Location Estimation in Large Indoor Multi-floor Buildings using Hybrid Networks

Kejiong Li, John Bigham, Eliane L. Bodanese and Laurissa Tokarchuk

School of Electric Engineering and Computer Science

Queen Mary, University of London, United Kingdom

{kejiong.li, john.bigham, eliane.bodanese, laurissa}@eecs.qmul.ac.uk

***Abstract*—This paper presents results for an approach for indoor location estimation that integrates received signal strength (RSS) data from both WiFi and GSM networks. Previous work has focused on relatively small indoor environments. In many potential applications, getting approximate location information, such as in which room the mobile user is, is adequate. A hierarchical clustering method is used to partition the RSS space. To choose the best transmitters in a partition, we assess the amount of RSS variance that is attributable to different base stations (BSs) or access points (APs) by transforming the RSS tuples into principal components (PCs). This allows us to retain most of the useful information of detectable transmitters in fewer dimensions. In our experiments, we collected WiFi and cellular RSS on the 2nd and 3rd-floor electronic engineering (EE) building in Queen Mary campus. The experiment results show that the proposed method can provide a good accuracy of room prediction, especially when we integrate WiFi RSS with GSM RSS together to do the positioning.**

I. INTRODUCTION

There are different methods to locate a user's location inside and outside, such as Time of Arrival (ToA), Time Difference of Arrival (TDoA), Angle of Arrival (AoA) and received signal strength (RSS). The ToA, TDoA and AoA methods are seen as uneconomic because they require additional hardware in a transmitter or receiver to locate a user, e.g. precision clocks (ToA and TDoA) and antenna arrays (AoA). Many localization systems utilize the signal strength values received from the base stations (BSs) or relay stations (RSs) or access points (APs) to estimate the location of a mobile user, based on deterministic or probabilistic techniques. RSS has been widely investigated principally in the context of indoor location estimation. This is because the data required to create the RSS database is readily collected from indoors. Though not as accurate as time-based methods, RSS fingerprint-based localization has the potential to overcome the limitations of traditional triangulation approaches, because it performs relatively well for non-line-of-sight circumstances where the alternative of modelling the nonlinear and noisy patterns of realistic radio signals is a challenging task. It requires less battery resources than receiving GPS signals and less run time computational resources than triangulation calculation. Furthermore, RSS-based methods do not require the cooperation of network operators. This paper shows that WiFi RSS and GSM RSS data can be integrated to enhance accuracy.

RSS-based fingerprinting localization typically involves two phases: *training* and *online estimation*. In the training phase, RSS is collected at known locations to form a location fingerprints database (a.k.a. radio map). The generated radio map consists of many location-RSS tuples. Every tuple is the location fingerprint with its corresponding location. In the online phase, new RSS observations measured at unknown positions are compared with all the fingerprints in the radio map to estimate their locations based on the preferred algorithm and distance function. These are outlined in Section II.

Fingerprinting techniques are especially appropriate for the range of frequencies in which GSM and WiFi networks operate. This is because [1] [2] the signal strength at those frequencies presents an important spatial variability. Regarding GSM technology, several research works use this technology for localization, especially in outside environment. For example, our previous work [3] utilized GSM-based fingerprinting for outdoor localization. We have collected RSS fingerprints from the 4-strongest GSM BSs, achieving 50th percentile accuracy of 5.3 meters in a city environment. While inside buildings, [2] has proposed an accurate GSM-based indoor localization system by making use of the wide signal-strength fingerprints (includes the 6 strongest GSM cells and readings of up to 32 additional GSM channels, most of which are strong enough to be detected, but too weak to be used for efficient communication), but with the need of dedicated and complex hardware. Many research works [4] [5] [6] have investigated WiFi RSS fingerprinting in a relatively small size of indoor environment for positioning. [4] represents the first fingerprinting system for indoor localization of portable devices. It localizes a laptop in the hallways of a small office building with accuracies of 2 to 3 meters, using RSS fingerprints from four 802.11 APs. Other work uses augmenting mechanisms to improve the accuracy of this technique, such as RFID [7] and Zigbee [8]. Methods that use auxiliary active RFID tags have been proposed for high indoor accuracy, but this is not ideal for general use in larger areas.

