Indoor Localization with a Crowdsourcing based Fingerprints Collecting

Yan Huang, Xiaohua Tian
Department of Electronic Engineering, Shanghai Jiao Tong University

Abstract-Fingerprint matching is adopted by a large family of indoor localization schemes to make up for the deficiency of GPS for having better indoor localization, however, there are still two main problems, the first one is about the localization accuracy, the second one is about calculating time. In this article, we give some basic ideas about how to improve the localizing accuracy and we propose a new algorithm to shorten the calculating time.

1 Introduction

With the expansion of the application domain of technology of wireless network, mobile network and pervasive computing, combining the advanced mobile network with location based service is becoming the most prospective and potential business. And the core mission of location based services is to determine the location of somebody or something.

GPS(Global Positioning System) is the most successful localization system till now, however, when it comes to the indoor locations or somewhere the buildings are dense, due to the receiver cannot capture enough strength of satellite signal, the GPS localizing quality cannot be guaranteed, so it's common application is seriously restricted.

So a new method called fingerprint based localization jump into view. The first fingerprint based localizing system were proposed by Bahl and his fellows in 2000, the process is shown in Fig. 1. The process can be divided into two parts: offline process and online process.

1. Offline phase: take several samples in one region and have the samples form the area at a definite density. Have all the available AP(Accessing Point), its MAC address, the signal strength and the location of the sampling point as one as one note in the database, these information are called location fingerprint, which can be shown in Fig. 2.

2. Online phase: to explain this phase in the simplest way, in this phase, when a customer come into a definite region, his mobile devices such as mobile phone or ipad receives the different RSS from different APs, then the devices send the information to the server, then the processor match the location in the database with our RSS and return the location to the customer.

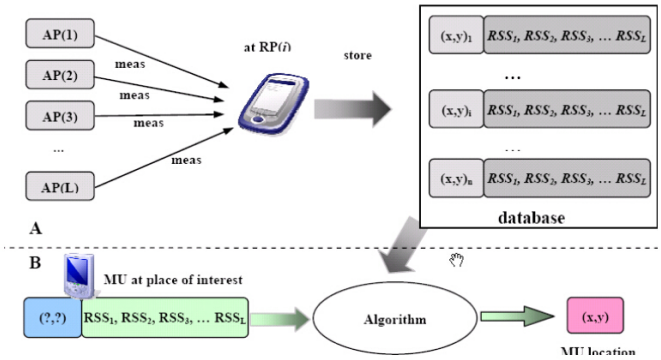

Fig. 1: Fingerprint Based Localizing System[2]

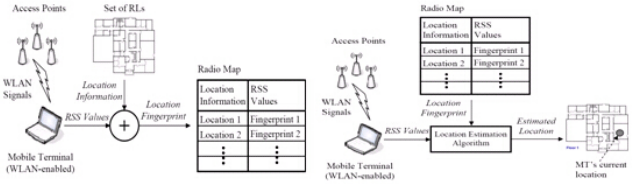

Fig. 2: Offline Phase[2]

2 Previous Work

There are two main problems restricting the fingerprint localization:

1:The location accuracy. The ability of devices to receive signal varies, but the type of mobile devices of customers is always not certain, so the information of RSS may have some error; The RSS of APs are changing overtime, which is also as a factor of error.

2:The calculating time. The calculating time means the time the customer need to wait for matching its information with the database, consider the situation that the customer is walking, if it's calculating at a low speed, then the location appeared may be the previous location of the customer which may also lead in some error.

2.1 Terms and Explanations

AP: Accessing point

RSS: Received Signal Strength

PCA: Principal Component Analysis KNN: K-Nearest Neighbor Algorithm

2.2 To deal with heterogenous devices

In order to get optimum reception fingerprint which is assumed to be more closely with the real fingerprint value for each location, we should first estimate RSS probability distribution having most reality, then choose the one who has maximum probability. As in crowdsourcing indoor location model, many common users will flush into to contribute for constructing the fingerprints database, thus there are sufficient statistic samples to estimate this distribution.

