A Low-Cost Outdoor Fingerprinting Localization Scheme For Wireless Cellular Networks

Abstract: This paper considers outdoor fingerprinting localization in LTE cellular Networks, which can localize non-cooperative user equipment (UE) that is unwilling to provide Global Positioning System (GPS) information. We propose a low-cost fingerprinting localization scheme that can improve the localization accuracy while reducing the computational complexity. Firstly, a data filtering strategy is employed to filter the fingerprints which are far from the target UE by using the Cell-ID, Timing Advance (TA) and eNodeB environment information, and the distribution of TA difference is analyzed to guide how to use TA rationally in the filtering strategy. Then, improved Weighted K Nearest Neighbors (WKNN) are implemented on the filtered fingerprints to give the final location prediction, and the WKNN is improved by removing the fingerprints that are still far away from the most of the K neighbors. Experiment results show that the performance is improved by the proposed localization scheme, and positioning errors corresponding to Cumulative Distribution Function (CDF) equaling to 67% and 95% are declined to 50 m and 150 m.

Keywords: fingerprinting localization; TA; filtering strategy; improved WKNN

PEI Dengke, XU Xiaodong, QIN Xiaowei, LIU Dongliang, and ZHAO Chunhua

(Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, Anhui 230027, China)

DOI: 10.12142/ZTECOM.201903007

http://kns.cnki.net/kcms/detail/34.1294.TN.20190920.2103.002.html, published online September 20, 2019

Manuscript received: 2019-05-18

1 Introduction

Benefiting from the fast growing of wireless techniques, location-based services (LBS) have made great development in various fields, such as emergency rescue, reconnaissance survey, and intelligent transportation systems. Several localization approaches have been analytically presented in literature for in-building and outdoor environments. Global Positioning System (GPS) is considered as one of the most well-known outdoor localization techniques [1], since it can provide acceptable localization accuracy as long as the target user equipment (UE) can receive navigation messages from at least four visible satellites. With the ubiquitous terrestrial wireless cellular network, LBS can also be provided by several eNodeBs through Time of Arrival (TOA), Time Difference of Arrival (TDOA) and Angle of Arrival (AOA) measurements. It appears apparently an important way to localize non-cooperative UE which is unwilling to provide GPS information, but the network has to be equipped with multiple antennas and additional location measurement units (LMUs) in general.

As a cost-saving counterpart, the pioneer framework of fingerprinting localization is proposed in [2], where no additional hardware modification is needed. The process of fingerprinting localization can be constituted by two phases, namely off-line construction of fingerprinting database and on-line localization. During the off-line phase, the received signal strengths from adjacent access points are tabulated as location-related fingerprints, and the fingerprints with the corresponding physical coordinates are stored in a database. While during the on-line localization phase, target user reports the observed signal strengths so that the user's location can be estimated by matching in the database. As for wireless cellular networks, UE at a specific place can receive information of multiple Reference Signal Received Powers (RSRPs) from multiple eNodeBs, and the tuple of these RSRP measurements is usually

This work was supported by the ZTE Industry-Academia-Research Cooperation Funds under Grant No. 20160722-01.

employed as the fingerprint. In literature, [3] presents a probabilistic Received Signal Strength Indicator(RSSI) based fingerprinting localization system for GSM wireless cellular networks, which needs a large amount of fingerprinting information to build the probabilistic model for a particular location. In this regard, the database construction becomes laborious and time costly, and the localization accuracy is not satisfactory since RSSI is the only information in use. For the sake of decreasing computational complexity as wells as increasing localization accuracy, [4], [5] and [6] use the Cell-ID as an additional information to form the fingerprint, and deterministic localization algorithms, such as Weighted K Nearest Neighbors (WKNN) are used, where fewer fingerprinting records are measured as compared with the probabilistic localization one. However, this kind of fingerprinting localization scheme may eliminate those physically much closer fingerprints so as to induce attenuated localization accuracy.

In this paper, we put forward a low-cost outdoor fingerprinting localization scheme for the fourth generation (4G) wireless systems, which is shown to have decreased positioning error and computational load in contrast to the counterparts. During the off-line phase, database construction can be simply implemented through the UE measurement report (MR) that is gathered at the network side. Different from the most existing works in which fingerprints just contain RSRP, the fingerprint used in this paper is comprised of RSRP, Cell-ID, and Timing Advance (TA). What's more, we also derive a theoretical analysis to demonstrate the usage of TA. Therefore, we would like to shed more light on the improvement during on-line localization phase, and the localization result is shown to be better than the related work [4] with the same dataset. The main technical contributions are listed as follows:

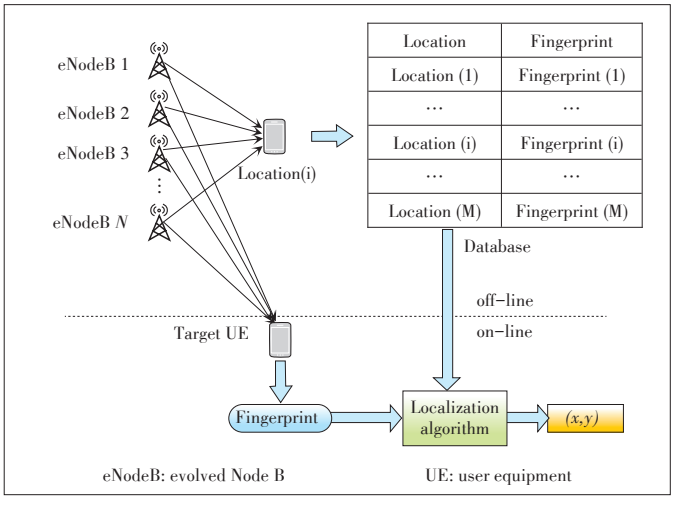
1) A database filtering strategy is proposed to exclude the fingerprints which are far from the target UE before matching, where we preserve only the fingerprints which have the same serving Cell-ID, approximate TA and approximate eNodeB environment for the target UE. Furthermore, the potential range of TA at a particular location is theoretically analyzed to help us construct a proper filtering threshold. In this way, the positioning error and computational load can be decreased to some extend since during the matching process, fingerprints with geographic positions closer to target UE have a larger probability to participate in the computation.

