Autonomous Construction of a WiFi Access Point Map Using Multidimensional Scaling

Jahyoung Koo and Hojung Cha

Department of Computer Science, Yonsei University,
134 Shinchon-Dong Sudaemoon-Ku, Seoul, Korea
koojh@cs.yonsei.ac.kr, hjcha@cs.yonsei.ac.kr

Abstract. To construct a WiFi positioning system, dedicated individuals usually gather radio scans with ground truth data. This laborious operation limits the widespread use of WiFi-based locating system. Off-the-shelf smartphones have the capability to scan radio signals from WiFi Access Points (APs). In this paper we propose a scheme to construct a map of WiFi AP positions autonomously without ground truth information. From radio scans, we extract dissimilarities between pairs of WiFi APs, then analyze the dissimilarities to produce a geometric configuration of WiFi APs based on a multidimensional scaling technique. To validate our scheme, we conducted experiments on five floors of an office building that has an area of 50 m by 35 m in each floor. WiFi APs were located within a 10m error range, and floors of APs are recognized without error.

Keywords: WiFi Access Point Map, Positioning, Autonomous and Unsupervised Learning, Multidimensional Scaling

1 Introduction

Since the early 2000s, research on WiFi positioning systems [1-9] has been actively carried out, and databases for positioning systems have been constructed in various places. The databases usually consist of RF fingerprints or positions of WiFi access points (APs). A WiFi-enabled device estimates its position by matching an observed radio scan against the RF fingerprints database [1] or by estimating distances from WiFi APs within its radio range [8, 9]. Most current WiFi positioning systems build their database by inputs from dedicated operators who gather received signal strengths (RSSs) transmitted by WiFi APs, along with ground truth information where RSSs were scanned, while driving outdoors or walking indoors. It is, however, not possible for a limited number of dedicated operators to cover all locations; some places are often not open to the public. Furthermore, generating ground truth information is cumbersome and incurs a high cost.

Recently, several studies have attempted to reduce the effort needed for the database construction. Additional devices are, for instance, used to generate ground truth information [10]. Inertial sensors, such as accelerometers and gyroscopes, make it possible to estimate trajectories of a device. The performance is, however, highly dependent on the precision and mounted positions of the sensors. Hence, it is difficult to produce meaningful data with a conventional device, such as a smartphone. In other studies, stochastic models were used to minimize the required amount of ground truth data [11, 12]. Generally, the initial database is constructed with a small amount of data with ground truth, then the database is expanded with more data without ground truth. The methods are devised for a particular database builder. Despite these efforts, constructing WiFi positioning systems especially in indoors that are usable by a wide variety of people in everyday lives has not yet been achieved.

Our goal is to achieve the ubiquity of WiFi positioning systems for all people at all times. To do so, a database should be built for all places that people can visit. We believe that the conventional method of data collection has limitations for reaching this goal. Hence, a scheme for ordinary people to join in building the database should be developed. Off-the-shelf smartphones have WiFi capability, and the number of these devices is dramatically increasing. Hopefully, WiFi APs can be scanned every one meter during people's daily lives. One critical problem is that most radio scans do not include information about the location of the scan.

In this paper, we propose a scheme to use radio scans gathered by smartphones to identify positions of WiFi APs. Ordinary smartphones users are not expected to move in a predefined route. Hence, ground truth information of radio scans is usually unknown. Each radio scan only includes identifiers of WiFi APs and their received signal strengths. In a scan, fortunately, radio signals from multiple APs are received. Our basic idea is to extract dissimilarities between pairs of WiFi APs from radio scans; then by analyzing the dissimilarities, produce the geometric configuration of WiFi APs. Here, we assume that we can acquire enough radio scans and that some of them are gathered in places beneficial for our scheme, for instance, a place near an AP.

To estimate the geometric configuration of WiFi APs, we adopt multidimensional scaling (MDS) techniques [13, 14, 15]. MDS has been successfully applied in the social sciences to find a spatial relationship from high dimensional data, which contains dissimilarity information between objects. In wireless sensor networks, MDS was used to estimate locations of sensor nodes through measuring all pairs of distance between nodes [16, 17, 18]. In case of WiFi AP, adopting MDS is not straightforward. Since WiFi APs cannot measure distances from other APs without changing their software, we take an approach to estimate dissimilarities between APs from radio scans gathered by smartphones. Dissimilarity does not mean exact distance. It is a measure of how close it is. Hence, a result of multidimensional scaling is relative positions of WiFi APs. Given at least three positions of APs or radio scans, relative positions can be transformed to absolute positions. In the research, we focus on accurately estimating relative positions of APs corresponding to the real geographic configuration. In addition, we expand our scheme to a multifloor environment by using three-dimensional MDS.

Knowing the positions of WiFi APs is useful in many aspects because of the following reasons: First, positions of APs are directly used to locate WiFi devices. Several systems provide positioning service based on positions of WiFi APs [3, 8]. Second, positions of APs can be used as supplementary data for constructing WiFi positioning systems [11, 19]. Third, the geometric configuration of WiFi APs provides the knowledge about the nature of networks, e.g., density, connectivity, interference properties, and models for simulations [20].

The rest of this paper is organized as follows: Section 2 presents related works. In Section 3, we propose a WiFi AP positioning algorithm using MDS. The experimental results are provided in Section 4. In Section 5, we provide information on future challenges. We conclude the paper in Section 6.

2 Related Work

According to the type of database, WiFi positioning algorithms can be categorized into two types: RF fingerprint-based and AP position-based. Some algorithms use both schemes.

In an RF fingerprint-based algorithm, a space is divided into grids. Dedicated operators gather radio scans at every grid. The gathered data from the entire grid construct an RF signal map in the space. When a mobile device scans radio signals at a position, the radio scan is searched in the signal map. The position of the best matching scan in the signal map is determined as the device's position. RADAR [1] is the representative research work on this method. Horus [2] is based on this mechanism, but the scheme uses a stochastic model to improve the performance. The commercial system Ekahau [7] also uses a similar algorithm. To achieve a high accuracy in RF fingerprint-based algorithm, we reduce the size of the grid and increase the number of scans in the grid. RSS fingerprint-based algorithm is known to be superior to AP position-based algorithm. The database construction process is, however, too laborious.

