Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

ResidualRanking: a robust Access-Point selection strategy for indoor location tracking

MinWang
Department of Computer Science and Technology
Harbin Institute of Technology Shenzhen Graduate School
Shenzhen, China
wangminhitsz@gmail.com

*ChunkaiZhang
Department of Computer Science and Technology
Harbin Institute of Technology Shenzhen Graduate School
Shenzhen, China
ckzhang@hotmail.com

***Abstract*—In this paper, we propose a robust approach to access point (AP) selection problem for the indoor location tracking. It takes the environments changes into account and makes use of residuals ranking algorithm to select those APs least sensitive to the environment changes in indoor location tracking, we call it ResidualRanking method, also we make an improvement of residual computing according to the properties of radio signals. Additionally, we present a location tracking system called BRR (Bayesian and Residuals Ranking) which is based on the Bayesian decision method and the ResidualRanking method we proposed. Finally, we make a comparison to the MaxMean AP selection method, and the experimental results indicate that the ResidualRanking method we proposed can achieve a better performance than the MaxMean method, also the proposed system BBR can get desirable results in the realist indoor location tracking.**

***Index Terms*—location, fingerprint, access-point.**

I. Introduction

As the ubiquitous computing becomes more popular, the importance of context-aware applications increases. This finally fuels the need to get the user's location estimation, with which the system can provide location-based services. On the other hand, the development of wireless technology and mobile computing devices also has promotes the growing commercial and research interests in developing various location estimation systems. Thus, many systems over the years have tackled the problem of determining and tracking the user position. The most popular ones are GPS and GSM, they are the main techniques used to deliver Location-Based Services (LBSs) in outdoor environments [7], and they have been developed maturely. However, the GPS and GSM system are not designed for the indoor environments where the environment is more complicated. Compared to the outdoor location techniques, the location estimation techniques in indoor environments are lagging behind. But the LBSs are of equal interest in indoor environments in a wide range of personal and commercial applications, such as location-based network management and security, medicine and health care, personalized information delivery, and context awareness. Thus the indoor location estimation is becoming a hot issue in recent years.

In indoor environments, much effort has been focused on the development of Radio Frequency (RF)-based location-estimation techniques using Received Signal Strength Indication (RSSI) by making use of popular infrastructures such as the IEEE 802.11b Wireless Local Area Networks (WLANs), and Radio Frequency Identification (RFID) based networks [12].

At present, fingerprint is becoming a popular method used in RF-based indoor location estimation [5]. Due to the complexity of the indoor environment, the signals in indoor environment is vulnerable to the noise and non-line of sight propagation (NLOS), thus the signal strength in indoor environment is not decays with log distance, and the geometry methods based on the propagation model such as the lateration and angulation cannot be applied in indoor location estimation. However, the fingerprint method making use of the fingerprints can get a desirable result in this complex indoor environment. The fingerprint method consists of an offline training phase and online location phase. In the offline phase, a radio map representing the RSSI distribution in location area is built. In the online phase, we make use of the real-time RSSI and the learned radio map to estimate a user's current location. However, its performance is often influenced by the lack of up-to-date information about the environment, and the fact that constructing the radio map is very time-consuming [8].

With regard to the architecture of location system, it can be mainly divided into server-based and client-based location systems [9]. As client-based location system is effective for preserving the privacy of clients, it is widely used in indoor location systems. However, client devices are usually small, maintained by constrained battery power, then a challenging problem is how to reduce the online computation of positioning while achieving a high accuracy at the same time. As the mobile software continues to grow up in complexity and power demand increases, it is critical to reduce the computational burden in the client side [3]. To tackle this problem, an access point (AP) selection method is needed. Suppose the set of APs covering the survey points is denoted as A and $|\mathrm{A}| = L$, the objective of AP selection is to determine a set A' needed in the positioning phase such that $|\mathrm{A}'| = d < L$.

At present, there are several ways to make AP selection, Youssef et.al bring up the MaxMean method [13], it rank APs in descending order of their average signal-strength values, and select the k strongest APs. Chen et.al's Info Gain method [1] works as selecting the most discriminative APs using the information entropy. Fang et.al [3] bring up

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

a novel concept, that is to make location fingerprinting in a decorrelated space making use of PCA, and it not only can reduce the computational complexity, but also reserving all the APs's information. Kushki et.al demonstrate the need for AP selection,and highlight the interplay among this component and the distance calculation step, and encourage future research in the area, its AP selection method is based on the principles of minimizing correlation between selected APs [6].