2) WKNN is employed to compute the geographic position of the target UE according to the fingerprinting database. K nearest neighbors are selected from the database according to the distance between target UE and their RSRP measurements. However, due to the intrinsic intense fluctuation of RSRP measurements, two adjacent fingerprints with the minimum Euclidean distance do not always mean that they are close to each other in geographic position. Therefore, we further improve the WKNN matching algorithm by removing those anomaly fingerprints that are still possibly far away from the others among K neighbors.

2 Architecture of Fingerprinting Localization

Obeying the two- phase strategy of off-line construction of fingerprinting database and on-line localization, **Fig. 1** depicts an architecture of fingerprinting localization for 4G wireless cellular networks. In this paper, the RSRPs and cellular network information from eNodeBs received by the measured UE are denoted as fingerprint, and every fingerprint and the corresponding geographic coordinates provided by GPS (AGPS) modules are stored in the database (Fig. 1). For clarity, the stored samples with sufficient information about geographic coordinate in the database are called reference points (RPs). In the fingerprinting database, detailed information about a RP can be expressed as **Table 1**, where x, y represent the longitude and latitude of the corresponding RP respectively, and the Cell-ID$_n$ stands for the Cell-ID of the severing eNodeB, and the value of the TA for the severing eNodeB is m. N is the number of total eNodeBs measured by all of the RPs in the database. RSRP$_i$ represents the RSRP of the ith eNodeB. Note that, for one RP, the number of detectable eNodeB is l, so for $N - l$ unreachable eNodeBs, the corresponding RSRPs will be set as a minimum level of detectable signal strength empirically [4], which is -140 dBm in this paper.

As shown in Table 1, almost all location-related parameters have been considered in this paper. But unfortunately, none of



▲**Figure 1. An architecture of fingerprinting localization for 4G system.**

▼**Table 1. Information of reference points**

Longitude	Latitude	Severing Cell-ID	TA	Detectable eNodeBs	Cell-ID$_1$	⋯	Cell-ID$_i$	⋯	Cell-ID$_N$
x	y	Cell-ID$_n$	m	l	RSRP$_1$	⋯	RSRP$_i$	⋯	RSRP$_N$

RSRP: Reference Signal Received Power TA: Timing Advance

them has an explicit relationship with the location of target UE in reality due to the complicated characteristics of the propagation channel. For instance, the instantaneous values of RSRP and TA from a single eNodeB are often sensitive to the environment so that they vary distinctively even for multiple observations at the same place. It is reported in [7] that RSRP, Cell-ID and TA information is used to classify UE, while in this paper, we would like to further perform a new application, in which more comprehensive information is used to calculate the geographical coordinates of UE.

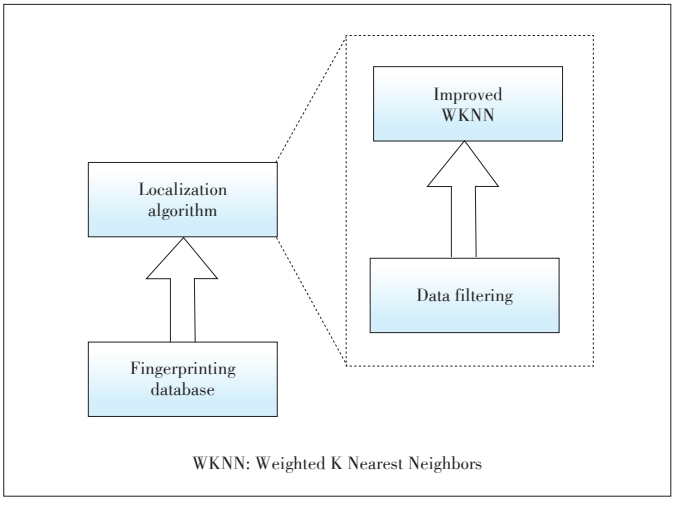
3 Proposed Fingerprinting Localization Scheme

In this section, the procedure of online fingerprinting localization will be introduced. **Fig. 2** illustrates the proposed scheme, which consists of a data filtering module followed by an improved WKNN algorithm.

3.1 A Data Filtering Strategy

As mentioned above, thousands of RPs have been stored in the fingerprinting database according to their MRs. Given a fingerprint from the target UE, it is computationally wasteful to use WKNN to figure out the most approximate RP in the whole dataset directly with a brute-force searching. On the other hand, RSRP information is commonly used by WKNN during the matching process, but in general RSRP fluctuates too sharply to obtain expected results. In this regard, a data filtering method should be introduced beforehand, and the main purpose of data filtering is to perform the matching process in a constrained region so as to decrease the localization error as well as the computational load. In the sequel, additional steady information is used to construct the constrained region, including Cell-ID, TA and eNodeB environment.