In an AP position-based algorithm, positions of APs are given a priori by a network operator; otherwise they need to be discovered through the process of war-driving. In this approach, there are several possible methods to estimate the position of a mobile device. The simplest method is to use proximity, where a mobile device takes its position from the nearest AP. The method using a Centroid or weighted centroid is also frequently adopted [3, 6]. Here, the position of a mobile device is estimated as the center of positions of APs visible to the device. If a radio propagation model is known, distances from positions of APs are calculated based on RSSs. Then, the position is estimated by multilateration [9]. Place Lab [3] and Skyhook Wireless [8] work mainly based on positions of WiFi APs. Performance of this kind is known to be relatively lower than RF fingerprint-based algorithm. Several researchers, however, tried to improve the performance of AP position-based algorithm close to the performance of RF fingerprint-based algorithm [11, 19]. Even though the performance of AP position-based algorithm is relatively low, this approach has several advantages. First, it only needs a small database, for instance, identifiers of APs and their positions. Second, system building efforts are relatively light. It does not need to survey every grid in a site; hence this algorithm is easily applied on a large scale. Third, it is more reliable in case ground truth data is inaccurate and networks frequently changes, such as addition of new APs [4, 6]. Therefore, to achieve the ubiquity of WiFi positioning system in anonymous network environments, these advantages need to be taken into consideration.

Several researchers have tried to reduce the labor in constructing a WiFi positioning system. Reducing the required number of ground truths is one approach.

Wang et al. [12] devised an algorithm for a multiple floor environments. In that research, WiFi radio scans were collected with ground truth data on a floor, while on other floors, radio scans were gathered without ground truth data. They say that floor plans in a building are usually identical, and although the signals can be quite different on different floors, some correspondences exist. Hence, by aligning the position-known radio scans with position-unknown ones, signal maps of adjacent floors are constructed. However, only a few adjacent floors can be successfully processed.

In [19], AP positions were known a priori. An RF signal map was generated from scanned data by mobile nodes, which know the location of scanned radio signals. Madigan et al. [11] took a similar approach, but they did not assume that the scanned positions of mobile node were known. WiFi-SLAM [21] is an approach to build indoor radio maps with WiFi radio vectors. The scheme estimates the movement of a device and simulates the latent-space locations of unlabeled signal strength data using Gaussian process latent variable modeling. The scheme can use radio scans only gathered by one person moving at a constant speed, and its computation requirement is too expensive to be executed in a smartphone.

In contrast, Woodman et al. [10] tried to generate ground truths by using inertial sensors. They accurately estimated the moving trajectory of a person. The sensors used in the experiment are, however, too expensive and are mounted on foot, which is the best position to estimate pose and direction. Hence, this method is not practical for the average individual. Inertial sensors built in current smartphones are not accurate enough to estimate a moving trajectory. Frequent changes of positions of smart phone, for instance at hand or in a pocket, make it worse.

EZ system [22] has a similar approach to ours. The system tries to eliminate system training efforts. EZ uses radio scans gathered in unknown locations and three scans in known locations, which are essential in the algorithm. The scheme is based on an RF propagation model. From radio scans, it estimates parameters of the RF propagation model, locations of APs, and locations of radio scans. The results show that the performance is comparable to previous indoor location systems. EZ, however, requires extensive calculation, which takes a few minutes to several hours in a high-end computer, and large storage for the radio scans. In a large site, the scheme actually needs more than three scans gathered in known locations.

There have been several studies using MDS in the positioning research field. MDS-MAP [16] is the representative research work based on MDS. It estimates positions of sensor nodes by using connectivity information and finding the shortest paths between all pairs of sensor nodes. Unlike WiFi APs, sensor nodes usually have the capability to receive a radio signal or ultrasound from neighboring sensor nodes. Hence, distances are easily measured. When the average connectivity degree is larger than 9, the simulation result shows that the localization error is below the radio range of a sensor node. Xian Ji et al. [18] also applied MDS to a sensor positioning field. MDS technique was used in a distributed manner to estimate a local map for each group of neighbor sensors; these maps are then aligned together based on the alignment method. The approach is claimed to accurately estimate a sensor's position in a network with an isotropic topology.

3 WiFi AP Positioning Using MDS

Through the analysis of dissimilarities between pairs of objects, MDS makes it possible to discover a geometric configuration of the objects in a low-dimensional space, usually two- or three-dimensional, where the configuration matches the original dissimilarities between the pairs of objects [13, 14, 15]. Dissimilarity between two objects reflects how far apart an object is from the other psychologically, perceptually, or other type of sense. In various research fields, MDS has been successfully applied to identifying the spatial relationship of objects. For instance, sociologists have used MDS to obtain the structure of groups and organizations, based on members' perceptions of one another and their interaction patterns. In the case of sensor node localization, distances between pairs of sensor nodes are used as dissimilarity information. By applying MDS to distances, relative positions of the sensor nodes are estimated. The main advantage in using the MDS for positioning estimation is that even though dissimilarity information is error-prone, it can generate relatively high accurate position estimation [18].

3.1 Positioning scheme

Consider a network of n devices, where each device is assumed to measure dissimilarities between itself and all other devices. The dissimilarity between devices i and j, when it is measured by device i, is represented by p_{ij}. X_i denotes the estimated position of device i, which is represented as (x_i, y_i, z_i) in three-dimensional spaces. The Euclidean distance between X_i and X_j is denoted as d_{ij}. The central motivating concept of multidimensional scaling is to find a configuration of devices $\{X_1, X_2, \ldots, X_n\}$ in some space such that Euclidean distances d_{ij} between the devices correspond to the dissimilarities p_{ij}. There are several variants to solve the problem. The basic form of the solution is to find the configuration that minimizes Equation (1).

$$\sqrt{\frac{\sum_i \sum_j [f(p_{ij}) - d_{ij}]^2}{scale\_factor}} \quad \textbf{(1)}$$

In the equation, f is a continuous parametric monotonic function in metric MDS. In nonmetric MDS, f is related only on the rank order [13]. The positions of the devices that MDS estimates are relative positions, so they are always subject to rotation, translation, and scaling. Given at least three positions of the devices, the relative positions can be transformed to absolute positions. The relative positions are also important for the understanding of the network topology by themselves.