The aim of this paper is to provide a novel hybrid RSS-based localization method to tackle indoor environments. The hybrid localization scheme combines the RSS measurement collected from WiFi networks and from GSM networks and estimates a mobile user location in an indoor multi-floor environment. Unlike other previous research, we are content to locate to a specific room or room segment for a mobile user rather than looking for very high location accuracy indoors. In our work, in order to choose a subset of transmitters, Principal Component Analysis (PCA) [9] technique is used to project the measured RSS into a transformed signal space. The basis in the transformed space can be viewed as the linear combination of each transmitter with different weights (a.k.a principle components (PCs)), which represent the different contributions of each transmitter. A hierarchical partitioning scheme is used to divide the mobile stations (MSs) in the training set into a tree of clusters according to the sequence of the transmitter labels sorted by their RSS values in a descending order. For example, the first level branches correspond to partitions where each particular transmitter is the strongest; all MSs with the same two transmitters in the same order of RSS form a second level branch from the root of the tree. At run time, for a new MS with given RSS tuple, we use the labels of the transmitters that cover this MS sorted by RSS and determine which cluster it belongs to, by finding the longest label match in the tree and then apply the weighted K-nearest neighbour (WKNN) algorithm to predict the room number for this MS in that cluster.

In this paper, the novel features that contribute to the greater accuracy are: (1) the PCA approach can assist to choose the best transmitters, because it can extract the useful information into the relatively lower dimensions by suitable transformation. The reduction of the data dimensions leads to a decrease in the computational complexity and avoids unnecessary calculations; (2) the proposed clustering scheme can give a good partitioning for localization; (3) integrating WiFi RSS with GSM RSS data can enhance positioning accuracy.

The rest of the paper is organized as follows. Section II reviews related work on the traditional transmitters selection approaches and location estimation methods using RSS measurements. Section III illustrates our proposed algorithm for transmitter selection and the room estimation using both GSM RSS and WiFi RSS in indoor multi-floor buildings in detail. Section IV presents the experimental evaluation of our proposed algorithm in a real environment. Section V concludes the paper and discusses directions for future work.

II. Related Work

Various techniques have been proposed to estimate a mobile user's location using RSS. We survey the related work for two different issues: (a) detectable transmitter selection methods and (b) location estimation methods.

A. Previous Work on Transmitters Selection

Choosing a subset of transmitters is an intuitive way to reduce the computational burden and storage requirement on the resource-weak devices. [5] [10] show how their smart AP selection methods can achieve good localization results as compared with using all the available APs in an indoor environment. In [5], the MaxMean approach is proposed to choose the K most important APs, which are defined to be those K APs having the highest average RSS. This mechanism unavoidably throws out the information of detectable but unselected APs, and also requires at least one AP that can communicate with every point in the grid. This makes the approach only suitable for small areas. [10] introduces the InfoGain algorithm for AP selection, which divides the indoor environment into n grid elements. Suppose m is the number of detectable APs. The signal strengths from the APs are collected in every grid G_j. The average value of signal strength in G_j from $AP_i (i = 1, ..., m)$ is defined as the value of the i-th feature of G_j. The main idea of InfoGain is to select the top K APs in terms of the "worth" of each AP feature in every grid element. The worth of each AP_i feature is calculated as the reduction in entropy by including the feature, which is given by $InfoGain(AP_i) = H(G) – H(G|AP_i)$. Here $H(G) = -\sum_{j=1}^{n} Pr(G_j) \log Pr(G_j)$ the entropy of the grid when AP_i's RSS value is not known, $Pr(G_j)$ is the prior probability of grid G_j and is treated as uniformly distributed, i.e. a user can be equally likely in any grid [10]. $H(G|AP_i) = \sum_v \sum_{j=1}^{n} Pr(G_j, AP_i = v) \log Pr(G_j|AP_i = v)$ computes the conditional entropy given AP_i's value. v is one possible value of signal strength from AP_i. The summation is taken over all possible values of AP_i. So they only need to focus on the value of $H(G|AP_i)$ which is the conditional entropy of grid G given AP_i's RSS value. Although positive results have been demonstrated for a small indoor environment, for a larger area such as an outdoor environment, it is difficult to determine the appropriate number of grid elements for the target area and the size of each grid element.