All the AP selection methods mentioned above are carried out offline based on the training data. This hampers the operation of the system over time since indoor environment are highly dynamic and APs can be easily moved or discarded. On the other hand, due to the dynamic properties of the indoor environments, some APs selected in the offline phase probably make a side effects to the accuracy in the online phase. For example, at some time, a barrier exists in the front of APi , and this also can be considered as that the radio map we got in the offline is outdated, so if we still use APi to make the estimation, the accuracy may be disturbed. To deal with this situation, here we propose a novel APs selection strategy that makes use of the residuals ranking method and works only in the online phase to select those APs least sensitive to the environment changes in indoor location tracking. We call this AP selection method ResidualRanking. Additionally, we present a BRR (Bayesian and Residuals Ranking) indoor location system which is based on Bayesian decision method and ResidualRanking method. Besides, the radio map is constructed in statistical method. The novelty of our works can be summaries as follows:

1) Compared to the previous AP selection methods carried out in the offline phase, we select APs for positioning in the online phase, that is more adapt to the indoor environment and makes the estimator more robust.
2) Our BRR location system can get a high accuracy while in a low computation by making use of the radio map constructed in statistical method .

The rest of this paper is organized as follows: section 2 introduces the Bayesian decision method used in location area. In section 3, we makes a detail description of the rational of our AP selection strategy. Section 4 makes a detail description of the BRR system we proposed. Section 5 presents an experiment and the results analysis. Section 6 concludes this paper and discuss our future work.

II. Bayesian decision method

Several statistical learning theory have been applied to the fingerprint techniques, such as Weighted K-nearest-neighbor (WKNN), Maximum Likelihood (ML), Multi-Layer Perceptron (MLP), Neural Network, and Support Vector Machinies (SVMs). In this paper, we adopt Bayesian decision method to make the location estimation. Bayesian decision method treat RSSI as the random variables, which are statistically dependent on the location and it employs Bayesian rule to make the location estimation [9]–[11], [13]. Given an observation variable o which is a vector of RSSI for a set of APs, then location l is calculated as:

$$l = \arg\max_{i} p(l|o) = \frac{p(o|l)p(l)}{p(o)} \tag{1}$$

Where $p(l|o)$ is the posterior distribution of the location l given an observation o. $p(l)$ is the prior probability of being at location l before knowing the value of the observation variable, here $p(l)$ gives a principled way to incorporate background information such as personal user profiles. $p(o)$ is the probability of observation o, but it does not depend on the location variable l, so it is only treated as a normalizing constant. $p(o|l)$ is called the likelihood function, because it gives the probability of the observation o given the assumed source of the observation. Assuming the RSSI from different APs are independent, then:

$$p(o|l) = \prod_{i=1}^{n} p(o_i|l) \tag{2}$$

In Bayesian decision method, the key problems are how to get the prior knowledge $p(l)$ and the likelihood function $p(o|l)$.

III. Rationale of Our AP selection strategy

Suppose there are n APs available in the deployment area, considering the bearing capacity of the client computing devices, we want to select m APs to make the position estimation in the online phase. As we know, there are altogether C_n^m different combinations for m APs, the objective here is to select m APs least sensitive to the environment changes. Here we make use of the residual ranking algorithm, rank the APs according to their residuals, and select the m APs whose residuals are the m smallest, and the AP selection method we proposed only works in the online phase.

Here we want to clarify that, due to the real-time requirement of the position output in the online phase, it is impossible to make our AP selection before each positioning operation. However, we know the signal distribution in the location area takes on a regional property, that also can be interpreted as that the nearby locations share the similar APs property, so in the indoor location tracking, the AP selection method we proposed works only at a fixed time interval when the APs selected at the previous time is outdated, otherwise, we use the APs selected at the previous time to make the estimation. Then how to make the AP selection online?

For n measurements from APi denoted by $m = [m_1, m_2 ....m_n]^T$ and a reference location $\overset{\wedge}{\theta}$, the residual of AP$_i$ can be generalized as $\sum_j (m_j - f_i(\overset{\wedge}{\theta}))^2$ [2], where $f_i(\overset{\wedge}{\theta})$ means the mean measurement of AP$_i$ stores in the radio map at location $\overset{\wedge}{\theta}$. So if $\overset{\wedge}{\theta} \approx \theta$, the residual reflects magnitude of the bias of AP$_i$. by computing the residual of each AP's measurements, and then rank the residuals, thus we can select the m APs whose residuals are the m smallest, then we use the m APs to make the location estimation in the next time interval, and we repeat this process at a fixed time intervals in the entire tracking process.

