WLAN Indoor localization

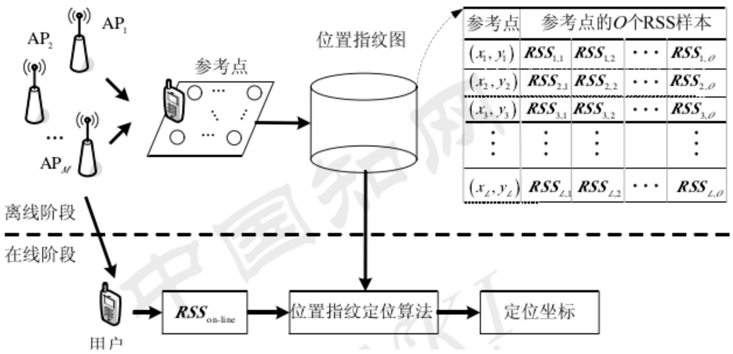
Zhang Yifan 5130309525

Abstract: With the rapid development of wireless local area network (WLAN) technology, an important target of indoor positioning systems is to improve the positioning accuracy while reducing the online calibration effort to overcome signal time-varying. A novel fingerprint positioning algorithm, known as the adaptive radio map with updated method based on hidden Markov model (HMM), is proposed. It is shown that by using a collection of user traces that can be cheaply obtained, the proposed algorithm can take advantage of these data to update the labeled calibration data to further improve the position estimation accuracy. This algorithm is a combination of machine learning, information gain theory and fingerprinting. By collecting data and testing the algorithm in a realistic indoor WLAN environment, the experiment results indicate that, compared with the widely used K nearest neighbor algorithm, the proposed algorithm can improve the positioning accuracy while greatly reduce the calibration effort.

Key words: HMM, wireless, accuracy

Introduction:

The emergence of wireless technology and mobile computing devices has promoted business development of various location estimation systems. In recent years, many researchers focus on offering perfect services by taking fully advantage of the mobility of the devices. Therefore, the location based services (LBS) have become a hot topic of mobile computing research area. In general, the mobile set needs to locate itself before enjoying the service based on location. It is well known that the existing global positioning system (GPS) may provide a simple and effective solution for such applications in the outdoor environment. In spite of the fact of GPS playing a vital role nowadays, limitations still exist. The main shortage is the invalidity to perform positioning inside the buildings due to the signal blockage in the indoor environment. Thus how to determine the exact location of a mobile device in buildings is still a hot issue to be solved.



According to the difference of positioning principles, indoor positioning technologies can be roughly classified into three types: the positioning technologies based on hardware device, the positioning technologies based on the signal transmission model and the positioning technologies based on the received signal strength indication (RSSI). The positioning technologies based on hardware device usually need certain special auxiliary equipment, and calculate the physical location according to the information from the equipment, which easily leads to a cost increase and process complication. The positioning technologies based on the signal transmission model calculate the distance between the sender and the receiver by calculating the time of signal transmission. Thus, it can get the receiver's physical location using a known emission source. Its limitation is that the signal does not always transmit along a straight line, because there are all kinds of obstacles between signal emission sources and receivers, which leads to complex reflection, refraction, diffraction and inaccurate positioning. The whole process of the positioning technologies based on RSSI can be divided into two phases: the training phase and the positioning phase. During the training phase, it is needed to collect RSSIs of access points (APs) with certain physical addresses at every reference point (RP), which will be stored in order to build a fingerprint database (FD). During the positioning phase, it is also needed to collect RSSIs to match information in the FD for the purpose of getting the optimum information and realizing the localization. The representative RSSI-based positioning technologies are based on the nearest neighbor (NN) algorithm or the k-nearest neighbor (KNN) algorithm, which are easily affected by the multipath effect. To sum up, the performance of indoor localization technologies still has a huge improvement space, especially in positioning accuracy, environmental adaptability, hardware cost, and energy consumption.