B. Previous Work on Location Estimation Approaches

Fingerprinting can be categorized into deterministic and probabilistic approaches. A simplest deterministic approach estimates the location of an observed RSS tuple as the average of the locations of the K-nearest neighbour (KNN) in RSS space which is measured as the Euclidean distance from this observed RSS to the training RSS tuples [4]. [7] and [11] improve the accuracy by using a weighted average of the coordinates of the K nearest training samples. The weight values are taken as the inverse of the Euclidian distance between the observed RSS measurement and its K nearest training samples. This method is referred to as weighted K-nearest neighbour (WKNN). The experimental results in [11] indicate that the KNN and the WKNN can provide a higher accuracy than the single nearest neighbour (NN) method, particularly when K = 3 and K = 4. However, when a high density radio map is available, i.e. there is a lot of training data, the simple NN method can perform as well as other more complicated methods [12]. Probabilistic techniques [5] [6] [13] [14] are used for modelling errors in the location estimation of RSS measurements in wireless networks. These methods use the training RSS tuples to construct conditional probability density functions of the fingerprint tuples given their locations, and utilize Bayes' theorem to compute the posterior probabilities of possible locations given a new RSS tuple. However, because of the complexity, the authors in [6] assume the elements of the fingerprint tuple are statistically independent from each other, which does not hold in a real environment, i.e. our data sets contain correlation coefficients greater than 0.9.

III. Proposed Location Estimation Scheme

We will mainly focus on two issues: (1) how to select the most representative detectable transmitters; (2) how to create clusters and estimate the position within clustering.

A. The Most Representative Detectable Transmitters Selection

In a real environment, a mobile user can receive signals from many detectable transmitters surrounding the area of the interest. For example, for our test-bed (the 2nd and 3rd-floor Electronic Engineering (EE) building in Queen Mary campus) 101 APs and 20 BSs for a particular operator can be detectable. The PCA scheme is used to choose the optimal subset of APs and BSs separately. Here we take the process of AP selection as an example.

Assume that in the training stage, a set of n MSs: the RSS measurements from all N neighbouring APs and its corresponding room number are collected. If one MS does not receive measurable signal strength from one typical AP, we set a default value -120 dBm, as it is the minimum signal strength. Let R denote the RSS measurements received by all the training data from APs in WiFi networks in the target environment, which is given by

$$R = \begin{pmatrix} \vec{r}_1 \\ \vdots \\ \vec{r}_i \\ \vdots \\ \vec{r}_n \end{pmatrix} = \begin{pmatrix} r_{11} & \cdots & r_{1,j} & \cdots & r_{1,N} \\ \vdots & \ddots & & & \vdots \\ r_{i,1} & & r_{i,j} & & r_{i,N} \\ \vdots & & & \ddots & \vdots \\ r_{n,1} & \cdots & r_{n,j} & \cdots & r_{n,N} \end{pmatrix}_{n\times N} = \begin{pmatrix} \vec{t}_1 \\ \vdots \\ \vec{t}_j \\ \vdots \\ \vec{t}_N \end{pmatrix}^T \quad (1)$$

Here $\vec{r}_i$ is a N-dimension row vector of RSS received by MS i from N APs, while $\vec{t}_j$ is the n-dimension column vector of RSS received by all the n MSs from AP j, i.e. a different viewpoint on the same data. $L = (l_1, ..., l_i, ..., l_n)$ consists of the room numbers, l_i is the room number where MS i is. The detectable AP selection process based on PCA can be divided into following steps:

Step 1: Calculate the matrix $\bar{R}$, each of its row vector is the mean value $(\bar{t}_1, ..., \bar{t}_j, .., \bar{t}_N)$ of the training RSS data in R from each AP, and then the $N \times N$ covariance matrix S of the training RSS data can be obtained. This describes the mutual dependence of the signal strengths received by any two MSs from different APs and can be expressed as $S = \frac{1}{n-1}(R-\bar{R})(R-\bar{R})^T$. In this case, $S = [s_{j,p}]$, where $s_{j,p} = \frac{1}{n-1}\sum_{i=1}^{n}(r_{i,j}-\bar{t}_j)(r_{i,p}-\bar{t}_p)^T, 1 \le j,p \le N$.

Step 2: Calculate the eigenvalues and eigenvectors of S. The eigenvalues contains the variances for the principal components (PCs) and the eigenvectors contains the linear coefficients for the principal components. Assume the eigenvalue $\{\lambda_1, ..., \lambda_N\}$ is in descending order, and $\vec{e}_i$ represents the normalized eigenvector associated with λ_i. Thus, the principal component coefficients can be defined as $A = [\vec{e}_1, ..., \vec{e}_N]$. So far, we have accomplished the principal components analysis itself.