The filtering strategy is divided into three steps. Firstly, we use Cell-ID to eliminate candidate RPs which have different severing eNodeB as the target UE, since UEs belonging to distinctive eNodeBs are generally far from each other to a large extent. However, there is at least one specific case that the Cell-ID based data filtering strategy is improper and should be skipped. That is, when the target UE is at the edge of an eNodeB, where the knowledge can be read by RSRP no more than a specified threshold in this paper, because the serving RSRP is below the specified thresholds mainly due to the high propagation loss caused by excessive distance between the eNodeB and the target UE. And the thresholds is set to -110 dBm according to [8].



▲**Figure 2. The proposed fingerprinting localization scheme.**

Secondly, TA is employed to further filtrate improper RPs within the candidates obtained through the first filtering step. Due to the immunity of TA against the variation of local propagation environment, we keep RPs in the candidate set which have approximate TAs with the target UE and remove the offenders according to the following rule

$$\left| \hat{t}\frac{UE}{TA} - \hat{t}\frac{RP}{TA} \leqslant \Delta T \right|, \tag{1}$$

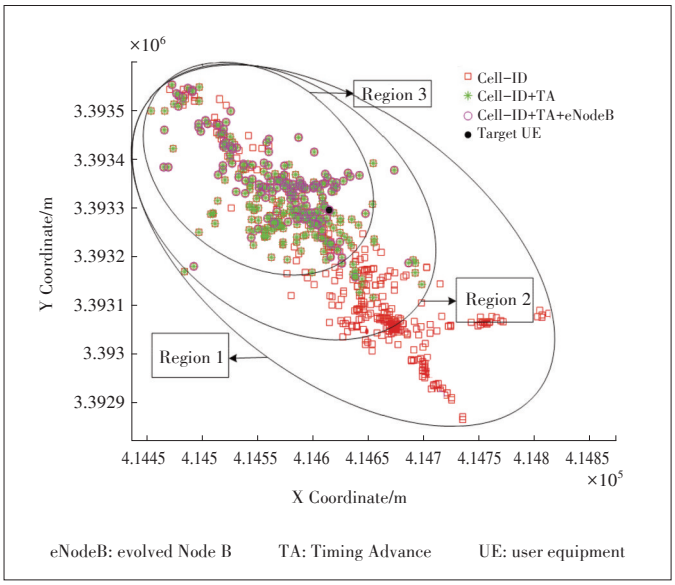
where $\hat{t}\frac{UE}{TA}$ and $\hat{t}\frac{RP}{TA}$ stand for measured TA of the target UE and certain RP respectively. And $\hat{t}\frac{UE}{TA} - \hat{t}\frac{RP}{TA}$ is called TA difference in this paper, which is an integer value and can be designed following the analytical result of TA difference in the next section.

The third and the last filtering steps are established based on the fact that the target UE and its neighbor RPs should be in similar eNodeB environment, and the similarity is reflected both in the number of total detected eNodeBs and the number of eNodeBs with the same Cell-IDs. In this work, the strong correlation between the target UE and RPs in the candidate set holds if the difference about the number of total detected eNodeBs is less than four, where at least three detected eNodeBs have the same Cell-IDs as the target UE [4]. After all, RPs violating this rule will be eliminated from the candidate set. Now the candidate set selected by the three filtering steps is defined as $\Re$.

To end this discussion, a numerical example is presented to show the effectiveness of the data filtering strategy in **Fig. 3**. We highlight three regions in this figure to clarify such a filtering process. It is clearly that the regions become more precise with further rigorous filtering conditions. The following matching process can now be performed on the candidate set $\Re$, therefore the localization error can be reduced and the number of candidates is decreased from ten thousands to less than one hundred for about 1 000 times.

3.2 Improved WKNN Algorithm

Conventional WKNN algorithm is widely used in fingerprint-



▲**Figure 3. Filtering Results.**

ing localization and has shown the superior performance due to many advantages [9], such as being easy to understand, being robust to noise, and high accuracy in localization. However, it is inefficient in terms of computational complexity if one performs WKNN directly on a large training dataset. Fortunately, this disadvantage can be alleviated by the data filtering strategy presented above.

Because of the drastic fluctuation of instantaneous RSRPs, RPs which have high fingerprinting similarity with the target UE may be far from the UE geographically. In this regard, among K neighbors chosen by WKNN for the target UE, some should be looked as erasable anomaly RPs to further improve the performance of localization accuracy. In the sequel, we present a way to remove these anomaly RPs from K neighbors and name the modified WKNN algorithm as an improved one.

When K nearest neighbors have been selected, the improved WKNN algorithm continues to compute the geographic distance for every pair of them. We assume that RPs which are far away from the others are regarded as anomalous RPs, and they should not participate in location calculation. The procedure of the improved WKNN algorithm is summarized as follows:

1) Calculate the distance between the target UE and each filtrated RP in the candidate set according to the following equation

$$e(k)=\left(\sum_{i=1}^{l}(UE_i-RP_i(k))^2\right)^{1/2},\ k=1,2,3,\ldots,M, \tag{2}$$

where $e(k)$ is the Euclidean distance between target UE and the kth RP in the sense of RSRP. M is the total number of RPs in the candidate set $\Re$, and UE_i represents for the measured RSRP of the detectable eNodeB by the target UE. l is the number of the detected eNodeBs by the target UE.

