Wireless indoor localization based on multispectral waterfall maps

Tianyu Zhang[*1], *Yifan Jia*[2], *Weiwei Jiang*[3], *Junnan Gao*[4], *Sijie Yan*[5], *Huadong Meng*[6]

Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
kevinzhangty@gmail.com, menghd@tsinghua.edu.cn

Abstract

Wireless indoor localization is becoming increasingly important with continuing technological advancements and the new demands of the business model. This paper proposes a wireless localization method based on the waterfall maps of GSM(Global System for Mobile Communications) spectrum and Wi-Fi spectrogram. The method adopts the naive Bayes classifier to classify various locations based on the frequency characteristics of those locations. The data in different frequency bands can be combined to perform grid anchor positioning based on frequency characteristic classification. Meanwhile, the indoor experiment with an anchor spacing of 1.2 m achieves a localization accuracy of over 90%. The experimental results demonstrate that combining the frequency bands of phones and Wi-Fi can be an effective strategy for indoor wireless localization.

1. Introduction

Numerous studies on wireless localization have been performed in recent years. With the development of mobile Internet, localization has played a vital role in location-based services. Indoor localization has become a basic technology for indoor navigation in the commercial district and enterprise mobility [1,2]. Indoor localization methods can be divided into parametric and non-parametric methods. Parametric methods, such as TOA, TDOA, and AOA, are based on the parameter channel model for localization [3]. However, these methods are not highly adaptable to the environment. Indoor localization methods based on channel transmission models are not effective because the indoor environment is complex due to reflections and other effects on radio signals between multifaceted barrier walls. Non-parametric methods are based on the spatial positioning of characteristics of fingerprint information. Non-parametric methods include fingerprint image localization technology and indoor location sensing using active RFID [4, 5]. Non-parametric methods also have several problems, including issues with mismatch, the information utilization rate, and redundant channel state information in the fingerprint database. Numerous localization engineering systems have been developed using non-parametric radio. E Martin from UC Berkeley proposed a positioning system based on the spectral characteristics of multi-format signal space, achieving a precision of approximately 2 m [2]. Park J who designed a localization method based on the Voronoi graph method, developed a system to verify the algorithm [6]. Qun Wan summarized the indoor localization systems of the current stage [7]. However, these systems do not efficiently use the spectral information of different frequency bands at different times, preventing the localization accuracy from meeting the actual requirements. The experimental system in the paper achieves higher localization accuracy than the existing methods by using the Bayes classification method according to the GSM(Global System for Mobile Communications) spectrum and Wi-Fi spatial fingerprint characteristics.

The second part of this paper describes the principle using naive Bayes classification and data fusion theory [8]. The application of the spectrogram in localization is also described. The third part describes the experimental environment, the laboratory equipment, the experimental process, the experimental results under various conditions, and an analysis of those results.

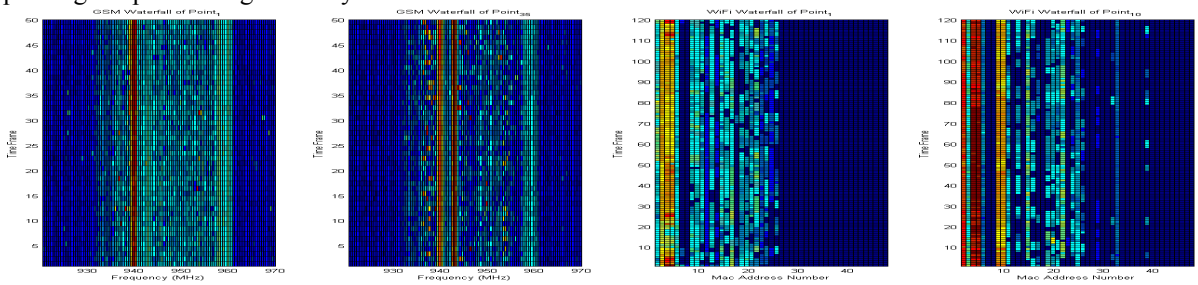
2. Principle of Fingerprint Localization