The residual computing method have a limitation, that is signals at different locations usually have different signal-to-interference ratio (SINR), thus different noise variance σ_i^2 . The residual should be weighted according to σ_i^2. thus the residual of AP$_i$ is given as following:

$$\sum_j \left(\frac{m_j - f_i(\hat{\theta})}{\sigma_i}\right)^2 \tag{3}$$

Here the location $\hat{\theta}$ plays an important role. Ideally, we should use the true location, but it is not achievable. Here we use measurement data from all available APs o determine the location estimation and use it as the approximated location. Then the ResidualRanking algorithm we proposed can be generalized as following:

Algorithm 1: ResidualRanking

Input : o,m
o: the observation vector
m: the numer of APs need to select
Output: M
M: the subset of selected APs

begin
get the reference location estimation using all the available APs
$\hat{\theta} \leftarrow LocationEstimate\ (o)$
compute the residuals of each APs
$S \leftarrow ResidualComputing\ (\hat{\theta}, o)$
rank the APs according to their residual
$Ranking\ (S)$
get the m smallset individuals of S to M
$M \leftarrow SelectAP\,(S, m)$
return M
end

This approach also contains the following two problems to solve:

1) how to choose a value of m APs to make the location estimation?
2) how to choose a value of the time interval T to make AP selection ?

In determining the best value m for the number of APs to select, we need to take into account 2 factors: (1) as m increases, the process of estimating the location becomes more complex and (2) we need a value for such that all locations are covered by at least by those APs most of time. The second factor is important because the number of APs at a given location is varying with time. Typical values for parameter can be found in section 5.

In determining the best value T for the time interval to make the AP selection, two aspects should be considered: (1) as the value of T decrease,more computation will be spent on the AP selection,this will be a burden for the terminal devises, also lose the motivation of AP selection. (2) too large a value of T can damage the performance of the location system, also will lose the robust feature of the AP selection. Typical values for parameter can alsob be found in section 5.

IV. BRR SYSTEM

The BRR positioning systme we proposed is based on Bayesian decision method and the ResidualRanking AP selection method we proposed. The BRR location system works in two phase: offline training phase and online location determination phase, the proposed system architecture is shown in Fig.1.

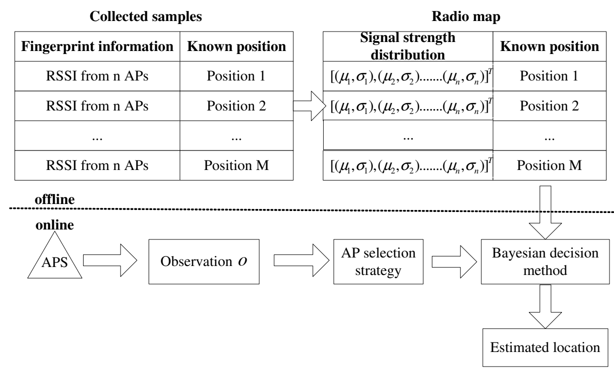

Fig. 1. the architecture of the proposed BRR location system.

A. offline training phase

During the offline phase, the main purpose is to construct a radio map which can represents the signal strength distributions of all the APs available in the deployment area.

Estimating the signal strength distribution

At each location in the set of training locations, we store the distribution of the signal strength of all available APs.

Here we construct the radio map in probabilistic method and make use of Gaussian distribution to represent the signal distribution at a known location [4]. Suppose altogether n APs available in the deployment area are independent, then at location l , the entry is denoted by a pair of a n-dimension vector $[(\mu_1, \sigma_1), (\mu_2, \sigma_2)...(\mu_n, \sigma_n)]^T$ and the known location l . Here (μ_i, σ_i) means the APi 's signal distribution at location l .

Then how to get the Gaussian distribution, here we use the Mean Variance Theory in statistics. At each known location l , we collect k samples, then each sample is a n-dimension vector, and the ith sample is denoted by $[R_{i1}, R_{i2}......R_{in}]^T$ where R_{ij} means the ith sample collect of APj and n is the number of APs. Then (μ_i, σ_i) which denotes the signal distribution of APi is computed as follows:

$$\mu_i = \frac{1}{k}\sum_{j=1}^{k} R_{ji} \tag{4}$$

$$\sigma_i^2 = \frac{1}{k-1}\sum_{j=1}^{k} (R_{ji} - \mu_i)^2 \tag{5}$$

B. Online location determination phase

The general idea of what happened during the determination phase is as follows: suppose the AP selected subset denoted by S, we get samples from all the APs at an unknown location, then if the S must be updated, we make the AP selection, otherwise, we use the APs in S and the radio map constructed in the offline phase to make the location estimation.