2) Sort $\{e(k)\}$ to choose K RPs as the nearest neighbors with K lowest distances; in this paper, the value of K equals to 7. Then we calculate the geographic distance between any two nearest neighbors and eliminate T anomaly RPs which are far away from the others for more than 300 m in general.

3) Compute the weight coefficients through

$$w_{\tilde{k}}=\frac{1/(e(\tilde{k})+0.001)}{\sum_{\tilde{k}=1}^{K-T}1/(e(\tilde{k})+0.001)}, \tag{3}$$

where $\tilde{k}$ represents the kth RP of the rest $K-T$ nearest neighbors after erasing anomaly. The resulting weight coefficients are used to proceed to calculate the geographic coordinate of the target UE. The relationship between the target UE and the rest $K-T$ nearest neighbors can be formulated as

$$\begin{cases}\tilde{x}_{UE}=\sum_{\tilde{k}=1}^{K-T}w_{\tilde{k}}x_{\tilde{k}}\\ \tilde{y}_{UE}=\sum_{\tilde{k}=1}^{K-T}w_{\tilde{k}}y_{\tilde{k}}\end{cases}, \tag{4}$$

where $(\tilde{x}_{UE},\tilde{y}_{UE})$ denotes the geographic coordinate of the target UE, and $(x_{\tilde{k}},y_{\tilde{k}})$ stands for the geographic coordinate of the kth neighbor.

4 Analysis of TA Difference

The threshold of TA difference plays an important role on the performance of the proposed data filtering strategy. To shed more light on how to design this threshold, we will give an analytical result in the following.

TA can be used as an approximation for TOA [10], but TA represents the round trip time. Therefore, $|t_{TA}^{UE}-t_{TA}^{RP}|/2$ equals to the discrete value of $|t_{TOA}^{UE}-t_{TOA}^{RP}|$, where t_{TA}^{UE} and t_{TA}^{RP} are the TA of the target UE and RP respectively. Similarly, t_{TOA}^{UE} and t_{TOA}^{RP} represent the TOA of the target UE and RP. Based on this perspective, if we can infer the fluctuation range of TOA, we can get the fluctuation range of the TA.

TA in [11] can be modeled by

$$\hat{t}_{TA}=t_{TA}+\xi, \tag{5}$$

where $\hat{t}_{TA}$ and t_{TA} are the measured TA and actual value of TA respectively, and t_{TA} can be seen as the mean value of $\hat{t}_{TA}$. ξ is the error corrupting the true value and can be modeled as a random variable obeying normal distribution. As for TOA, it is reasonable to construct the observation model as

$$\hat{t}_{TOA}=t_{TOA}+\omega, \tag{6}$$

where $\hat{t}_{TOA}$ and t_{TOA} are the measured TOA and actual value of TOA respectively and ω can also be modeled as a Gaussian variable.

$$\omega\sim N(0,\sigma^2). \tag{7}$$

Given two independent identically distributed random variables $\hat{t}_{TOA1}$ and $\hat{t}_{TOA2}$ obeying (6), we can easily get the Gaussian distribution of the TOA difference:

$$\hat{t}_{TOA1} - \hat{t}_{TOA2} \sim N(0, 2\sigma^2). \quad (8)$$

In the same way, given two independent identically distributed random variables $\hat{t}_{TA1}$ and $\hat{t}_{TA2}$, TA difference shall obey the Gaussian distribution with standard deviation $\hat{\sigma}$ and $\hat{\sigma}^2 = 8\sigma^2$ can be gotten according to the relationship between TOA and TA. It appears that if the standard deviation of the measured TOA can be calculated, the distribution of TA can thus be acquired. Hence in the next step, we attempt to explore an approximate solution about the standard deviation of the measured TOA.

In the multipath channel environments, TOA means the signal arrival time about first path, so the fluctuation range of the TOA cannot be greater than τ, the delay spread of the propagation channel. In nature, the delay spread τ can be interpreted as the difference arriving time between the earliest multipath component (typically the line-of-sight component) and the latest multipath component. On the other hand, according to the three-sigma rule, more than 99.73% samples lie within a band around the mean in a normal distribution $N(\mu, \sigma_1^2)$ with a width six standard deviations. In short, random variable distributes in range $[\mu - 3\sigma, \mu + 3\sigma]$ with an asymptotic probability to 1. Therefore, the relationship between τ and the fluctuation range of the TOA can be described as

$$\tau \geqslant \hat{t}_{TOA\_max} - \hat{t}_{TOA\_min} = 6\sigma, \quad (9)$$

where $\hat{t}_{TOA\_max}$ and $\hat{t}_{TOA\_min}$ mean the maximal TOA the minimum TOA respectively. In this paper,τ can be calculated according to [12]

$$\tau = exp(A \times L + B), \quad (10)$$

where $A = 0.038, B = 2.3$, and L is path loss (dBm). In this way, we can obtain the upper bound of the standard deviation of the measured TOA, which is $\tau/6$, and the standard deviation of TA difference can then be described as $\hat{\sigma} \leqslant \sqrt{2}/3\, exp(A \times L + B)$. Taking the upper bound as the variance, the distribution of TA difference can thus be formulated by

$$\hat{t}_{TA1} - \hat{t}_{TA2} \sim N(0, (\frac{\sqrt{2}}{3} exp(A \times L + B))^2). \quad (11)$$