Radio signals are affected by the complexity of the indoor environment due to blocking and reflection by multiple walls. As a result, indoor positioning systems based on the channel transmission model do not perform well. In this paper, we classify the locations according to the features (the fingerprint) of each frequency at various locations for positioning. The method proposed in this paper can be divided into two processes: generating the training dataset and processing the test data. First, we gather the data for each frequency at different positions to build a fingerprint database. Based on these data, the distributions of the frequency characteristics are set at different locations using the naive Bayes method. Frequency characteristics distributions are used to classify the frequency characteristics of the data under positioning to determine the positioning. This paper conducted a localization test at the selected interval anchor point but not at other positions. However, this work can be regarded as an important foundation for continuous positioning because the anchor point is designed to be intensive (the distance between two adjacent anchors is 1.2 m).

Frequency characteristics with anisotropic space, which can be used for classification, are stable over time. According to the measurement results, the strength of different frequencies at the same location remains stable for a certain duration. To achieve a higher level of accuracy, the device repeats the test within a short period in time at the same location

978-1-4673-5225-3/14/$31.00 ©2014 IEEE

to obtain data frames, which record the strength of signals at different frequencies. (For example, in the experiment, one frame of training data is collected every 25 ms, with 20 frames for each location, and one frame of testing data is collected every 25 ms, with 180 frames for each location.) The information in different frames would be gathered to form the waterfall map. For the GSM signal (Figure 1 (a-b)), the horizontal axis of the map (Figure 1 (c-d)) represents frequency, and the vertical axis represents time; for the Wi-Fi signal, the horizontal axis of the graph represents the different AP, and the vertical axis represents time. Each row represents the energy value of different frequencies at a specific time, and each column represents the energy value of a specific frequency over a short period of time. Each location corresponds to a spectrum waterfall map that includes the fingerprint information of each frequency. The waterfall map is a vector that changes over time. Therefore, repeated measurements will reduce the disruption caused by the randomness of the strength, thus improving the positioning accuracy.



(a) (b) (c) (d)

Fig. 1 Waterfall Maps of GSM and Wi-Fi spectrum signals

After a training set is obtained, the strengths of the GSM and Wi-Fi signals are used as the feature of training set $A_1, A_2, \ldots, A_{M_1}, B_1, B_2, \ldots, B_{M_2}$. Under the assumption that the prior probability distribution is a Gaussian distribution $X \sim N\left(\mu, \sigma^2\right)$. The location and scale parameters can be estimated by the sample means and variances respectively. Based on the probability density function, the features of training set $A_1, A_2, \ldots, A_{M_1}, B_1, B_2, \ldots, B_{M_2}$, which belong to different Bayes classifiers, will be exported to positioning in the Bayes method (see Fig. 2) . Now we describe the localization algorithms used in the paper. The proposed method uses the naive Bayes classifier to match fingerprints. Bayes classifiers are widely used in classification problems. To construct a special Bayes classifier with a given set of pre-classified instances, assume $A_i\left(i = 1, 2, \ldots, M_1\right), B_j\left(j = 1, 2, \ldots, M_2\right)$ are features that take values $a_i\left(i = 1, 2 \ldots, M_1\right), b_j\left(j = 1, 2, \ldots, M_2\right)$ respectively. Those features will be used to predict the value c of class C. According to the Bayes rule, the probability that an instance is in class C is as follows:

$$P_{C|A_1,\ldots,A_{M_1},B_1,\ldots B_{M_2}}(c|a_1,\ldots,a_{M_1},b_1,\ldots b_{M_2}) = \frac{P_{A_1,\ldots,A_{M_1},B_1,\ldots B_{M_2}|C}(a_1,\ldots,a_{M_1},b_1,\ldots b_{M_2}|c)P_C(c)}{P_{A_1,\ldots,A_{M_1},B_1,\ldots B_{M_2}}(a_1,\ldots,a_{M_1},b_1,\ldots b_{M_2})}$$

where A represents the frequency of GSM and B represents the Wi-Fi access point (AP). The naïve Bayes model assumes that all attributes are independent within a given class, i.e.,

$$P(a_1,\ldots,a_{M_1},b_1,\ldots,b_{M_2}|c) = \prod_{j=1}^{M_2}\prod_{i=1}^{M_1} P(a_i|c)P(b_j|c) \tag{1}$$