As we have made a detailed description of AP selection strategy in section3. Here we only introduce the location estimation method.

Location determination

Here Bayesian decision method is employed to get the location estimation. From section 2.3 we know the main problems of Bayesian decision method are to get the prior knowledge $p(l)$ and the likelihood function $p(o|l)$, here o is the vector of testing samples. Suppose we collect k samples at each test point, then o=$o_1, o_2 ... o_k$, where o_i denotes the ith sample.

We employ kernel method to get the likelihood function $p(o|l)$, A probability mass is assigned to a "kernel" around the mean signal strength vector μ at the location l which is stored in the radio map which we constructed in the offline phase. Thus the resulting density estimate for likelihood function $p(o|l)$ is a mixture of k equally weighted density functions, where k is the number of testing vectors in l:

$$p(o|l) = \frac{1}{k} \sum K(\mu : o_i) \quad (6)$$

Where $K(\mu : o_i)$ denotes the kernel function. Here we use the Gaussian kernel:

$$K_{Gauss}(\mu; o_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(\mu - o_i)^2}{2\sigma^2})$$

Where σ is an adjustable parameter that determines the width of the kernel, here we make use of the σ stores at location l. Then how to get the prior probability $p(l)$, here we make use of the number of samples collected at each location in the training phase to compute this prior probability.

V. Experimental Results

In this section we present the experiments and the results analysis. To demonstrate the performance of the AP selection strategy proposed, we make a comparison of ResidualRanking method and MaxMean method, all systems were implemented in the same environment for fair comparison.

We perform our experiments in a WLAN environment of 30×20 meters which is located on the first floor of the four-story building. To eliminate the effect of randomness of human behavior, in this study, we divide the experimental area into 150 grids, where each grid measures 2×2 meters. A total of 8 APs were detectable throughout the floor, each providing a overall coverage of the environment. the area of the experiment is shown in Fig. 2. In the training phase, we collect samples at each training point in this location area, each sample is thus an 8-demensional signal strength vector and is collected at the center of each grid. Unfortunately, realization of a uniform grid may not always be practical in an indoor environment because of the presence of walls, furniture, and other obstructions, preventing measurements in certain areas. Thus a grid result causes the algorithm to produce a variable resolution in different areas.

Test measurements are collected on different days, time of day, and by different persons than the training set, to capture a variety of environmental conditions. This method can also reflects the mismatch between training and testing conditions in the real-life operation of the system. In this experiment, the test measurements are collected randomly during the routine activities.

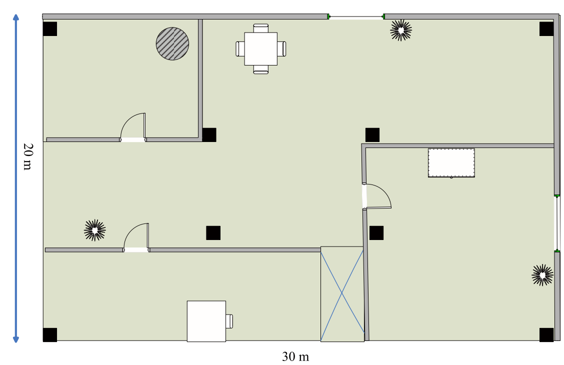

Fig. 2. the area we perform the experiment, the black dots show the location of the APs

To get a more desirable system, we do a lot of experiments on the number of training samples collected at each training point and the number of testing samples collected at each testing point. Here we use all 8 APs to experiment. The results are shown as Fig.3. The results shows that: for the number of training samples collected at each training points, we find that 60 samples can get a more desirable results than the others, too small and too large training samples both can not get desirable results. It is not true that more training samples can get a more accurate result, that is because too large a number of samples makes the radio map over fitting to the training samples, so in the training phase, getting a proper quantity of training samples is very important. Besides, we can find that too few of testing samples collected at each test point also can not get a good result. For example, when we collect 1 sample at each testing point, then the results explicitly lag behind from the other test sets. The results from 2, 3, 4 test sets are similar, but we can see that when the number of the testing samples collected at each testing points is 3 and 60 training samples collected at each training point, the experiment can get the most favorite results. Thus, this configuration is implemented in the following experiments.