Defining $\Delta T = \delta\hat{\sigma}$ as a threshold of TA difference for data filtering strategy, and the corresponding confidence intervals as γ, we have:

$$p(|\hat{t}_{TA}^{UE} - \hat{t}_{TA}^{RP}| \leqslant \Delta T) = p(|t_{TA}^{UE} - t_{TA}^{RP}| \leqslant \delta\hat{\sigma}) =$$

$$\int_{-\delta\hat{\sigma}}^{\delta\hat{\sigma}} \frac{1}{\sqrt{2\pi}\,\hat{\sigma}} exp(\frac{-x^2}{2\hat{\sigma}^2})\, dx =$$

$$\int_{-\infty}^{\delta\hat{\sigma}} \frac{1}{\sqrt{2\pi}\,\hat{\sigma}} exp(\frac{-x^2}{2\hat{\sigma}^2})\, dx - \int_{-\infty}^{-\delta\sigma} \frac{1}{\sqrt{2\pi}\,\hat{\sigma}} exp(\frac{-x^2}{2\hat{\sigma}^2})\, dx =$$

$$erf(\frac{\delta}{\sqrt{2}}) = \gamma, \quad (12)$$

where δ and γ are positive, and γ should be no greater than 1. Apparently, δ can be given by:

$$\delta = \sqrt{2}\, erf^{-1}(\gamma). \quad (13)$$

In this paper, we set $\gamma = 0.9$, which means that 90% RPs that are at the same location as target UE will be selected through this TA difference threshold and that $\delta = 1.65$ according to (13), thus the filtering rule can be described as

$$|\hat{t}_{TA}^{UE} - \hat{t}_{TA}^{RP}| \leqslant 1.65 \times \frac{\sqrt{2}}{3} exp(A \times L + B). \quad (14)$$

From (14), the path loss of the actual environment has a great influence on the threshold ΔT. To compute the path loss, Standard Propagation Model (SPM) [13], [14] is considered in this paper, which can be formulated by

$$L_{SPM} = K_1 + K_2 \times \log(d) + K_3 \times \log(h_{te}) + K_4 \times (Diffract) + K_5 \times \log(d) \times \log(h_{te}) + K_6 \times h_{te} + K_7 \times f(clutter), \quad (15)$$

where K_1 is the constant offset related to frequency; d is the distance between the receiver and the transmitter (m); K_2 is the multiplying factor for $\log(d)$; h_{te} is the effective height of the transmitter antenna (m); K_3 is the multiplying factor for $\log(h_{te})$; K_4 is the multiplying factor for diffraction calculation;. K_5 is the multiplying factor for $\log(d) \times \log(h_{te})$; h_{te} is the effective height of the receiver mobile antenna; K_6 is the multiplying factor for $\log(h_{te})$; $f(clutter)$ is the average of weighted losses due to clutter.

For users within a particular cell, the maximum pass loss often happens when they are at the cell edge. In an LTE system, h_{te} is about 30 m usually, and K_3, K_5, K_6 are often set separately with default values 5.83, −6.55, and 0. K_4 and K_7 can be zero each as in [13] and [15]. K_1 and K_2 are adjusted in our situation as $K_1 = 25, K_2 = 37$. Under the consumption of d=400 m for the effective cell radius in an urban area, the maximum path loss can eventually be calculated as 104.7 dBm according to (15). Therefore, the resulting threshold. $\Delta T = 1.65 \times \hat{\sigma} =$ 414 ns. In LTE system, TA is practically quantized in bit period T_b with $T_b = 520$ ns. Therefore, the quantized threshold used in this paper should be set as 1 TA. By this filtering rule we can select most of the fingerprints who are close to the target UE, and exclude the fingerprints who are far from the UE geographically.

5 Experimental Results

To validate the performance of the proposed fingerprinting localization scheme, some experiments will be presented in this section by using the experiment data collected at some districts in Sichuan Province, China. There are 42 588 MR messages with exact geographic coordinates. 20% of them are selected randomly as testing data, and 80% of them are used to construct database. The experiments include three parts. The first part demonstrates the contributions of filtering on localization accuracy and computational burden. The second part com-

pares different matching algorithms and demonstrates the superiority of the improved WKNN. The third part tests the localization scheme by doing experiments in different districts with different experimental data.

5.1 Effect of the Data Filtering Strategy

We examine the effect of the proposed filtering strategy on improvement of localization accuracy by using different filtering strategies in the experimental area. The experimental area is 35 000×35 000 m^2, containing 34 070 RPs and 8 518 test data (target UEs). Different filtering strategies are as follows:

1) Select all the RPs which have the same Cell-ID with the target UE in the filtering process

2) Select all the RPs which have the same Cell-ID and similar TA with the target UE in the filtering process

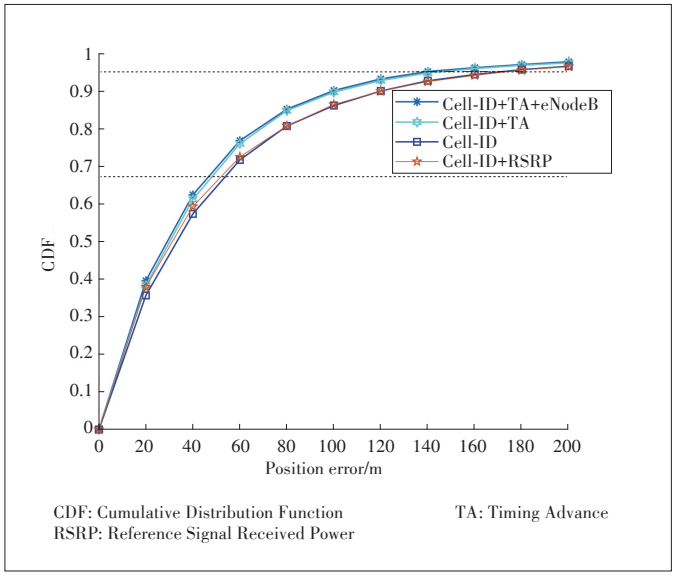
3) Select all the RPs which have the same Cell-ID, similar TA and similar eNodeB environment with the target UE in the filtering process

4) Use the filtering strategy in [4], where TA information is not considered.