Problems analysis:

Firstly, we implement several rounds of experiments in the real indoor environment to collect and analyze data about the signal fluctuations at the reference position caused by multipath effects. During the process of data collection, the factors of object moving and blocking are fully taken into account to make the test more realistic and credible. The data collection, measurement and analysis are mainly based on wireless routers as APs and mobile terminals with the Wi-Fi function. The mobile terminals can move freely, and scan the RSSI values of APs continuously. Three real indoor environments are constructed with Wi-Fi access points deployed. Two of them are quadrate rooms, and another one is a long and narrow corridor of a building. We collect data of signals in both scenarios, and analyze the relationship between mean values of signals and variances of RSSI. We tested how some orientation affects the RSS readings due to blocking and reflection of radio signals by human bodies, including user's facing orientation, holding phone positions, number of samples taken at each location and time of the day. Experimental results in show that the positioning result varies with these factors changing. In our experimental environments, five APs are deployed and the fingerprint database is built. The distances of RPs are set 0.6 m and 1 m for testing positioning results. Each RP collects 300 values of RSSI during the training phase. During the following data analysis process, eight positioning test points are taken to calculate variances and mean values. At each point, we collect 20 sets of data. We also take the time of the day into consideration and implement experiments in different time of the day, as the RSSI varies with the time changing.

This is the software we used.

序号	坐标	p2psearch	403	FAST	DEJURE	DORM503
1	(0.3,1.2)	-27	-54	-62	-66	-67
2	(0.3,2.8)	-42	-25	-74	-69	-64
3	(0.6,0.9)	-24	-57	-57	-58	-64
4	(2.4,6.6)	-54	-66	-75	-73	-73
5	(0.1,6.6)	-64	-59	-62	-75	-75
6	(0.1,-1.5)	-47	-60	-68	-79	N/A
7	(3.2,4.2)	-48	-60	-75	-79	-75
8	(2.4,3.2)	-41	-45	-65	-60	-66
9	(1.5,1.7)	-28	-45	-64	-62	-68
10	(1.8,4)	-45	-51	-72	-64	-65

This is the primitive work. Which shows the difficiency of the job.

According to the analysis above, we can draw the conclusion that the variances of RSSI vary with the differences of RSSI values and distances. If a mobile terminal is far from an AP, the measured RSSI at that point is low, has a large variance and becomes unreliable. If KNN is used to calculate fingerprint distances, and different RSSIs will be given different confidence based on the measured mean values and variances, the effect of large variance may be reduced.

Fingerprint analysis:

The KNN-based localization algorithm is based on RSSI. In the training phase, RSSIs of APs are collected and stored in the fingerprint database. In the positioning phase, Euclidean distances between RSSIs stored in the database and RSSIs just measured will be calculated, and adopted as fingerprint distances. Then the smallest k values of fingerprint distances will be chosen with their corresponding coordinates to calculate the average value as the positioning result. The KNN-based localization algorithm defines the fingerprint distance as:

$$D(j) = \sum_{i=1}^{N} (r_i - s_{ij})^2$$

where N is the total number of AP stations, ri is the ith signal value measured in the positioning phase, sij is the ith AP’s RSSI at the jth RP, and D(j) is the fingerprint distance between RSSIs measured in the positioning phase and the training phase at the jth RP.

From the fingerprint distance definition of KNN, it is easy to find that all RSSIs collected are with the same configuration. There is obviously a shortcoming that if an RSSI collected is distant to its expected mean value due to the multipath effect, the final fingerprint distance will be larger, but other collected values are still normal. This occasional error will make KNN unable to choose the optimal k RP to achieve accurate localization.

Since the error of measurement cannot be avoided completely, we need to reduce the effect of error measurement and rely on these RSSI values close to those values collected in the training phase and stored in the fingerprint database to make positioning results accurate. Due to the reason above, we define a new improved fingerprint distance as follows:

$$D(j) = \sum_{i=1}^{N} d_i (r_i - s_{ij})^2$$

where di is the adjusting weight assigned to each measured RSSI value ri. Credibility is introduced here for evaluating the credible level of the accuracy of fingerprint distance. When calculating fingerprint distance, it is necessary to assign high credibility to accurate measured RSSI values and low credibility to inaccurate ones. Thus, the positioning results can be adjusted effectively. We have two key problems to deal with: one is how to judge whether a measurement is accurate or not; the other is how to get the optimum adjusting weight to efficiently reduce the error of fingerprint distance.