To evaluate the ResidualRanking method we proposed, here we make a comparison with the MaxMean method which is works on selecting the strongest APs to make the location estimation. Here we set the time interval T of ResidualRanking method to 2 second. Fig. 4 and Fig.5 report the mean and variance of error of different number of AP selected respec-

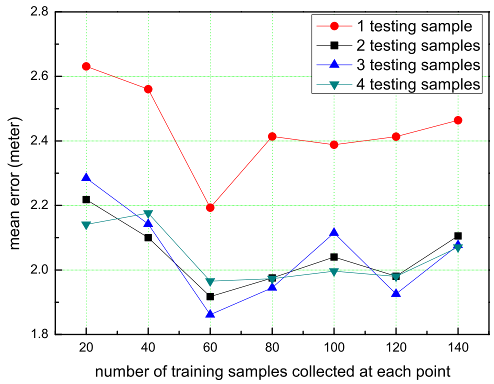

Fig. 3. the effect of the number of training and testing samples collected at each point.

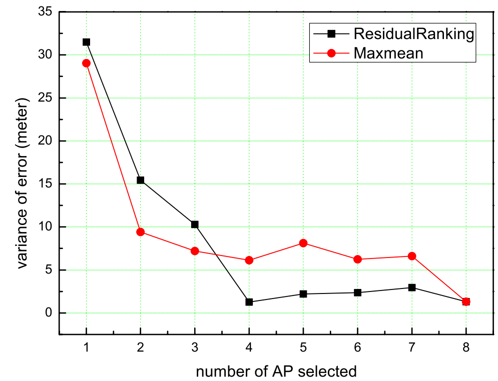

Fig. 5. Variance of error versus number of APs selected

tively. Both the results show that the ResidualRanking method we proposed outperforms the MaxMean method significantly when more than 4 APs are selected. Especially, when 4 APs selected out of 8 APs, the ResidualRanking method we proposed works very well, the mean error is 2.04 and the variance is 1.26, that is very delighted for us to see. The results show that the testing point nearly can be located to the nearest training point. But when less than 4 APs selected, our method works not very well, but this has nearly no negative impact to our method, that is because nearly all the fingerprint method can not get a desirable results when too little APs used to make the location estimation.

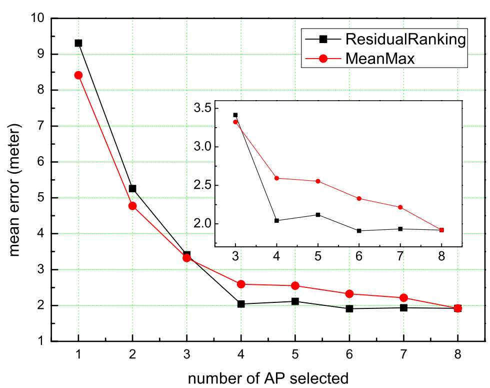

Fig. 4. Mean of error versus number of APs selected

After showing the comparison results, we want to make a deeper study on the ResidualRanking method we proposed, and we also set the time interval T to 2 second. The results in Fig 6 report the mean and variance of error with respect to the number of AP selected in ResidualRanking method we proposed. We can see that the smallest variance of error can be get when 4 APs are selected, and the difference between the mean error is not very distinct when more than 4 APs are selected. In conclusion we get that selecting 4 APs out of 8 APs is the most favorite choice in this experimental area.

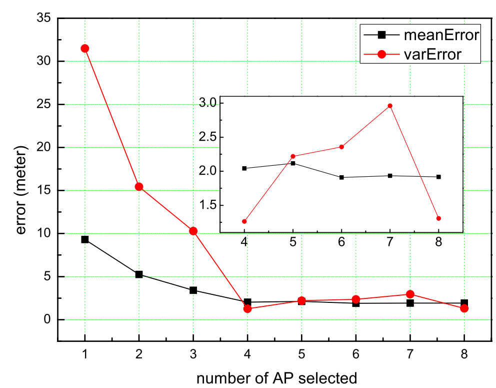

Fig. 6. mean and variance of error versus number of selected APs

Then we want to show the relationship between the time interval T and the performance of BRR system. The time interval T changes from 1 second to 8 second, here set the number of AP selected to 4. The experiment result is shown in Fig.7

From the results shown in Fig.2, we can see that as the T increases, the mean error of the experiments increase flatly, but we can see that when T under 3, the performance of the system changes not very significantly, but when T above 4, the performance of the system declines significantly, so in this experimental environment, $T = 3$ is a proper configuration.