Step 3: Hence, to put the PCA to use, we need to know what proportion each principal component represents of total variance, which can be expressed as

$$\omega_i = \frac{\lambda_i}{\sum_{i=1}^{N}\lambda_i} \quad (2)$$

According to (2), we can remove the PCs that contribute little to the variance, and project the entire dataset to a lower dimensional space, but retain most of the information. Fig. 1 gives an example of how to choose the optimal PCs on the two floors EE building in Queen Mary campus in WiFi networks. Since in our measurements, there are 101 APs detected on the two floors and 20 APs are stable. This figure shows how much variance in the dataset is explained by which PC (by the bars shown in Fig. 2) and how much variance is explained by the successive PCs (as seen by the blue line shown in Fig. 2). We can observe that the 16 first PCs can capture most of the variability in the signal strengths from the stable 20 APs. Since the principal components are orthogonal, the amount of total variance expressed by the first 12 PC can give us over 95% of the information about RSSs from the 20 Aps. In this way, the reduced dimensions can be denoted as Q, which is 12 here. The optimized principal component is $A_{opt} = [\vec{e}_1, ..., \vec{e}_Q]$.

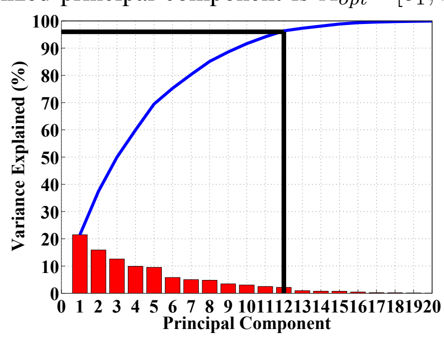

Fig. 1. The cumulative variance accounted for by successive PCs in WiFi networks on the 2nd and 3rd-floor EE building in Queen Mary campus

Step 4: Calculate component loadings of each AP to the Q largest PCs.

$$u_{ij} = \sqrt{\lambda_i}e_{ij}(i \in [i,Q], j \in [i,N]) \quad (3)$$

Here u_{ij} is the component loading of the j-th AP on the i-th PC, and e_{ij} is the j-th element of $\vec{e}_i$. Within each PC, one AP with the largest absolute value of component loading is chosen.

B. Training Stage

Once the optimal APs and BSs are determined, we want to build a radio map during the training stage. Let q be the optimal number of transmitters, which contains q_1 WiFi APs and q_2 GSM BSs. Let $Z = \{z_1, ..., z_i, ...z_n\}$ be the set of the n training MSs, where $z_i = \left[\vec{r}_i^{\,wifi}, \vec{r}_i^{\,gsm}, l_i\right]$, $\vec{r}_i^{\,wifi}$ and $\vec{r}_i^{\,gsm}$ are the WiFi RSS tuples and the GSM RSS tuples for the i-th training MS respectively and l_i denotes its corresponding room number.

The training stage analyses the RSS of WiFi and GSM separately with the same procedures, which are illustrated in Algorithm 1 below. For WiFi measurements of every training data z_i (step 1), we pick out the strongest w RSS measurements in descending order, which can be expressed as

$$\left\{ r^{wifi}_{i,i_1^W} \geq r^{wifi}_{i,i_2^W} \geq ... \geq r^{wifi}_{i,i_w^W} \mid i_1^W, i_2^W, ..., i_w^W \in [1, q_1] \right\} \quad (4)$$

Here $i_1^W, i_2^W, ..., i_w^W$ are the ID series of the chosen w WiFi transmitters respectively (step 2). Similarly, we can get the corresponding IDs of BSs as $i_1^G, i_2^G, ..., i_g^G$ by choosing the strongest g RSS of GSM transmitters in descending order (step 3). Then these IDs are concatenated into a new series: $[i_1^W, i_2^W, ..., i_w^W, i_1^G, i_2^G, ..., i_g^G]$ (step 4), which also can be regarded as the ranking pattern $\mathcal{P}_i^D$ of z_i of dimension $D(D = w + g)$. By repeating choosing sequential number of continuous IDs in $\mathcal{P}_i^D$ from the starting ID (step 5), we can get different ranking patterns $\mathcal{P}_i^d(d \in [i, D])$ for z_i in every dimension (step 6). Each ranking pattern $\mathcal{P}_i^d$ corresponds to a specific cluster C_i^d. Afterwards, the training data with the same ranking pattern are assigned into the same cluster (step 7). Repeat the steps and finally we can obtain ranking patterns of every training data in every dimension.

