska L. Comparative analysis of spatial interpolation methods for creating radio environment maps[C]// Telecommunications Forum. IEEE, 2012:334-337.

[2] Parkinson B W, Gilbert S W. NAVSTAR: global positioning system-ten years later[J]. Proceedings of the IEEE, 1983, 71(10):1177-1186.

[3] Wang X, Gao L, Mao S, et al. CSI-based Fingerprinting for Indoor Localization: A Deep Learning Approach[J]. IEEE Transactions on Vehicular Technology, 2016, 66(1):763-776.

[4] Wang X, Gao L, Mao S. CSI Phase Fingerprinting for Indoor Localization with a Deep Learning Approach[J]. IEEE Internet of Things Journal, 2016, 3(6):1113-1123.

[5] Sun X, Gao X, Ye G L, et al. Single-Site Localization based on a New Type of Fingerprint for Massive MIMO-OFDM Systems[J]. IEEE Transactions on Vehicular Technology, 2018, 67(7):6134-6145.

[6] Wang X, Liu L, Lin Y and Chen X. A Fast Single-Site Fingerprint Localization Method in Massive MIMO System[C]// 2019 11th International Conference on Wireless Communications and Signal Processing.

[7] Kupper A. Location-based Services: Fundamentals and Operation[J]. Jcms Journal of Common Market Studies, 2005, 49(5):923-947.

[8] Kushki A, N Plataniotis K, N Venetsanopoulos A. Kernel-based Positioning in Wireless Local Area Networks[J]. IEEE Transactions on Mobile Computing, 2007, 6(6): 689-670.