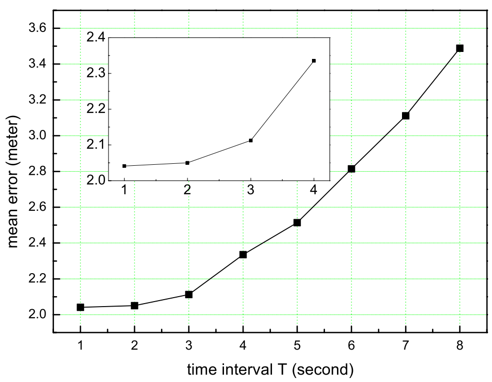

Fig. 7. the relation between the mean error and the time interval T

VI. CONCLUSIONS AND FUTURE WORKS

A. Conclusions

In this paper, we presented the design, implementation, and the evaluation of the ResidualRanking AP selection method we proposed, besides, we implement the BRR location system which is based on Bayesian decision method and the ResidualRanking method we proposed.

The AP selection method we proposed works only in the online phase. It makes use of the residual ranking algorithm to select those APs least sensitive to the environment changes in indoor location tracking, that it has a robust attribute in some extent, also the experiments show that the ResidualRanking method we proposed outperforms the MaxMean method in the experiment.

The BRR location system we proposed is based on Bayesian decision method and the ResidualRanking method we proposed. Besides, we introduce Mean Variance Theory to produce the radio map in the offline phase, that it reduces the online computation burden in some ways, and it accelerates the location determination in the online phase. Finally, the experiments show that the BRR location system works well in the indoor location tracking.

B. Future Works

From this works we know that the environment changes influence the AP selection problem greatly, we believe making AP selection dynamic in the online phase is still deserved to be explored. In future, we will work deeply into this AP selection problem. On the other hand, as a result of the close relationship between AP selection method and location determination method, we will view them as a whole, our ultimate goal is to develop a good performance location system in the indoor location.

REFERENCES

[1] Y. Chen, Q. Yang, J. Yin, and X. Chai. Power-efficient access-point selection for indoor location estimation. *IEEE Transactions on Knowledge and Data Engineering*, pages 877–888, 2006.

[2] L. Cong and W. Zhuang. Nonline-of-sight error mitigation in mobile location. *IEEE Transactions on Wireless Communications*, 4(2):560–573, 2005.

[3] S.H. Fang, T.N. Lin, and P.C. Lin. Location Fingerprinting In A Decorrelated Space. *IEEE Transactions on Knowledge and Data Engineering*, 20(5):685–691, 2008.

[4] K. Kaemarungsi and P. Krishnamurthy. Properties of indoor received signal strength for WLAN location fingerprinting. In *Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004. The First Annual International Conference on*, pages 14–23, 2004.

[5] MB Kjae7rgaard. A Taxonomy for Radio Location Fingerprinting. *Lecture Notes in Computer Science*, 4718:139, 2007.

[6] A. Kushki, K.N. Plataniotis, and A.N. Venetsanopoulos. Kernel-based positioning in wireless local area networks. *IEEE Transactions on Mobile Computing*, pages 689–705, 2007.

[7] M. McGuire, KN Plataniotis, and AN Venetsanopoulos. Data fusion of power and time measurements for mobile terminal location.

[8] J.J. Pan, J.T. Kwok, Q. Yang, and Y. Chen. Accurate and low-cost location estimation using kernels. In *INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE*, volume 19, page 1366. LAWRENCE ERLBAUM ASSOCIATES LTD, 2005.

[9] T. Roos, P. Myllymaki, and H. Tirri. A statistical modeling approach to location estimation. *IEEE Transactions on Mobile Computing*, 1(1):59–69, 2002.

[10] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen. A probabilistic approach to WLAN user location estimation. *International Journal of Wireless Information Networks*, 9(3):155–164, 2002.

[11] A. Smailagic and D. Kogan. Location sensing and privacy in a context-aware computing environment. *IEEE Wireless Communications*, 9(5):10–17, 2002.

[12] J. Yin, Q. Yang, and L.M. Ni. Learning adaptive temporal radio maps for signal-strength-based location estimation. *IEEE Transactions on Mobile Computing*, 7(7):869–883, 2008.

[13] M.A. Youssef, A. Agrawala, and A.U. Shankar. WLAN location determination via clustering and probability distributions.