▼ **Table 2. Positioning error with different filtering strategies**

Location error / Filter criteria	67% error/m	95% error/m	Mean error/m	Time/s
Cell-ID	52.6	170.66	58.7	562
Cell-ID+TA	46.8	146.8	52.4	401
Cell-ID+TA+eNodeB	44.4	139.2	48.7	104
Cell-ID+RSRP	50.3	169.45	57.2	204

RSRP: Reference Signal Received Power TA: Timing Advance



CDF: Cumulative Distribution Function TA: Timing Advance
RSRP: Reference Signal Received Power

▲ **Figure 4. Positioning error with different filtering strategies.**

Table 2 shows the positioning errors and runtime of the different filtering strategies, where 67% error and 95% error mean the positioning errors corresponding to CDF = 67% and CDF = 95%, the mean error is the mean positioning error, and the time is the total runtime for the 8 518 test data. **Fig. 4** is the CDF of the positioning error, where the legend Cell-ID, TA and eNodeB represent the remaining RPs after the corresponding step of data filtering. We use Cell-ID+RSRP to stand for the filtering strategy presented in [4]. From Fig. 4 and Table 2, it is clear that the positioning error of the proposed scheme has been reduced obviously by adding the TA filtering criterion, and the time spent for positioning has been reduced obviously by adding the eNodeB environment information. By this way, the mean positioning error declines to 50 m, and the positioning error corresponding to CDF = 67% and CDF = 95% are less than 50 m and 150 m, which have achieved the localization requirements of Federal Communication Commission (FCC), the United States, whereby mobile operators should be able to determine the location of a UE within 50–100 m in 67% cases and 150–300 m in 95% cases. And the filtering strategy presented in [4] has a similar positioning errors as the proposed one where only the first data filtering step is performed.

5.2 Performance of the Improved WKNN

Here we compare different matching algorithms to examine the performance of improved WKNN. Firstly, we compare different matching algorithms containing WKNN, Support Vector Machine (SVM), and Back Propagation Neural Networks (BPNN). The areas of experimental regions are chosen as 100×100 m^2, 400×400 m^2 and 1 000×1 000 m^2, respectively. The localization results of different matching algorithms are shown in **Table 3**, where we can see that WKNN shows superior performance in all of the experiments. The performance of SVM and BPNN can be acceptable when the experimental area is small, but with the increase of the experimental area, the performance of BPNN and SVM becomes worse than WKNN. More clearly, the mean localization error for WKNN is 33 m, such as the experimental results with the region of 1 000×1 000 m^2, but the

▼**Table 3. Comparison of different matching algorithms**

		67% error/m	95% error/m	Mean error/m
100 × 100 m^2	WKNN	23	43	18
	SVM	23	45	19
	BPNN	25	56	22
400 × 400 m^2	WKNN	27	79	27
	SVM	29	73	27
	BPNN	39	77	35
1 000 × 1 000 m^2	WKNN	30	100	33
	SVM	46	155	49
	BPNN	70	191	70

BPNN: Back Propagation Neural Network WKNN: Weighted K Nearest Neighbors
SVM: Support Vector Machine

mean localization errors for SVM and BPNN are 49 m and 70 m respectively.

Then the contrastive experiment between WKNN and improved WKNN has been operated. The area of experimental region is 35 000 ×35 000 m^2, containing 34 070 RPs and 8 518 test data (target UEs). Localization results of WKNN and improved WKNN are showed in **Table 4**. Just as expected, the localization performance is improved by using improved WKNN, especially with the CDF=95%, which testifies that the improved WKNN is effective for removing the anomalous RPs who are harmful to localization accuracy.

5.3 Application to Other Districts

In this section, we conduct experiments in different districts to test the robustness of the proposed localization scheme. The experimental results in **Table 5** show that the positioning errors corresponding to CDF = 67% are almost the same between District A, District B, and District C, and the positioning errors corresponding to CDF = 95% have a little difference but also in an acceptable range, and the mean errors also keep jarless between different districts. Generally speaking, there is no significant difference in positioning errors between different districts, which declares that the localization scheme proposed in this paper is robust, and the parameter setting is reasonable. Furthermore, the experimental results can meet the FCC requirements in all of the experimental districts.