Error analysis:

Measurement errors can be found during the positioning phase. However, the level of error is not given directly by the measurement value itself. So we need to get a large amount of detailed information from the fingerprint database. As the above experiment analysis shows, there exists a linear relationship between the variance and the mean value as

$$Var = Mean \times a + b$$

where *Var* is the variance of measurement value, *Mean* represents the mean value of measurement, and a, b are parameters of the linear relationship equation. In the positioning phase, measurement values are calculated with to get *Var*.

Adjusting weight:

The basic idea of adjusting weight is to assign higher weights to measurements with smaller errors, and assign lower weights to those with bigger errors. So we should separate RPs close to APs from those distant to APs in order to choose the k-nearest RPs more correctly. According to the relationship between the variance and the mean value of RSSI, variances tend to be large while RSSI mean values are small, and can be calculated with (3). In order to reduce errors caused by large variances, small weights should be assigned to them, and the reciprocal of variances can be used as the weight. It is necessary to point out that although there is only variances in (4), the calculation of variances needs mean values. Though the calculation of variances used the approximate linear relationship and the variance cannot be exactly accurate, this method is surely qualitative to reach the target of adjusting error. The calculation method of adjusting weight is as follows:

$$d_i = \frac{1/Var_i}{\sum_{i=1}^{N} 1/Var_i}$$

where Var_i is the variance of RSSI measured from the ith AP, and the total weight should satisfy

$$\sum_{i=1}^{N} d_i = 1.$$

Process:

The mobile terminal scans and collects signals from N APs at any RP within the positioning area, gets several sets of RSSI values, and then gets the mean value of each RSSI. The process of VFDA includes the following steps:

Step 1 Estimate the variances of RSSIs with the mean values of measurement.

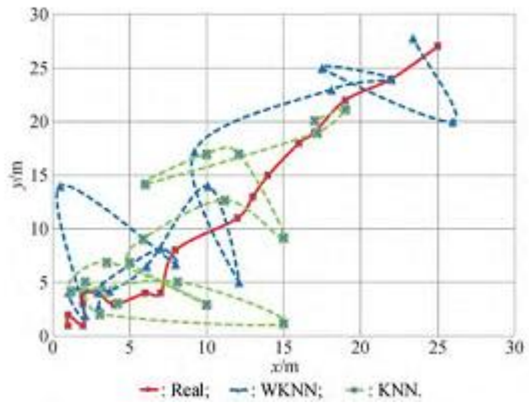
Step 2 Calculate the adjusting weights.

Step 3 Calculate the fingerprint distances.

Step 4 Obtain the physical location of the mobile terminal.

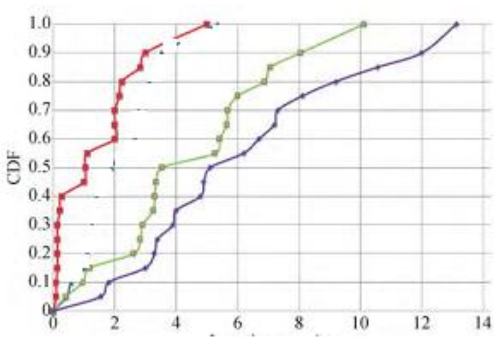
To sum up, we obtains fingerprint distances via calculating the distance, uses arrays to store parameters a and b of each AP, and travels through the list of the mean values of signals, multiplied by the adjusting weights, and finally gets the corrected fingerprint distances.

Results:



Then we need to calculate the distance error:

$$error = \sqrt{(x_l - x_r)^2 + (y_l - y_r)^2}$$



Distance error

Conclusion:

Unstable RSSI values caused by the multipath effect lead to the inaccuracy of indoor localizations. In order to solve this problem, we propose a method based on the relationship between variances and mean values to adjust fingerprint distances. From the experimental results and the analysis, we find out that the mean values increase with the variance decreasing, which means that the mean values of RPs far from APs are smaller than the mean values of RPs close to APs. From the further study on the relationship between the mean value and the value for certain AP, we find out that this relationship is approximately linear.