Algorithm 1

Requried:

$Z = \{z_1, ..., z_i, ..., z_n\}$: training data set

w, g: the number of transmitters to choose from optimal APs and BSs respectively (the matching length)

Steps:

1: **for** $i = 1$ **to** n **do**
2: Sort the strongest w WiFi RSS in descending order and get corresponding ID series of APs : $[i_1^W, i_2^W, ..., i_w^W]$
3: Sort the strongest g GSM RSS in descending order and get corresponding ID series of BSs: $[i_1^G, i_2^G, ..., i_g^G]$
4: Combine the two ID series together to get the ranking pattern in dimension $(D = w + g)$.
$$\mathcal{P}_i^D = [i_1^W, i_2^W, ..., i_w^W, i_1^G, i_2^G, ..., i_g^G]$$
5: **for** dimension $d = 1$ **to** D **do**
6: The ranking pattern in dimension d is:
$$\mathcal{P}_i^d = \mathcal{P}_i^D(1 : d)$$
7: $\forall j < i$, **if** $\mathcal{P}_i^d = \mathcal{P}_j^d$, **then** $C_i^d \equiv C_j^d$
8: **end for**
9: **end for**

C. Online Localization Stage

Given a new MS m with observed RSS measurement $\vec{r}_m$ from q transmitters including q_1 APs and q_2 BSs, we want to estimate which room this MS belong to, which is illustrated in Algorithm 2 below. First we sort the RSS of WiFi and GSM in descending order separately, and obtain an ID sequence of the length of w and g (step 1 and 2), both of which are concatenated into a new one that can be defined as the ranking pattern $\mathcal{P}_m^D$ of MS m of dimension D (step 3). Then the estimation process of room ID is carried on recursively within the d dimensions (step 4). MS m's ranking pattern $\mathcal{P}_m^d$ is obtained by choosing a sequence from the beginning element in $\mathcal{P}_m^D$ with the length of d (step 5). MS m is regarded as belonging to the cluster which has the same ranking pattern with m in dimension d (step 6), and the room ID of MS m can be estimated by applying WKNN method using the room IDs of training data in the same cluster (step 7). If there is no cluster that satisfies this condition, we continue to search in a lower dimension (step 4).

Algorithm 2

Requried:

$Z = \{z_1, ..., z_i, ..., z_n\}$: training data set

w, g: the number of transmitters to choose from optimal APs and BSs respectively (the matching length)

$\mathcal{P}_i^d$: all the ranking patterns of training data in every dimension. $(i \in [1, n], d \in [1, D])$

$\vec{r}_m$: the MS m's RSS measurements from WiFi and GSM networks.

Steps:

1: Sort the strongest w WiFi RSS of m in descending order and get corresponding ID series of APs: $[m_1^W, m_2^W, ..., m_w^W]$
2: Sort the strongest g GSM RSS of m in descending order and get corresponding ID series of BSs: $[m_1^G, m_2^G, ..., m_g^G]$
3: Concatenate the ID series obtained in step 1 and 2 as the ranking pattern of m in dimension $(D = w + g)$.
$$\mathcal{P}_m^D = [m_1^W, m_2^W, ..., m_w^W, m_1^G, m_2^G, ..., m_g^G]$$
4: **for** dimension $d = D$ **to** 1 **do**
5: $\mathcal{P}_m^d = \mathcal{P}_m^D(1 : d)$
6: **if** $\exists i$ **satisfies** $\mathcal{P}_m^d \equiv \mathcal{P}_i^d$ **then**
7: $m \in C_i^d$
8: Apply WKNN to estimate the room ID of m using the RSS and Room IDs of training data Z in C_i^d.
9: **Break**
10: **end if**
11: **end for**

IV. Experimental Evaluation

The proposed algorithm is evaluated on a real indoor environment. The measurements are collected on the 2nd and 3rd-floor of EE building in the Queen Mary campus, as shown in Fig. 2. We collect both GSM RSS data and WiFi RSS data in each room in this area by a mobile app on an Android smartphone and their corresponding location information are labelled with the room number. The downloadable data can be found at [15].