For any i and j, $P(a_i|c) = N\left(a_i, \mu_{A_i}^{(c)}, \sigma_{A_i}^{2(c)}\right)$, $P(b_j|c) = N\left(a_j, \mu_{B_j}^{(c)}, \sigma_{B_j}^{2(c)}\right)$. The naive Bayes classifier predicts an instance using the formulation: $\underset{c \in \Omega_c}{\arg\max} P(c)\prod_{j=1}^{M_2}\prod_{i=1}^{M_1} P(a_i|c)P(b_j|c)$ .Given a list of the strengths of the training points $S_{m,n}^{(l)}$, $1 \le m \le M, 1 \le n \le N, 1 \le l \le L$, where m is the index of the characteristics, M is the capacity, n is the index of discrete time, and N is the amount of discrete time for training, l is the order of the location and L is the total number of locations. Given a list of the strengths of the testing points $t_{i,j}^{(l)}$, $1 \le i \le M, 1 \le j \le P, 1 \le l \le L$,s the index of the characteristics, M is the total number of characteristics, j is the index of discrete time, P is the amount of discrete time for testing, l is the index of the location and L is the total number of locations.

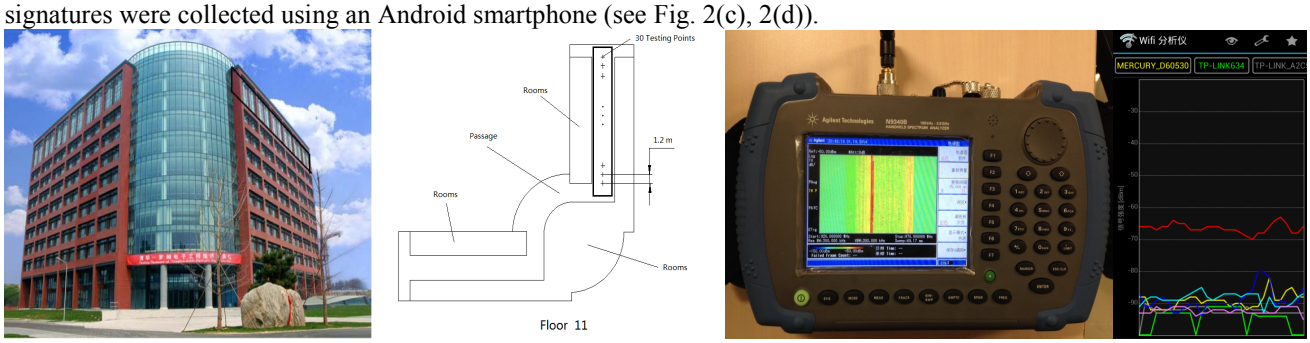
Considering that $t_{i,j}$ can be combined by different series of characteristics $t_{i,j}^{(S_1)}$, $t_{i,j}^{(S_2)}$. S_k denotes the series of characteristics, where $1 \le k \le Q$ and Q is the quantity of the series of characteristics. Our algorithm classifies the instances using the following formulation:

$$P\left(\mathrm{LOC}=l\left|\left\{\mathbf{t}_{\mathbf{i,j}}^{(s_k)}\right\}_{1\le i\le M,1\le j\le P,1\le k\le Q}\right.\right)=c_1\cdot P\left(\left\{\mathbf{t}_{\mathbf{i,j}}^{(s_k)}\right\}_{1\le i\le M,1\le j\le P,1\le k\le Q}\left|\mathrm{LOC}=l\right.\right)\cdot P\left(\mathrm{LOC}=l\right)=c_2\cdot\prod_{k=1}^{Q}\prod_{j=1}^{P}\prod_{i=1}^{M}P^{(s_k)}(\mathbf{t}_{\mathbf{i,j}}|\mathrm{LOC}=l)$$