The class-conditional probability of RSS value presenting as fingerprint value can be estimated from uploaded training data in different ways, most previous work treat this probability distribution as a Gaussian distribution, however this estimation method is not appropriate in crowdsourcing indoor location model, for the reason of heterogeneous devices. Probability distribution of different devices is not only different at mean value and variance, but also distribution shapes. Mean value and variance can be transformed across diverse devices through linear transformation, but distribution shape cannot be transformed. Distribution shapes variance across diverse devices is the key challenge resulting in linear transformation not efficient for applying with different types of devices, thus we adopt kernel density estimate[16] method to estimate the class-conditional probability distribution as following equation shown:

$$P_X^k(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-x_i}{h})$$

In this equation, kernel estimate function K(x) uses the standard Gaussian function, and h represents kernel bandwidth. using Kernel density estimation will eliminate distribution shape difference across diverse devices, which hints us to use Kernel density estimation method to estimate RSS probability distribution in crowdsourcing model and further based on this estimated probability distribution to extract fingerprint value across diverse devices.

2.3 To shorten the calculating time

Previous work introduce an algorithm called MMC-KNN algorithm, the author classified all of the points in the database in several parts using specific algorithm which we may not introduce too much here, in a word, the algorithm divide all the points in a floor into M clusters, as is shown in Fig. 3.

The KNN algorithm can be shown in Fig. 4 Then we find K nearest node and make out which cluster the most matched points belongs to, and at last, give the most matched point in a specific cluster back to the customer.

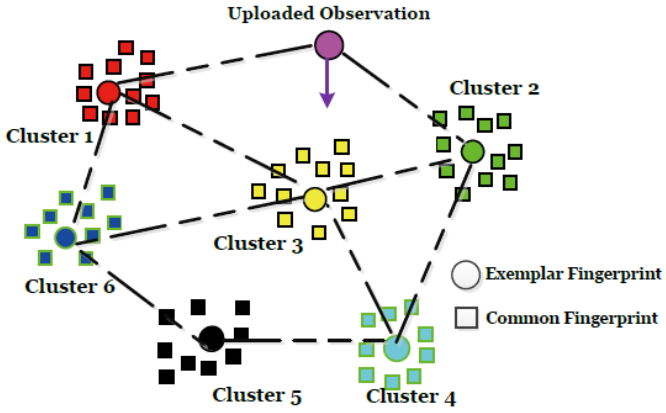

Fig. 3: MMC-KNN algorithm[1]

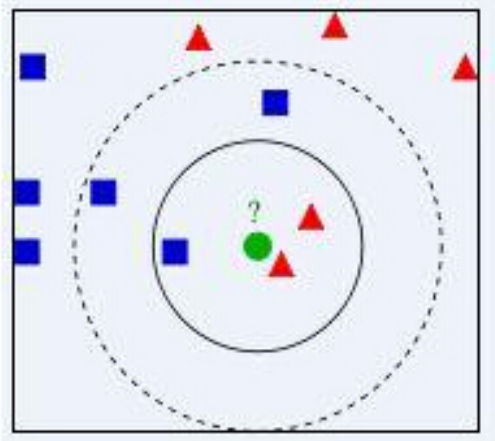

Fig. 4: KNN algorithm[4]

3 My work

3.1 To avoid the changing signal strength

In most of the previous work, they didn't take the variation of RSS overtime into consideration, However, we make a software to detect the RSS overtime and draw a graph as is shown Fig. 5, and we can find out that the RSS is changing over time actually and its swing cannot be omitted. A single signal receiving strength swings

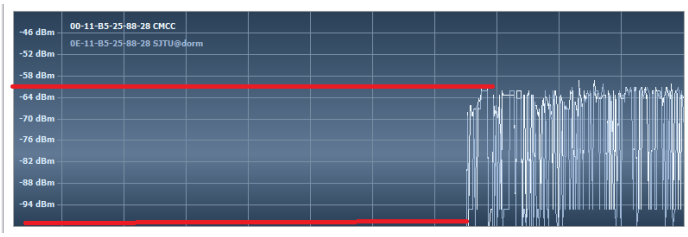

Fig. 5: RSS overtime

from -60dbm -94dbm over a short period of time.

So an overlapping average to decrease fluctuation is needed.