6 Conclusions

In this paper, we present an outdoor fingerprinting localization scheme for wireless cellular networks, which is low-cost and high-efficiency with localization accuracy meeting the FCC requirements for the case of urban environments. In the off-line database construction phase, we use the MRs obtained from the network side instead of laborious war driving. In the online localization phase, firstly, a data filtering strategy is proposed to improve the performance of localization, which selects the RPs that own the same Cell-ID, similar TA and eNodeB environment with the target UE, and the distribution of TA difference has been analyzed to guide how to use TA rationally in the filtering strategy. Secondly, WKNN has been improved by computing the geographic distance for every pair of K nearest RPs, and the RPs who are far away from the others should not participate in location calculation. Some experiments have been conducted and experimental results demonstrated that the localization scheme proposed could decrease the localization error as well as the computational load, and the mean positioning error declines to 50 m, and positioning errors corresponding to CDF= 67% and CDF = 95% are less than 50 m and 150 m.

In the future, we will investigate how to further improve the outdoor localization accuracy and analyze the bound of outdoor localization accuracy for wireless cellular networks.

▼**Table 4. Comparison between WKNN and the improved WKNN**

Experimental areas	Matching algorithm	67% error/m	95% error/m
35 000 ×35 000 m^2	WKNN	46	147
	Improved WKNN	44	139

WKNN: Weighted K Nearest Neighbors

▼**Table 5. Localization in different districts**

	67% error/m	95% error/m	Mean error/m
District A	44	140	50
District B	44	132	48
District C	43	127	50

References

[1] IBRAHIM M, YOUSSEF M. CellSense: A Probabilistic RSSI-Based GSM Positioning System [C]// Global Telecommunications Conference. Miami, USA, 2010. DOI:10.1109/GLOCOM.2010.5683779

[2] YOUSSEF M A, AGRAWALA A, SHANKAR A U. WLAN Location Determination via Clustering and Probability Distributions [C]//IEEE International Conference on Pervasive Computing & Communications. Fort Worth, USA, 2003. DOI: 10.1109/PERCOM.2003.1192736

[3] BSHARA M, ORGUNER U, GUSTAFSSON F, et al. Fingerprinting Localization in Wireless Networks Based on Received-Signal-Strength Measurements: A Case Study on WiMAX Networks [J]. IEEE Transactions on Vehicular Technology, 2010, 59(1): 283–294. DOI: 10.1109/tvt.2009.2030504

[4] FAYAZ S, SARRAFIAN S. Location Service for Wireless Network Using Improved RSS-Based Cellular Localization [J]. International Journal of Electronics, 2014, 101(6): 763–778. DOI: 10.1080/00207217.2013.794492

[5] SCHROTH G, HUITL R, CHEN D, et al. Mobile Visual Location Recognition [J]. IEEE Signal Processing Magazine, 2011, 28(4): 77–89. DOI: 10.1109/MSP.2011.940882

[6] ZHANG J, HALLQUIST A, LIANG E, et al. Location-Based Image Retrieval for Urban Environments [C]//18th IEEE International Conference on Image Processing. Brussels, Belgium, 2011. DOI: 10.1109/ICIP.2011.6116517

[7] SHI L, WIGREN T. AECID Fingerprinting Positioning Performance [C]//IEEE Global Telecommunications Conference. Honolulu, USA, 2009. DOI: 10.1109/GLOCOM.2009.5425928

[8] GOMEZ-ANDRADES A, BARCO R, SERRANO I, et al. Automatic Root Cause Analysis Based on Traces for LTE Self-Organizing Networks [J]. IEEE Wireless Communications, 2016, 23(3): 20–28. DOI: 10.1109/MWC.2016.7498071

[9] MOGHTADAIEE V, DEMPSTER A G. Indoor Location Fingerprinting Using FM Radio Signals [J]. IEEE Transactions on Broadcasting, 2014, 60(2): 336–346. DOI: 10.1109/TBC.2014.2322771

[10] YOST G P, PANCHAPAKESAN S. Improvement in Estimation of Time of Arrival (TOA) from Timing Advance (TA) [C]// IEEE International Conference on Universal Personal Communications. Florence, Italy, 1998. DOI: 10.1109/ICUPC.1998.733714

[11] ROTH J D, TUMMALA M, MCEACHEN J C, et al. Maximum Likelihood Geolocation in LTE Cellular Networks Using the Timing Advance Parameter [C]// IEEE International Conference on Signal Processing & Communication Systems. Montreal, Canada, 2017. DOI: 10.1109/ICSPCS.2016.7843379

[12] ITU. Propagation Data and Prediction Methods for the Planning of Short-Range Outdoor Radio Communication Systems and Radio Local Area Networks in the

Frequency Range 300 MHz to 100 GHz [S], 2013.
[13] De Freitas P R, FILHO H T. Parameters Fitting to Standard Propagation Model (SPM) for Long Term Evolution (LTE) Using Nonlinear Regression Method [C]// IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA). Annecy, France, 2017: 84–88. DOI: 10.1109/CIVEMSA.2017.7995306
[14] RANI M S, BEHARA S, SURESH K. Comparison of Standard Propagation Model (SPM) and Stanford University Interim (SUI) Radio Propagation Models for Long Term Evolution (LTE), International Journal of Advanced and Innovative Research (IJAIR), 2012, 1(6): 221–228
[15] XU H, SHI C, ZHANG W, et al. Field Testing, Modeling and Comparison of Multi Frequency Band Propagation Characteristics for Cellular Networks [C]// IEEE International Conference on Communications. Kuala Lumpur, Malaysia, 2016. DOI: 10.1109/ICC.2016.7510961

Biographies

PEI Dengke (pdke1@mail.ustc.edu.cn) received the B.S. degree from China University of Geosciences, China in 2016. She is currently pursuing the B.E. degree at University of Science and Technology of China. Her research interest is localization and deep learning.