According to the method of measuring, we can define different types of dissimilarity. The simplest measure of dissimilarity is proximity, which discriminate whether two devices are within their communication ranges. We set p_{ij} as 1 in case device i receives a signal from device j. Otherwise, p_{ij} is set to any large value. The most complex measure is the Euclidean distance. Distance is known to be exponentially related to RSS according to the radio propagation characteristic.

Equation (2) is frequently adopted model that shows the relationship between the received signal strength and distance.

$$P_r = P_0 - 10n\log_{10}(d/l_0) + X_\sigma \quad (2)$$

,where P_r is the RSS in dB, P_0 the signal strength at distance l_0 from the transmitter, and n is the pathloss exponent. X_σ represents the shadow noise and is modeled as a normal random variable with the standard deviation σ dB [23]. d is distance between nodes. Typically, l_0 is 1 m. If we have a priori knowledge on these parameters, we can use the distance as dissimilarity. Even though dissimilarities p_{ij} and p_{ji} are actually the same, the two values are measured differently in real environments. Hence, we need to adjust them to be symmetrical. Dissimilarity of itself, p_{ii} is set to 0. The set of dissimilarities is called a dissimilarity matrix, D=[p_{ij}].

3.2 Dissimilarity matrix of WiFi APs

With respect to WiFi APs, there are several limitations to the method introduced in the previous section. First of all, WiFi APs usually do not have the capability of measuring distances between WiFi APs. To support this, WiFi APs' software needs to be modified. This modification is impractical in the real environment, since numerous numbers of APs are already deployed. Second, various algorithms are not applicable to WiFi APs due to the difficulty in making software changes. Hence, we need to devise a new method to obtain dissimilarities between WiFi APs without modifying already deployed WiFi APs. Our basic idea is to infer dissimilarities between WiFi APs from RSSs scanned on smartphones.

Let us assume that we have several radio scans gathered by a smartphone in a site. Each radio scan includes received signal strengths from visible APs at a position. In the case where n number of WiFi APs are scanned, a scan is represented as $scan_i$ = $\{(AP_1, rss_1), (AP_2, rss_2), \ldots, (AP_n, rss_n), timestamp\}_i$, where AP_n is the identifier of n_{th} AP and rss_n is its received signal strength. The scan set is denoted as $SCAN = \{scan_1, scan_2, \ldots, scan_m\}$, where total m radio scans are gathered. For the dissimilarity between AP i and a smartphone, we denote p_i, which is a function of rss_i, $p_i=g(rss_i)$. The dissimilarity between AP i and AP j, which is p_{ij}, is estimated based on p_i and p_j by a certain rule, that is, $p_{ij}= f(p_i, p_j)$. g and f are functions indicating the relationship between dissimilarity and RSS. These are explained later. Here, we have two arguments. One is how to extract p_{ij} from rss_i and rss_j. The other is how to manipulate several radio scans with the same AP list.

To discuss this further, we use an example as shown in Figure 1. A smartphone scans RSSs from AP i and AP j at a position. Dissimilarities p_i and p_j are calculated based on the measured rss_i and rss_j. For the convenience of explanation, we assume that parameters of the radio propagation model are given a priori, and the Euclidean distance is calculated as dissimilarity. Here, p_i is smaller than p_j. In the scan, there is no information about directions of APs. Hence, from the viewpoint of the smartphone, the positions of APs may be points on the circles. Hence, we cannot find the exact dissimilarity p_{ij}. As seen in the figure, minimum dissimilarity between AP i and AP j is p_i- p_j and maximum is p_i+p_j. Then, we may select one value between p_i-p_j and p_i+p_j as dissimilarity. Specifically, the dissimilarity is a distance between Y_i and a point of the inner circle. It is reasonable to take an average of all possible distances. For instance, when p_i equals p_j, p_{ij} is about $1.37 \times p_i$. In case $p_i >> p_j$, p_i may be used as p_{ij}. On the estimation of the dissimilarity, we can predict the extent of possible errors. In the figure, the error of an estimated dissimilarity is $2 \times p_j$ at worst. This means that error is possibly proportional to the smaller one between p_i and p_j.

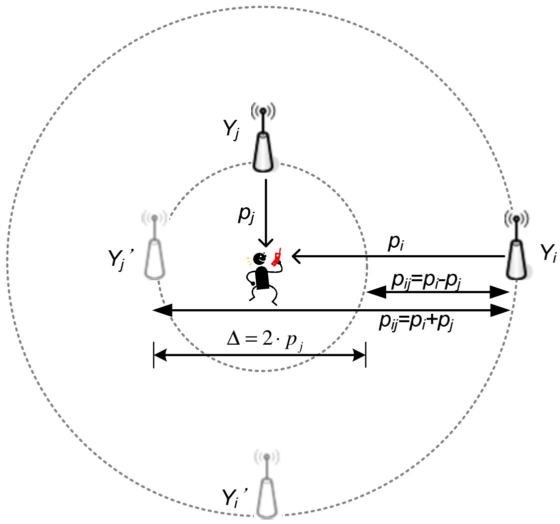

Fig. 1. At a point, a smartphone measures received signal strengths from two AP, Y_i and Y_j. The dissimilarities p_i and p_j are measured distances, and two circles show locatable positions of AP *i* and *j*, respectively. Then, the candidate of p_{ij} lies between p_i-p_j and p_i+p_j. $2 \times p_j$ is the extent of possible errors of the dissimilarity p_{ij}.