$$AVGRSS(t) = \frac{1}{n}\sum_{i=0}^{n-1} RSS(t-i)$$

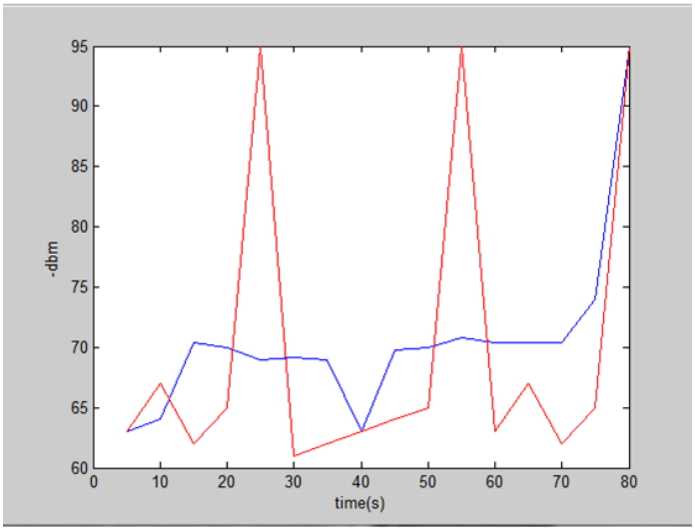

Fig. 6: Comparison between overlapping and un-overlapping singal

Easy to find in Fig. 6, the RSS after overlapping is much more smooth.

3.2 For shortening calculating time

Now we introduce a new algorithm to deal with the problem of finding the most matched point in one cluster in a short time. We use PCA algorithm which is mostly used in image processing, however, the called "fingerprint" really have some thing to do with image processing, because the RSS in one cluster are highly related, so we can use the RSS in one point as one "fingerprint", and then find the most matched "fingerprint" in the database. In this process, we covert the digit arithmetic into matrix arithmetic.

3.2.1 Step 1

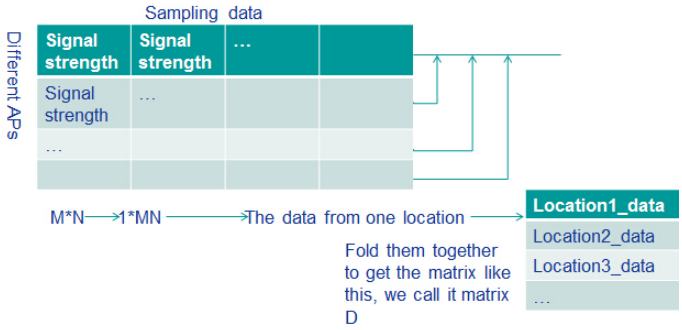

Fig. 7: Comparison between overlapping and un-overlapping singal

Like the graph of Fig. 7,At first, we save information of one point in one matrix, and we convert the matrix into a one-line matrix, and we put the information of the second point into the second line, then we save the information of all the points in one cluster in a matrix in this order.

3.2.2 Step 2

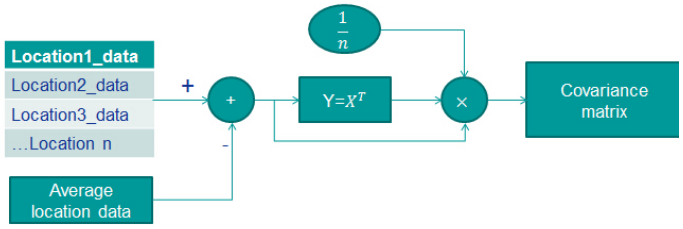

Fig. 8: Comparison between overlapping and un-overlapping singal

Like the graph of Fig. 8, some mathematic processing are finished, through the matrix calculation above, we get the covariance matrix which services for the next step.

3.2.3 Step 3

Fig. 9: Comparison between overlapping and un-overlapping singal

Like the graph of Fig. 9, after getting covariance matrix, we calculate the eigenvalue and eigenvector of the matrix, then we introduce the following formula as the mapping function from the information of points to the eigen dimain.

$$U_i = \frac{1}{\sqrt{L_i}} DV_i (i = 1, 2, ...r)$$

Now we have the mapping function and we can project the database to the eigen domain, and when a customer get into a specific area and receives some RSS from APs, it projects to the eigen domain too, through comparison in the eigen domain, we can largely shorten the calculating time because in the domian we can also omit some redundancy and unnecessary information while keeping the rate from dropping too much.

4 FUTUREWORK

In conclusion, I find an algorithm to match the RSS to the database consuming less time and I make some efforts to deal with the time varying signals. And now some work are still waiting to be completed.

1.Write an android program to test my work

2.Use a continuous series of samples(a cycle) instead of a point in the database to deal with varying signals.

5 REFERENCES

[1] Indoor Localization with a Crowdsourcing based Fingerprints Collecting-Zhengyong Huang

[2] Real time f ingerprint positioning algorithm based on spatial diversity and trajectory continuity-LIU Xingchuan, ZHANGSheng, XU Liqiang, LIN Xiaokang

[3] WLAN localization-China mobile communications co., LTD., institute of business

[4] KNN algorithm-http://www.baike.com/wiki/KNN