where LOC is the location to be estimated, and c_1, c_2 are constants. Assume that the existing probability in each location is equivalent, and then the time frames of testing data are the same for different signal frequency bands. $P^{(s_1)}\left(\mathbf{t}_{\mathbf{i,j}}|\mathrm{LOC}=l\right)$ and $P^{(s_2)}\left(\mathbf{t}_{\mathbf{i,j}}|\mathrm{LOC}=l\right)$ respresent $P\left(a_i|\mathrm{c}\right)$ $P\left(b_j|\mathrm{c}\right)$ in Eqn. (1) respectively. In this paper, GSM signal and/or Wi-Fi signal are used. If one signal band is used, then Q=1. If two signal bands are used together, Q=2.

3. Experiments

To evaluate the proposed approach , we conducted experiments on the 11th floor of the Electronic Engineering Department building of Tsinghua University, Beijing, China (see Fig. 2(a), 2(b)). GSM radio signatures were collected by an Agilent N9340B Handheld RF Spectrum Analyzer produced by Agilent Technologies in Chengdu, China. Wi-Fi signatures were collected using an Android smartphone (see Fig. 2(c), 2(d)).



(a) (b) (c) (d)

Fig. 3 Experiemental environment and facilities

We profiled 30 anchor points on the 11th floor of the building, as shown in Fig. 2, with an anchor spacing of 1.2 m. Using the proposed device driver and API, we collected 20 samples of the Wi-Fi and GSM signatures at each location, which are used as the training set, with 4 samples per millisecond. To test the system performance, we used 180 samples as a test set that was collected after a certain duration. The determination parameters must be defined to analyze the experimental results. The accuracy of the single frame is the ratio of the correct frames to the total testing frames at the same anchor point. The accuracy of the dual-band integration is the result using the combination of Wi-Fi and GSM signals in the form of a distribution. The localization accuracy is the ratio of the number of correct judgments to the total number of judgments, where the predicted value that appears most often is regarded as the judgment, using the plurality rule at an anchor point. Table 1 provides the statistical results of the judgements of the single frame using RSSI values for both GSM and Wi-Fi signals and dual-band integration signals by Bayes methods.

Table 1 Statistical results of the judgments of the 180 frames in the test set.

LOC	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
GSM	111	147	174	120	144	143	142	135	121	40	146	152	122	63	122
Wi-Fi	145	146	177	152	163	138	144	64	179	134	76	176	148	144	145
GSM+Wi-Fi	144	156	179	169	170	170	170	82	167	138	76	175	161	161	177
LOC	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
GSM	108	138	118	138	138	112	148	99	77	134	84	120	83	149	70
Wi-Fi	153	175	125	178	145	140	171	169	124	179	113	176	131	158	155
GSM+Wi-Fi	166	176	157	171	158	161	176	179	133	176	132	154	135	170	173

Table 2 Localization accuracy judged by the plurality rule.

Frame	Duration	GSM	Wi-Fi	Dual-band integration
45	11ms	92.5%	94.17%	95.83%
60	15ms	94.44%	92.22%	96.67%
180	45ms	96.67%	96.67%	96.67%

4. Analysis and Discussion

The results in Tables 1 and 2 illustrate that the three localization methods each reach an accuracy of nearly 90%. At a single position, the accuracies of the Wi-Fi-based localization at Points 3,6,8 and 11 are lower than those of the GSM-based localization. The Wi-Fi based localization produces errors at Points 8. These errors may be due to the instability of portions

of the Wi-Fi AP, which cause the similarity between Points 8 and 9. Dual-mode localization based on joint data is used to increase the probability of accurate judgment and thus increase the accuracy.

According to the results above, the accuracy of localization increases at certain positions when using joint data, whereas it decreases at other positions. The following rule is established to restrict the occurrence of such mistakes: if the localization results based on the GSM signals and Wi-Fi signals are consistent but differ from the result based on the combination of both types of signals, the results of the non-combined data should be used. Such a rule can optimize the localization. For example, when using 20 frames of GSM signals and 10 frames of Wi-Fi signals as training set, after implementing this correction, the accuracies of the localization at Points 1, 2, 9, and 26 are all improved. As an example, at Point 2, the accuracy is 61.67% without this rule, which is lower than the 81.67% accuracy obtained using the GSM signals and the 63.89% accuracy obtained using the Wi-Fi signals. After this rule is implemented, the accuracy increases to 63.89%, which is the same as the accuracy obtained using the Wi-Fi signals. The rule above can prevent misjudgments due to joint data and increase the accuracy of the localization.