The performance of the proposed localization method is compared with the KNN method in [4] and the KDE method in [6], which assumes RSSs are independent statistically. For the KDE method, we build the RSS probability density for every room. Three forms of RSS are used, viz. GSM RSS, WiFi RSS and both of them (a.k.a hybrid RSS). In the target indoor environment, we collected 500 samples of GSM and WiFi signal strengths in 15 different rooms on the two floors. In the experiments, the data are randomly divided into two sets. The first set is treated as training data and is a collection of 400 samples. Their corresponding room location information is known. The other set contains the other 100 samples and is used for room estimation test using only the RSS values,

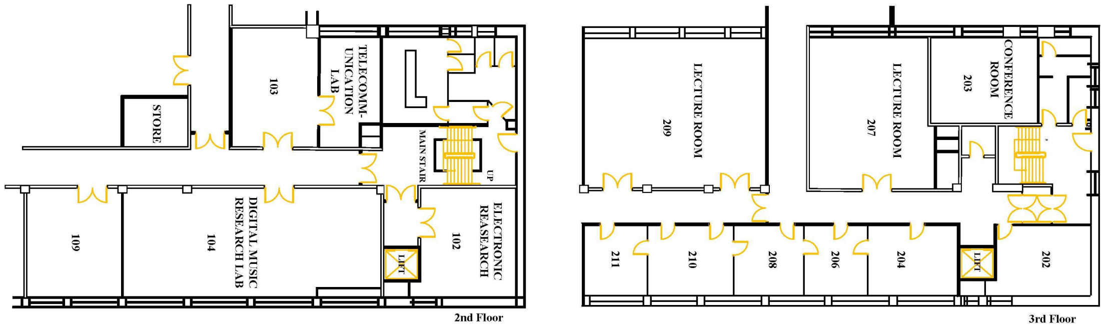

Fig. 2. The layout of the experimental test-bed

with their correct room number subsequently only used for validation.

A. The Effects of Transmitters Selection Methods

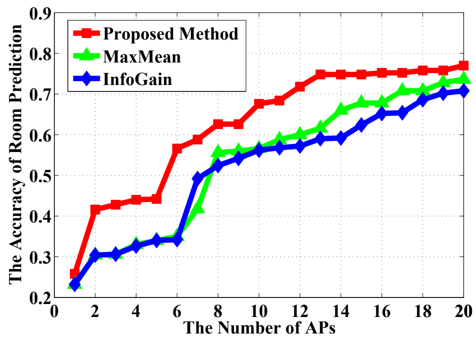

Fig. 3. The average accuracy of room prediction versus the number of APs

To better illustrate the effects of transmitters selection methods, we take the WiFi networks for example. We compare our approach to select the best number of APs from the 20 stable APs with the MaxMean [5] and InfoGain [10] approaches. For the InfoGain approach we take every room as a grid element. The performance is evaluated in terms of the average accuracy of room estimation, which is defined as the cumulative percentage of estimations within specified errors. Fig. 3 shows the accuracy comparison between MaxMean, InfoGain and our proposed transmitter selection method. It can be clearly seen that our approach significantly outperforms the traditional methods under the same numbers of the APs. For example, when using 12 APs, our approach reports 71.8% accuracy of room estimation while those of MaxMean and InfoGain are 60% and 57.2% respectively. Likewise, the proposed transmitter selection approach performs better than the other two methods for the GSM networks. In this comparison, we choose 12 APs and 4 BSs as the best subset respectively after using PCA.

B. The Effect of the Matching Length in Clustering

The proposed method needs to create different clusters according to ranking patterns with the different number of dimensions during the training stage. To balance the trade-off between computational complexity and estimation accuracy, we need to find the best maximum matching length to create clusters. Fig. 4 reports that the room estimation accuracy versus the highest maximum matching length used in clustering in WiFi networks. This corresponds to the depth of the tree constructed during training phase. Seen from Fig. 4, we can