In the estimation of dissimilarity, another issue should be considered. In the beginning we assumed that we could get many scans in various positions. As a result, we have several scans including the same APs. Figure 2 illustrates the situation. Four scans are gathered at positions m_1, m_2, m_3, and m_4, and their dissimilarities measured by the smartphone are ($p_{j(1)}$, $p_{i(1)}$), ($p_{j(2)}$, $p_{i(2)}$), ($p_{j(3)}$, $p_{i(3)}$), and ($p_{j(4)}$, $p_{i(4)}$), respectively. Considering we estimate dissimilarity p_{ij} from them, estimating four dissimilarities, $p_{ij(1)}$, $p_{ij(2)}$, $p_{ij(3)}$ and $p_{ij(4)}$ and averaging them is one possible solution. However, since most scans are not measured on the shortest path connecting two APs, its value is usually larger than a real distance. Hence, we select and use a radio scan that is the most proper to infer a real distance from. Measurement positions m_1 and m_3 compared, $p_{j(1)}$ is equal to $p_{j(3)}$, but $p_{i(1)}$ is larger than $p_{i(3)}$. Then, m_3 is selected as a candidate. Compared with m_2, $p_{j(2)}$+$p_{i(2)}$ is smaller than $p_{j(3)}$+$p_{i(3)}$, then m_2 is newly selected. Since m_2 and m_4 are on the shortest path, $p_{j(2)}$+$p_{i(2)}$ is similar to $p_{j(4)}$+$p_{i(4)}$. Hence, m_2 or m_4 may be equally selected. Until now we did not consider that received signal strengths are disrupted by noise. In reality, the measured signal strength is not always consistent with the expected distance. Hence, in the selection of a radio scan, we take an approach to reduce an expected error. One clue is the extent of possible errors shown in Figure 1. It is proportional to the value of the smaller one of dissimilarities between

APs and a smartphone. Therefore, we finally select the radio scan with the smallest dissimilarity value, which in this case is m_2.

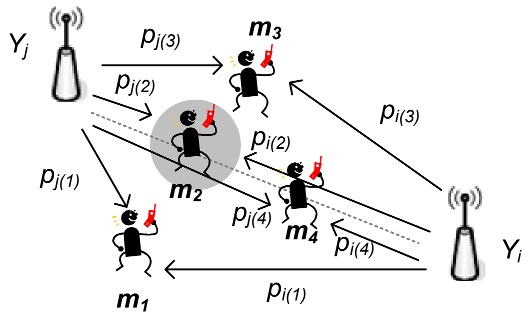

Fig. 2. At several points the received signal strengths from two APs, Y_i and Y_j are measured. The dashed line shows the shortest path between the two APs. Here, $p_{j(2)}<p_{j(1)}=p_{j(3)}<p_{j(4)}$ and $p_{i(4)}\ p_{i(3)}<p_{i(2)}<p_{i(1)}$. This shows the effect of measurement positions.

Now, we discuss the functions g and f that converts received signal strength to dissimilarity. In real environments, we do not often know parameters of the propagation model in Equation 2. Several previous researchers assumed a practical value for them. However, such assumptions are frequently fragile in many situations. Furthermore, even at the same distance, measured RSSs are different due to environmental factors including multipath fading. So, we do not try to convert received signal strength to a real distance. Instead, we quantize received signal strength into several levels. Considering the characteristic of exponential decaying of received signal strength, displacements between quantization thresholds should not be uniform. In the case where the range of received signal strength of WiFi APs is from –20dBm to –90dBm, our thresholds are –55, –70, –80, –85, and –90 in five levels, and its dissimilarities are 1, 2, 3, 4, and 5. When there is no scan including any pair of APs, we determine that they are not connected and set the dissimilarity between the two APs as any large value, for instance 9999. This quantization makes dissimilarities more robust to small variation of RSS.

In a large site, many pairs of APs are outside the communication range of the other; hence a significant portion of dissimilarities remain unmeasured. To solve this problem, we use a graph-based algorithm. Nodes in the graph represent APs, and weighted edges represent dissimilarities between two APs. In the graph, we find the shortest paths of all pairs. For the dissimilarity between unconnected APs, the distance of the shortest path is used. As the number of APs increases, we get a larger number of dissimilarities, which raises information for understanding the configuration of the APs. As a consequence, positions are more accurately estimated in a large site.

3.3 Multifloor AP positioning

Until now, radio scans are implicitly assumed to be gathered from a single floor. In general, people live in multifloor environments. Their radio scans are actually gathered from multiple floors. In the existing studies, floors of radio scans were labeled by user intervention. To encourage grass-roots radio gathering, however, the explicit labeling should not be requested to normal users. In some radio scans, RSSs from APs on adjacent floors are included. Hence, dissimilarities between APs in different floors are also obtained. We process radio scans gathered on several floors altogether, then generate a dissimilarity matrix, which includes both dissimilarities between inter-floor APs and dissimilarities between intra-floor APs. Theoretically, MDS has no restriction on the dimension of an analysis result. One constraint is that its performance is affected by the number of dissimilarities. So, if there are enough WiFi APs in a site, applying three-dimensional MDS to multifloor dissimilarities is possible. As a rule of thumb, there should be at least twice as many dissimilarities as parameters to be estimated, to assure an adequate degree of statistical stability [13]. That is, at least 12 APs are needed for a three-dimensional MDS.

Signals from an AP reach other floors following several paths. They penetrate through floors directly or detour through outside windows. Hence, in dissimilarities between inter-floor APs, different types of attenuations are included, and this breaks the consistency between a real distance and its dissimilarity. As a result, the estimated configuration is easily distorted. To overcome this, we divide multifloor AP positioning into three steps: *three-dimensional MDS*, *clustering of APs* and *multiple single-floor MDS*. In the first step, all the scans gathered from multiple floors are processed together to produce dissimilarities. Three-dimensional MDS is applied to the dissimilarities, and we get the three-dimensional positions of APs. Although three-dimensional MDS estimates a distorted configuration of APs, APs from the same floor tend to be placed nearby. In the second step, we distinguish floors of APs. If the number of floors is k, we cluster APs into k clusters based on the three-dimensional positions of APs. We assume that the number of floors is given. If whole floors are scanned, it is the number of floors of a building. APs in the same cluster are considered as being placed on the same floors. In the final step, APs from the same floor are processed using two-dimensional MDS separately. The process identifies positions of APs on multiple floors from radio scans without information on which floor scans are gathered.

3.4 Autonomous collaboration

To discover positions of WiFi APs universally, the participation of anonymous smartphone users is important; but this is not easily achieved if the technical barrier for participation is high. For instance, asking for ground truth for a scan or exact ranging to an AP hinders user participation. The proposed scheme only requires radio scan data; hence collaboration among smartphone users can be actively occurred. Here, we need to note that smartphones have different radio characteristics, which affect the dissimilarity estimation.