XU Xiaodong received his B.Eng. degree and Ph.D. degree in electronic and information engineering from University of Science and Technology of China (USTC) in 2000 and 2007, respectively. Since 2007, he has been a faculty member with the Department of Electronic Engineering and Information Science, USTC, where he is currently an associate professor. His research interests include array signal processing, wireless communications, and information-theoretic security.

QIN Xiaowei received the B.S. and Ph.D. degrees from the Department of Electrical Engineering and Information Science, University of Science and Technology of China (USTC) in 2000 and 2008, respectively. Since 2014, he has been a member of staff in Key Laboratory of Wireless-Optical Communications of Chinese Academy of Sciences at USTC. His research interests include optimization theory, service modeling in future heterogeneous networks, and big data in mobile communication networks.

LIU Dongliang received the B.S. degrees from the Telecommunications Engineering College, Beijing University of Posts and Telecommunications (BUPT), China, in 2004. Since 2006, he has been an Algorithms Researcher of the Algorithms Department in ZTE Corporation. His research interests include positioning theory, network optimization theory and big data in mobile communications.

ZHAO Chunhua received the M.S. degrees from the College of Electronic and Communication Engineering, Harbin Institute of Technology (HIT), China in 2002. Since then, she has been working on RRM algorithms design and network optimization in mobile communications. She has been an algorithms researcher at the Algorithms Department of ZTE Corporation since 2009. Her research interest is big data analytics for network optimizationeas of experimental regions are chosen as 100×100 in mobile communications.

←From P. 22

Accurate Object Detection and Semantic Segmentation [C]//IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 580–587. DOI: 10.1109/CVPR.2014.81
[35] GIRSHICK R. Fast R-CNN [C]//IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 1440–1448. DOI: 10.1109/ICCV.2015.169
[36] TANG X, DU D K, HE Z, et al. Pyramidbox: A Context-Assisted Single Shot Face Detector [C]//European Conference on Computer Vision (ECCV). Munich, Germany, 2018: 797–813. DOI: 10.1007/978-3-030-01240-3_49
[37] ZHANG Z, LUO P, LOY C C, et al. Facial Landmark Detection by Deep Multi-Task Learning [C]//European Conference on Computer Vision. Zurich, Switzerland, 2014: 94–108. DOI: 10.1007/978-3-319-10599-4_7
[38] CHEN D, REN S, WEI Y, et al. Joint Cascade Face Detection and Alignment [C]//European Conference on Computer Vision. Zurich, Switzerland, 2014: 109–122. DOI: 10.1007/978-3-319-10599-4_8
[39] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [DB/OL]. (2014-09-04). https://arxiv.org/abs/1409.1556
[40] YANG S, LUO P, LOY C C, et al. Wider Face: A Face Detection Benchmark [C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vega, USA, 2016: 5525–5533. DOI: 10.1109/CVPR.2016.596
[41] YANG B, YAN J, LEI Z, et al. Aggregate Channel Features for Multi-View Face Detection [C]//IEEE International Joint Conference on Biometrics. Clearwater, USA, 2014: 1–8. DOI: 10.1109/BTAS.2014.6996284
[42] YAN J, ZHANG X, LEI Z, et al. Face Detection by Structural Models [J]. Image and Vision Computing, 2014, 32(10): 790–799. DOI: 10.1016/j.imavis.2013.12.004
[43] MARKUS N, FRLJAK M, PANDZIC I S, et al. Object Detection with Pixel Intensity Comparisons Organized in Decision Trees [DB/OL]. (2013-05-20). https://arxiv.org/abs/1305.4537
[44] LI H, LIN Z, BRANDT J, et al. Efficient Boosted Exemplar-Based Face Detection [C]//IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 1843–1850. DOI: DOI: 10.1109/CVPR.2014.238
[45] LI J, ZHANG Y. Learning Surf Cascade for Fast and Accurate Object Detection [C]//IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA, 2013: 3468–3475. DOI: 10.1109/CVPR.2013.445

Biographies

GUO Da received the B.Eng. from the Computer Engineering College, JiMei University, China in 2018. He is currently a master student at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China. His research direction is face detection and recognition based on deep learning.

ZHENG Qingfang received the B.S. degree in civil engineering from Shanghai Jiao Tong University, China in 2002 and Ph.D. degree in computer science from Institute of Computing Technology, Chinese Academy of Science, China in 2008. He is currently the chief scientist of video technology with ZTE Corporation. His research interests include computer vision, multimedia retrieval, image/video processing, with a special focus on low power embedded application and large-scale cloud application.

PENG Xiaojiang (xj.peng@siat.ac.cn) received his Ph.D. from School of Information Science and Technology from Southwest Jiaotong University, China in 2014. He currently is an associate professor at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China. He was a postdoctoral researcher at Idiap Institute, Switzerland from 2016 to 2017, and was a postdoctoral researcher in LEAR Team, INRIA, France, working with Prof. Cordelia Schmid from 2015 to 2016. He serves as a reviewer for IJCV, TMM, TIP, CVPR, ICCV, AAAI, IJCAI, FG, Image and Vision Computing, IEEE Signal Processing Letter, Neurocomputing, etc. His research focus is in the areas of action recognition and detection, face recognition, facial emotion analysis, and deep learning.

LIU Ming received the M.Sc. degree from Harbin Engineering University, China in 2011. He is currently a senior engineer with ZTE Corporation. His research interests include object detection, tracking and recognition.