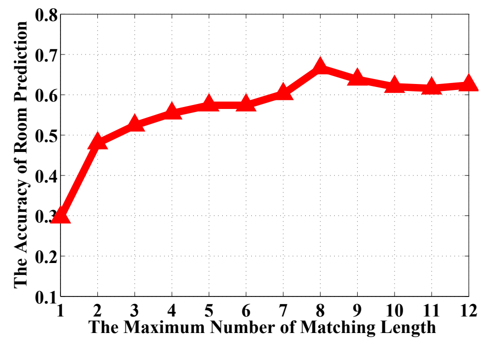

Fig. 4. The accuracy of room prediction versus the maximum number of matching length in clustering in WiFi networks

see the estimation accuracy increases as the matching length increases from 1 to 8. However, the predictive accuracy does not monotonically improve along with the increasing matching length. When the maximum allowed matching length is set as 8, inclusion of additional RSS leads to worse rather than better performance. Similarly, for GSM, the best maximum number length allowed is set as 4.

C. Positioning Performance

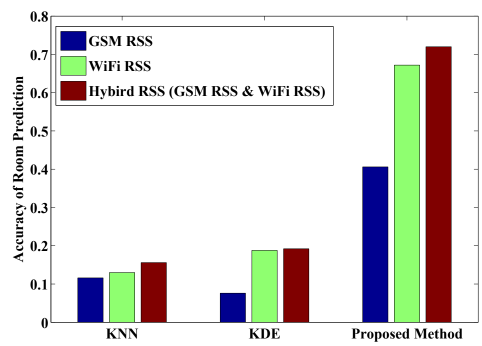

Fig. 5. The room accuracy results for different alrgothims in three forms of RSS

Fig. 5 compares the estimation accuracy of the three different algorithms by using GSM RSS, WiFi RSS and both of them. We perform 10 trials for every algorithm and plot the mean value. In each trial, we used the same number of training data and test data. From Fig. 5, we can see that the proposed localization method significantly outperforms the two traditional methods, especially when hybrid RSS data (WiFi RSS and GSM RSS) is used. For instance, when integrating WiFi RSS with GSM RSS, our proposed method can achieve 72% accuracy of correct room prediction, whereas the

KNN and KDE methods report 15.6% and 19.2% respectively. Furthermore, we observe that all the three methods based on hybrid RSS data perform better than those based on only GSM signal strength or WiFi signal strength at a certain extent. In addition, it clearly shows that using WiFi data can achieve better accuracy than using GSM data. This is because the variation of GSM signal strengths in different rooms is smaller than that of WiFi signal strengths and there are more APs available.

D. Comparing cluster-based PCA and Global PCA

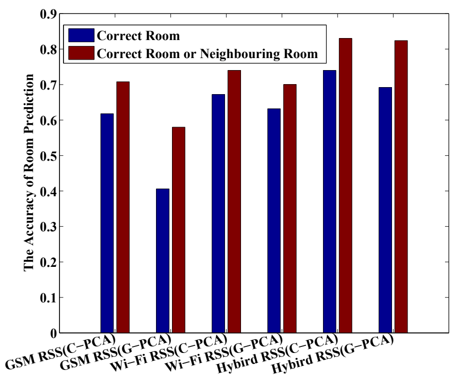

Fig. 6. The room accuracy results comparisons between cluster-based PCA and Global PCA methods in three forms of RSS

As we mentioned before, the target of our research is to perform location estimation in a large indoor areas. Here it might not be a reasonable way to use global PCA to choose a best subset of transmitters relevant to all possible locations. We cannot necessarily neglect any one of detectable transmitters because each of them takes the important responsibility in the region where it covered. So we compared using global PCA method (a.k.a G-PCA) that we have described in Section III, and another approach where PCs are selected within each cluster (a.k.a C-PCA) from the full set of transmitters. In each cluster we use the best transmitters, i.e. those that account for most of the variability in the data. Both methods are tested by using different subsets of the RSS, as shown in Fig. 6. This figure not only shows the correct room prediction accuracy, but also illustrates the accuracy of obtaining either the correct room or a neighbouring room. A marked improvement in accuracy is found using C-PCA, especially for the GSM RSS. The reason is that the chosen PCs can be quite different in each cluster after a suitable transformation, and C-PCA does not require the selection of a single relevant subset, which G-PCA does. Therefore C-PCA is more scalable. This would be more apparent in a larger area.