We discuss three cases of collaboration. In the first case, several users gather enough radio scans in a site and then estimate their own dissimilarity matrices. Each matrix reflects the different radio characteristics of smartphones. Classically, three-way MDS processes multiple dissimilarity matrices at a time. A representative three-way MDS is individual difference scaling, INDSCAL [13, 14], which aggregates the matrices based on common characteristics and produces a collaborated result. It, however, requires intensive computation. It is known that the result from INDSCAL is similar to the result from MDS for averaged data of all the dissimilarity matrices [13].

Second, users might gather enough radio scans to generate dissimilarity matrices at different sites. Then, AP positioning is performed individually and the estimated positions are patched together. In this case, radio characteristics of smartphones hardly affect the results.

Finally, it is the most concerning case that users gather radio scans sporadically in a large site, so individual scans are not sufficient to estimate a dissimilarity matrix. In this case, we estimate one dissimilarity matrix from the entire set of radio scans shared by all users. The problem is that signal strengths have different characteristics according to the device gathering them. Fortunately, this is mitigated by using a small quantization level and the fact that radio characteristics between smartphones do not differ as much as between different types of devices. In the case where we use proximity information only, this problem is almost eliminated.

4 Results

To validate the proposed algorithm, we conducted experiments in real environments and in simulation. To gather WiFi scans, we implemented software scanning WiFi signals on a smartphone, HTC Hero, running Android 1.5. With the smartphone, we gathered WiFi scans in a 20-story office building. Scans were performed for five floors, from the 10th floor to the 14th floor. As shown in Figure 3, each floor is 35 m by 50 m size and about 3 m in height. The outer wall is surrounded by glass. In the middle of each floor, there are restrooms, elevators, stairways, and meeting rooms, which are mainly constructed with concrete. The places where peoples work are divided by soft partitions. Each floor has about six WiFi APs attached on walls. The positions of WiFi APs are manually collected. The smartphone was placed on a hand or in a trouser pocket. On a scan, BSSID, SSID and RSS are stored. For analysis, some of scans were gathered with their ground truths. In experiments, only BSSIDs and RSSs are used. The strongest signals from APs are distributed from –23 dBm to –48dBm. Only two APs have signal strengths stronger than –30 dBm. The weakest signal strength was –96 dBm. In a scan, APs are listed in order of signal strengths. The AP with the strongest signal is listed first. From the scans, we found that two BSSIDs were transmitted by an AP. We gathered a total of 54,821 scans and found 57 APs in the scans whose maximum signal strength was stronger than –50 dBm. We expect that APs with maximum signal strength weaker than –50 dBm are located on other floors. Gathered scans are processed in a notebook computer running Windows XP (CPU: Intel(R) Core(TM) 2 Duo CPU 1.83 GHz; memory size: 2GB). We implemented a program to estimate dissimilarities in C. As a tool of MDS, we used the Matlab function *mdscale*, which performs nonmetric MDS by default. Results of MDS are relative configurations of WiFi APs. To evaluate accuracies of configurations, we need to match estimated configurations to the real configuration of APs. We performed this by using the Procrustes analysis, which finds the isotropic dilation, translation, reflection, and rotation that best match one configuration with another. We also used the Matlab function *procrustes*.

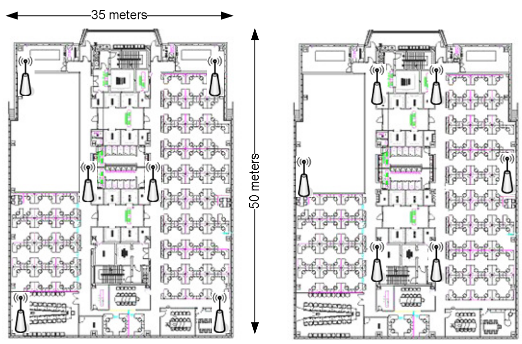

Fig. 3. Floor plans for the building. Left shows an even-numbered floor, and right shows an odd-numbered floor. Each floor has six APs setup by a network operator. Even-numbered floors have different deployments from odd-numbered floors.

4.1 Single-floor AP positioning

The performance of the proposed algorithm is affected by the way dissimilarities are obtained. First, we used five-level quantized RSS as dissimilarity. A brief summary of the procedure used to calculate the dissimilarity is as follows: first, whether the strongest RSS in a scan is larger than a threshold, which is –50 dBm in our implementation, is examined. Conceptually, the examination judges whether the scan is gathered near an AP. If it is satisfied, dissimilarities to all the APs in the scan are calculated. Otherwise, the scan is discarded. Dissimilarities are set as 1, 2, 3, 4, and 5 when RSS is larger than –47 dBm, –63 dBm, –75 dBm, –83 dBm, and –90 dBm, respectively. This process is repeated for all scans. If smaller dissimilarities are found in other scans, dissimilarities are updated as the new ones.

In the calculated dissimilarities, there were many mismatched pairs of p_{ij} and p_{ji}. Many factors cause this phenomenon. It is well known that symmetries of radio propagations are not guaranteed due to environmental effects, even though AP hardware is the same. In our case, dissimilarities between APs are measured indirectly by a smartphone. p_{ij} may be calculated from a scan measured nearest to AP *i*. In contrast, p_{ji} may be calculated from a measurement nearest to AP *j*. Positions of the two scans are not likely to have the same geometric conditions. One of the two scans might be closer to the line connecting two APs, and its dissimilarity has a smaller value. From this insight, we take the lower value as the dissimilarity in case the two dissimilarities are different.

Second, we used proximity information as dissimilarity. Simply, we decide a device is connected when a signal from an AP is visible. In this case, we set the dissimilarity to 1. However, this makes multidimensional scaling frequently fail to estimate positions of APs. In a small set of APs, this information is too limited to distinguish the configuration of APs. In the worst case, whole dissimilarities have the same value, for instance 1. To reduce this, we set a threshold to tighten the condition of proximity. In the experiments, we use –70 dBm as a proximity threshold. Since dissimilarities between unconnected APs are calculated by the shortest path-finding process, we obtain the effect that the amount of information increases.