In addition to the naive Bayes classification, the KNN algorithm (based on Euclidean distance k=3) and the algorithm based on the Bhattacharyya distance have been considered, and the results of these three methods are compared (See Fig. 5). The comparison indicates that the naive Bayes classification achieves a higher level of accuracy than the other two methods.

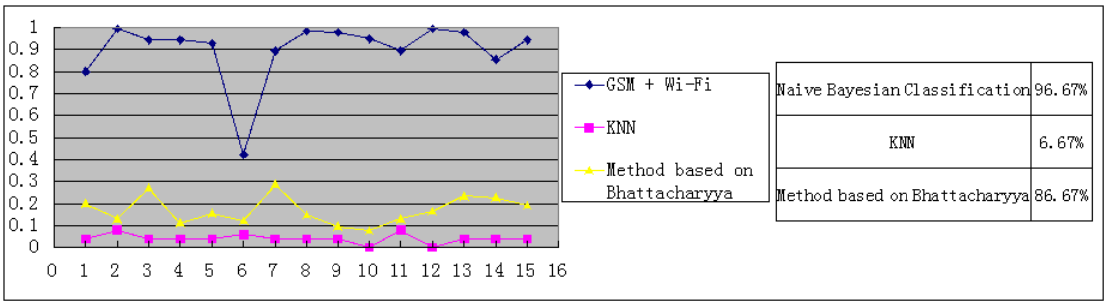


Fig. 5 Comparison of accuracies of the three algorithms in terms of success percentage

5. Conclusion

By using the distrbutions of frequency characteristics obtained by training data, naive Bayes classification can be applied to feature classification and thus to localization. Based on the naive Bayes classification, GSM-based localization, Wi-Fi-based localization, and joint localization are performed, with accuracies in terms of success percentage of 66.85%, 83.52%, and 87.52% respectively in the single-frame analysis, and 96.67%, 96.67%, and 96.67% respectively in the multi-frames analysis. When combining GSM signals and Wi-Fi signals for joint analysis, even though the accuracy in terms of success percentage decreases at several positions, it increases at the majority of positions (66% positions). The experiments demonstrate that highly accurate classifying localization can be achieved by adopting naive Bayes classification with a dual-mode joint distribution as the likelihood function.

Due to the obvious redundancies of spectrum data, in further work, attempts can be made to extract additional distinguishable features to reduce the amount of data and achieve a more effective localization.

References

1. Do, J., Matthew Rabinowitz, and Per Enge, "Performance of hybrid positioning system combining GPS and television signals," *Position, Location, And Navigation Symposium, 2006 IEEE/ION*. IEEE, 2006: 556-564.
2. Martin, E, "Multimode radio fingerprinting for localization," *Radio and Wireless Symposium (RWS), 2011 IEEE*. IEEE, 2011: 383-386.
3. Wong, Carl Monway, "Investigation of wireless local area network facilitated angle of arrival indoor location," University of British Columbia, 2008.
4. Kang Hongwen, Efros, A.A., Hebert, M, Kanade, "Image matching in large scale indoor environment," *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*. IEEE, 2009.
5. Ni, L. M., Liu, Y., Lau, Y. C., & Patil, A. P. "LANDMARC: indoor location sensing using active RFID."*Wireless networks* 10.6 (2004): 701-710.
6. Park, Jun-geun, et al, "Growing an organic indoor location system," *Proceedings of the 8th international conference on Mobile systems, applications, and services*. ACM, 2010: 271-284..
7. Wan Q. Indoor Localization. China: Publishing House of Electronic Industry, 2012.
8. Friedman, Nir, Dan Geiger, and Moises Goldszmidt, "Bayesian network classifiers," *Machine learning* 29.2-3 (1997): 131-163.