V. Conclusion

In this paper, we have proposed a novel hybrid RSS-based room estimation approach in multi-story indoor environment. The use of PCA method to investigate the different contributions of the different transmitters to choose the subset of transmitters can reduce the computational burden and storage, as well as improve the estimation accuracy. To evaluate the performance of our proposed cluster-based deterministic algorithm, we collected real RSS from both WiFi and GSM networks on two floors EE building in Queen Mary campus. The results indicate using the hybrid RSS can improve the estimation accuracy in multi-story building compared with the traditional algorithms. The cluster based approach allows for extensibility to larger areas as a predefined subset of relevant transmitter is not selected for the whole area, rather it is specific to each cluster.

In future work, we will incorporate the mobile users' movement trajectories to further improve the accuracy of location estimation. We are currently validating mechanisms for integrating WiFi RSS with GSM RSS and have obtained significant accuracy improvements in a large scale multi-storey indoor environment, e.g. Stratford Westfield shopping mall in London. The corrections described in this paper can help enhance the accuracy. Moreover, we also have been developing an approach for adjustment according to temperature, humidity and user density.

References

[1] V. Otsason, A. Varshavsky, A. LaMarca and E. de Lara, "Accurate GSM Indoor Localization," in *Proc. of the 7th Int. Conf. on Ubiquitous Computing*, Tokyo, Japan, Sept. 2005, pp. 141-158.

[2] E. Martin, O. Vinyals, G. Friedland and R. Bajcsy, "Precise Indoor Localization Using Smart Phones," in *Proc. of the 18th International Conf. on Multimedia*, Firenze, Italy, Oct. 2010, pp. 787-790.

[3] K. Li, P. Jiang, E.L. Bodanese and J. Bigham, "Outdoor Location Estimation Using Received Signal Strength Feedback," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 978-981, July 2012.

[4] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system", in *Proc. of IEEE 19th Annu. Joint Conf. of the IEEE Comput. and Commun. Soc.*, Tel Aviv, Israel, Mar. 2000, vol. 2, pp. 775-784

[5] M. Youssef, A. Agrawala and A. U. Shankar, "WLAN Location Determination via Clustering and Probability Distributions", in *Proc. of the 1st IEEE Int. Conf. on Pervasive Computing and Commun.*, Dallas-Fort Worth, Texas, USA, 2003, pp. 143-150.

[6] T. Roos, P. Myllymki, H. Tirri and P. Misikangas, "A Probabilistic Approach to WLAN User Location Estimation", *Int. J. of Wireless Inform. Networks*, vol. 9, no. 3, pp. 155-164, July 2002.

[7] L. M. Ni, Y. Liu, Y. C. Lau and A. P. Patil, "LANDMARC: Indoor Location Sensing Using Active RFID", *Wireless Networks*, vol. 10, pp. 701-710, 2004.

[8] A.S.-I. Noh, W.J. Lee, J.Y. Ye, "Comparison of the Mechanisms of the Zigbee's Indoor Localization Algorithm Software Engineering," in *Proc. of 9th Int. Conf. on Software Eng., Artificial Intell., Networking, and Parallel/Distributed Computing*, Thailand, Aug. 2008, pp. 13-18.

[9] I.T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer-Verlag, 2002, pp. 1-9.

[10] Y. Chen, J. Yin, X. Chai and Q. Yang, "Power-Efficient Access-Point Selection for Indoor Location Estimation," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 7, pp. 877-888, July 2006.

[11] B. Li, J. Salter, A. G. Dempster and C. Rizos, "Indoor Positioning Techniques Based on Wireless LAN", School of Surveying and Spatial Information Systems, UNSW, Sydney, Australia, 2006.

[12] V. Honkavirta, T. Perala, S. Ali-Loytty and R. Piche, "A Comparative Survey of WLAN Location Fingerprinting Methods", in *Proc. of the 6th Workshop on Positioning, Navigation and Commun.*, Hannover, Germany, Mar. 2009, pp. 243-251.

[13] A. Haeberlen and A. Rudys, "Practical robust localization over large-scale 802.11 wireless networks", in *Proc. of the 10th ACM Int. Conf. on Mobile Computing and Networking*, 2004, pp. 70-84.

[14] A. Kushki, K. N. Plataniotis and A. N. Venetsanopoulos, "Kernel-based Positioning in Wireless Local Area Networks", *IEEE Trans. on Mobile Computing*, vol. 6, no. 6, pp. 689-705, June 2007.

[15] K. Li and P. Jiang, *Open Google Project*, [Online]. Available: http://code.google.com/p/location-estimation-trials/.