Table 1. Experimental results of AP position estimation (error in meter).

	All APs (57 APs)		Duplicated APs removed (32 APs)	
	five-level quantization	Proximity only	five-level quantization	Proximity only
10th floor	9.22	8.93	5.34	8.53
11th floor	5.94	10.47	5.66	12.64
12th floor	9.49	14.64	**11.64**	13.84
13th floor	5.72	6.18	**5.48**	4.79
14th floor	8.77	12.32	10.08	12.41
Average	**7.85**	**10.50**	**7.64**	**10.44**

The results of our experiments are shown in the Table 1. In each floor, positions of 12 or 13 unique BSSIDs are estimated. The average estimation error when using a five-level quantization is 7.85 m. When proximity information is used, the average error is increased to 10.50 meters. In the case of proximity only, most of the dissimilarities have a value of 1 or 2, which is not a distinguishable condition. If the area of a site is larger, dissimilarities will have various values and the performance will be enhanced. This is proved in a simulation experiment. Among the BSSIDs, 25 are duplicated, and we removed these duplicated BSSIDs. As a result, six or seven BSSIDs remain for each floor. Again, we experimented with the refined set. The performance did not differ appreciably. From the result, we conclude that only the number of physically unique APs affects the performance. In the table, we show that performances of even-numbered floors are better than odd-numbered floors. There are two reasons for this. First, effective areas, which we define as the size of the convex hull of APs, are different. The effective area of an even-numbered floor is much smaller than the area of the odd-numbered floor. This differentiates the actual connectivity degrees, which represent how many APs are connected to an AP, and it is known that performance of MDS increases as the connectivity degree increases [16]. Second, the topology of even-numbered floor is beneficial. The topology of even numbered floors is round, but the topology of odd-numbered floors is more complex. By examining the layout of an even-numbered floor, we found many concrete walls in the middle, and concrete is difficult for radio signals to penetrate. Even though two APs in the middle are close, it is hard to identify small dissimilarities between them.

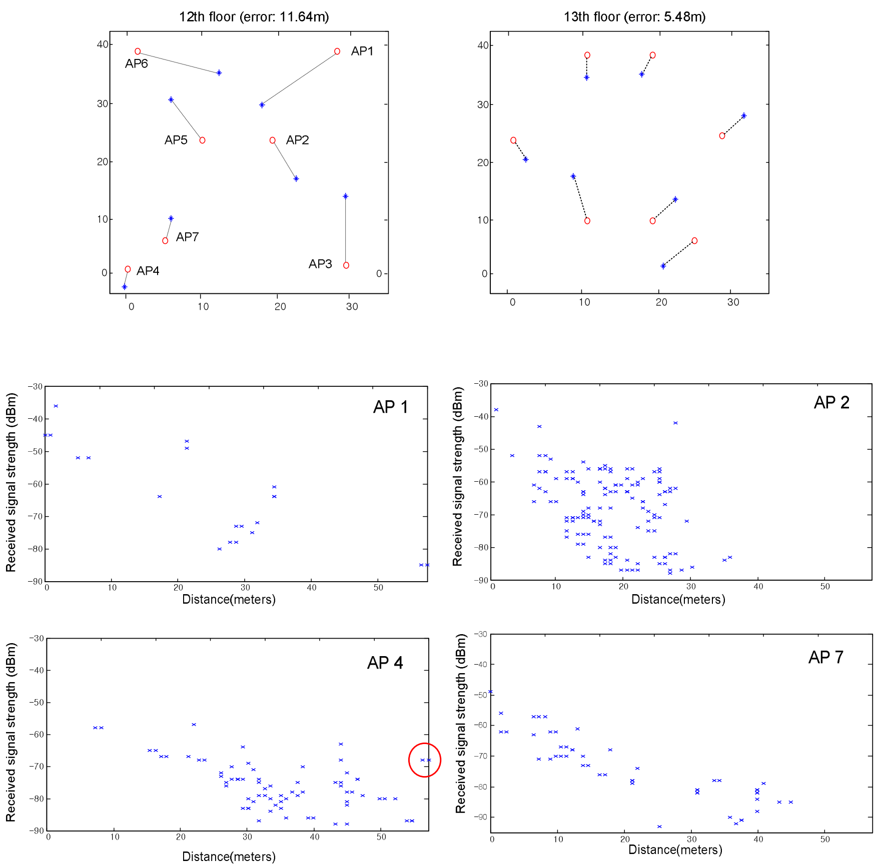

Fig. 4. The two graphs on the top show the estimation results of the 12th and 13th floors. "o" denotes the real position and * denotes the estimated position. Four graphs on the bottom depict scanned RSSs at various distances from four APs on the 12th floor, AP1, AP2, AP4, and AP7, respectively.

On the 13th floor, estimation performance is good (error of 5.48 m), and on the 12th floor estimation performance is poor (error of 11.64 m). Figure 4 shows the estimation results of the two floors. On the 13th floor, the configuration of estimated positions is similar to the original configuration. On the 12th floor, however, the estimated configuration is distorted compared with its original configuration. A noticeable error is that the position of AP1 moves too far inside. By analyzing the dissimilarities of the 12th floor, we found that dissimilarity between AP1 and AP4 was lower than it should have been. To identify the reason for this, we investigated all the scans gathered from the 12th floor.

The four graphs on the bottom in the Figure 4 show scatter plots of scanned RSSs according to the distances from AP1, AP2, AP4, and AP7, respectively. The range of distances from the APs is from zero to 56 m. Since AP2 is located in the middle of the site, its range is only from zero to 35 m. Among signals from AP4, there are unexpected large RSSs over a long distance (circled in red). At even 56 m from AP4, RSS was –68 dBm, and was measured around AP1. This caused a decrease in the distance between AP1 and AP4. Unlike other APs, AP7 was not set up by a network operator; it was a personal AP. Its signals were relatively low and were measured only in limited areas. Hence, dissimilarities to AP7 were largely found by the shortest path algorithm.

4.2 Multi-floor AP positioning

Even though expansions of algorithms to multifloor environments were mentioned, previous researches rarely conducted real experiments on them. In this study, we conducted a multifloor experiment. Radio scans gathered on five consecutive floors are used without floor information, but the number of floors is given.

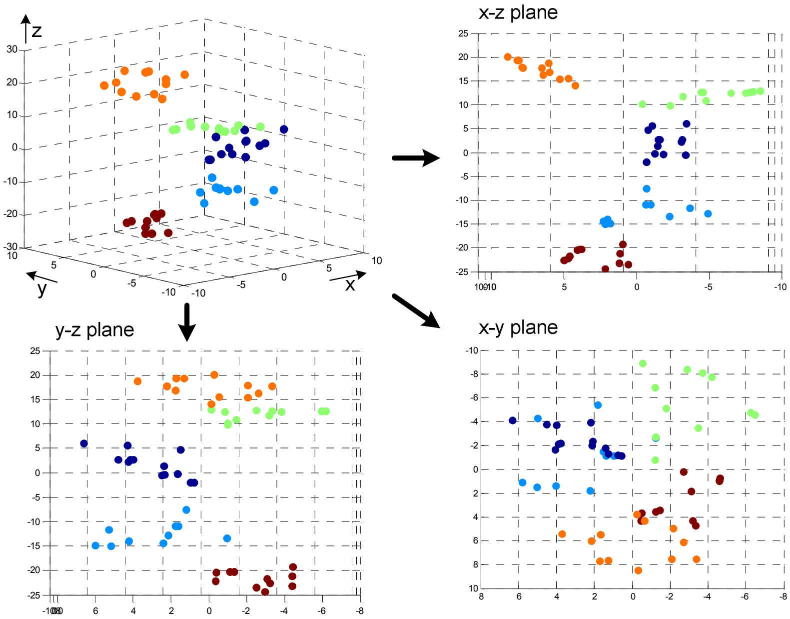

Fig. 5. A total of 57 APs on five floors are processed at the same time. Each dot shows the estimated positions of an AP. The positions are divided into five clusters. Each color shows a different floor.

All the scans are simultaneously processed to calculate dissimilarities between all pairs of APs on five floors. A total of 57×57 dissimilarities are estimated. We then applied three-dimensional MDS to the dissimilarities. The results are shown in Figure 5. Each dot denotes estimated positions of APs. We then divided the estimated positions of APs into five clusters. The Matlab function *cluster* was used for clustering. Each color in the figure shows a different cluster. In the *x–z* plane and *y–z* plane, we can see five distinctive clusters along the *z*-axis. To our astonishment, the clustering results show a 100% match to the real floors. This result verifies that we can identify the floor of AP with high accuracy using only radio scans collected on multiple floors.

In contrast to floor identification, two-dimensional positions on each floor are inaccurate. The *x–y* plane in Figure 5 depicts overlaid positions of all the floors. Each floor is biased to a different direction. When a radio signal passes through a floor, it is severely attenuated. This is frequently described by a floor attenuation factor [23]. However, since the outer wall of the building is made of glass, radio signals from other floors are easily received around the windows with only small attenuations. Even though dissimilarities are affected by these factors, they are not easily handled. Based on the clustering result, AP positioning is separately performed only with APs that are expected to be located on the same floor. This produces the same result as the single-floor experiment. The aligning problem between floors is still left, but this can be handled by radio matching methods [12].

4.3 Simulation analysis

To analyze the proposed algorithm, we conducted simulation experiments. To generate signal maps, we used radio propagation model as shown in Equation 2. *Po* at 1 m from an AP, *n* the pathloss exponent and the standard deviation of shadow noise are –27dBm, 3.4, and 9 dBm, respectively. The values are extracted from the radio scans used in our real experiments by a curve-fitting method.

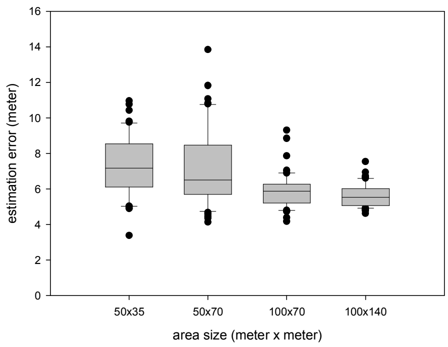

Fig. 6. While densities of APs remain the same, the size of areas is increased. The performance is enhanced as the size increases.

First, we analyzed the effect of the size of a site. We used the same density as our real experiment. In the 50×35 m^2 area, seven APs are used whose positions are randomly generated. In a doubled area (50×70 m^2), the number of APs is doubled. In areas of 100×70 m^2 and 100×140 m^2, 28 and 56 APs are used, respectively. Figure 6 shows the results. Even though the densities of APs are the same, the performance is enhanced and stable as the size of the area increases.

As the size increases, two benefits are apparent: connectivity degrees in an AP increase, and as the total number of APs increases, the number of dissimilarities also increases. These two facts increase the amount of information about the configuration of APs.

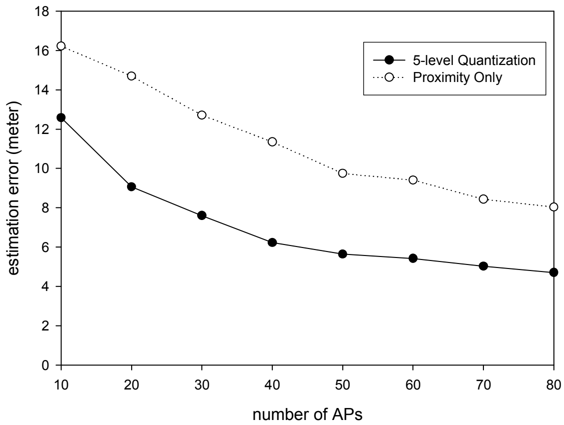

Fig. 7. Increase of connectivity degrees enhances the performance of AP positioning.

In contrast to the previous simulation, we fixed the area of site as 100×140 m^2, and increased the number of APs. Figure 7 shows that as the number of APs increases, the performance is enhanced. When the number of APs is large enough, the error of using proximity information only is around 9 m, and the error of five-level quantization is around 5 m.

5 Discussion and Future Work

We are concerned that the performance of the proposed scheme is suitable for location-based services. Outdoors, it may be enough to distinguish buildings, but indoors some location services will require more accurate results. We offer several possibilities to enhance the performance. First, performance improves as the size of an area widens. Second, an increased number of WiFi APs improves performance. In addition, a soft AP, which is a smartphone that acts like a WiFi AP temporarily, can be included to improve performance. Third, preprocessing radio scans can be applied. As shown in Figure 4, inconsistent signal strengths distort the geometric configuration of APs. We can use several stochastic methods, such as Kalman filter and Particle filter, to reduce this effect. We proposed an algorithm to estimate dissimilarities, but it can be replaced by a better method if there is one. In addition, there are many types of MDS techniques. We may use or develop a more appropriate version of an MDS technique. Despite all these variations, our overall scheme is still maintained.

In our research, we mentioned that relative positions can be transformed to absolute positions if there are at least three absolute positions of APs or radio scans. The important point is how easy it is to get three absolute positions of APs or radio scans. In general, many APs are seen outdoors where GPS signal can be received [24]. This means that we can frequently obtain absolute positions of radio scans at a boundary of a building. With the positions, absolute transformation can be performed. This is a practical situation.

The computational cost of the proposed algorithm is relatively low. In our experiment, execution of *mdscale* to 57×57 dissimilarities takes 3 s. In our opinion, the algorithm can be implemented in a smartphone or in a server. Hence, anyone can construct their own WiFi positioning system when needed.

With multifloor positioning, we distinguished floors of APs. Positions of APs are separately estimated in each floor, and then alignment of APs between adjacent floors is performed. If we can properly handle the floor attenuation factor and environmental effects across floors, we can remove all additional procedures. As our future works, we will improve the estimation algorithm of the dissimilarity matrix so that it is more robust to environmental effects including multifloor and perform an experiment of multiple users in a large site.

6 Conclusions

In this paper, we proposed an algorithm to determine a geographic configuration of WiFi APs from radio scans. Our method does not require ground truths for radio scans. Hence the assumption that we can obtain enough radio scans to build WiFi positioning system can be realized by the algorithm. We lowered the barrier of building a WiFi positioning system. Anyone using a smartphone can build their own positioning system or contribute information from their radio scans to help build a large positioning system.

To validate the feasibility of the proposed algorithm, we conducted real experiments in a 20-story office building. We gathered radio scans from five consecutive floors. On each floor, six or seven APs were found, and the average estimation error was 7.64 m. Based on the simulation result, the performance of the proposed algorithm can be enhanced as the size of the area widens. For multifloor, we are able to distinguish APs on different floors with high accuracy.

Acknowledgments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No.2010-0000405).

References

1. P. Bahl and V. N. Padmanabhan: RADAR: An in-building RF-based user location and tracking system. In Infocom (2000)
2. M. Youssef and A. Agrawala: The Horus WLAN Location Determination System. In MobiSys (2005)
3. Anthony LaMarca, Yatin Chawathe, Sunny Consolvo, Jeffrey Hightower, Ian Smith, James Scott, Tim Sohn, James Howard, Jeff Hughes, Fred Potter, Jason Tabert, Pauline Powledge, Gaetano Borriello, Bill Schilit: Place Lab: Device Positioning Using Radio Beacons in the Wild. In Pervasive (2005)
4. Y. Cheng, Y. Chawathe, A. LaMarca, and J. Krumm: Accuracy characterization for metropolitan-scale wi-fi localization. In Mobisys (2005)
5. M. Kim, J. Fielding, and D. Kotz: Risks of using AP locations discovered through war driving. In Pervasive (2006)
6. Mike Chen, Timothy Sohn, Dmitri Chmelev, Dirk Haehnel, Jeffrey Hightower, Jeff Hughes, Anthony LaMarca, Fred Potter, Ian Smith and Alex Varshavsky: Practical Metropolitan-scale Positioning for GSM Phones. In Ubicomp (2006)
7. Ekahau. http://www.ekahau.com/
8. Skyhook Wireless. http://ww.skyhookwireless.com/
9. A. Savvides, C.-C. Han, and M. B. Srivastava: Dynamic fine-grained localization in Ad-Hoc networks of sensors. In Mobicom (2001)
10. Oliver Woodman and Robert Harle: RF-Based Initialisation for Inertial Pedestrian Tracking. In Pervasive (2009)
11. D. Madigan, E. E., R. P. Martin, W. Ju, P. Krishnan, and A. Krishnakumar: Bayesian Indoor Positioning Systems. In Infocom (2005)
12. H. Wang, and et al: Indoor Localization in Multi-Floor Environments with Reduced Effort. In Percom (2010)
13. JB Kruskal and M Wish, "Multidimensional Scaling," Sage Publications (1978)
14. MAA Cox, TF Cox. Multidimensional scaling. Handbook of data visualization, Springer-Verlag (2008)
15. I. Borg and P.Groenen: Modern Multidimensional Scaling, Theory and Applications, Springer (1997)
16. Y. Shang, W. Ruml, Y. Zhang, and M. Fromherz: Localization from Mere Connectivity. In. MobiHoc (2003)
17. Y. Shang and W. Ruml: Improved MDS-Based Localization. In Infocom (2004)
18. Ji and H Zha: Sensor positioning in wireless ad-hoc sensor networks using multidimensional scaling. In Infocom (2004)
19. H. Lim, L. Kung, J. C. Hou, and H. Luo: Zero-Configuration, Robust Indoor Localization: Theory and Experimentation. In Infocom (2006)
20. A. Subramanian, P. Deshpande, J. Gaojgao, and S. Das: Drive-by localization of roadside WiFi networks. In Infocom (2008)
21. Brian Ferris, Dieter Fox, Neil Lawrence: Wi-Fi-SLAM Using Gaussian Process Latent Variable Models. In IJCAI (2007)
22. Krishna Kant Chintalapudi, et al.: Indoor Localization Without the Pain. In Mobicom (2010)
23. Rappaport, T. *Wireless Communications — Principles and Practice.* Prentice-Hall, Englewood Cliffs (1996)
24. B. Hoffman-Wellenhof, H. Lichteneeger, and J. Collins: Global Positioning System: Theory and Practice (4th ed.). New York: Springer-Verlag (1997)