

# An Improved Location Estimation Method for Wifi Fingerprint-based Indoor Localization

Jungmin So, Joo-Yub Lee, Cheal-Hwan Yoon and Hyunjae Park

*Dept. of Computer Engineering, Hallym University  
jsol@hallym.ac.kr*

## **Abstract**

*The Accurate indoor localization is a challenging task due to the absence of GPS. Among numerous proposals, Wifi fingerprint-based localization is one of the most promising approach, since most buildings are nowadays equipped with Wifi access points for wireless network coverage. Due to the nature of Wifi access points in which any user can deploy and manage their own, fingerprints from some access points lead to estimation errors. Location estimation algorithms should consider these factors and be able to locate users with low error distance. Finding the nearest neighbor using Euclidean distance in signal space is most widely used method in estimating location. However, this paper shows that Euclidean distance is prone to error when unstable access points are present. Also, Euclidean distance does not differentiate strong signals and weak signals, which can also mislead location estimation. We propose a different way to determine the nearest neighbor, which penalizes signals from unstable access points, and signifies strong signals compared to weak signals. Experiments with real measurements show that the proposed algorithm reduces mean error distance by 57% and 90-percentile error distance by 64% compared to the Euclidean distance method.*

**Keywords:** *indoor localization, Wifi fingerprint, location estimation*

## **1. Introduction**

With the wide spread of smartphones, location-based services are becoming prevalent. In order to enable location-based services, accurately positioning users in an indoor environment is essential. Indoor localization is challenging compared to outdoor localization, since GPS signals are not available indoors. Also, higher accuracy is typically needed in indoor environment than outdoor, since users should locate stores in a shopping mall, rooms in an office building, or even isles in a grocery store.

Numerous approaches have been proposed that find user locations without aid of GPS. Among these approaches, Wifi-based localization has gained major attention, due to the fact that most buildings are now getting equipped with Wifi access points. Using these access points as location indicators removes the need for additional infrastructure cost. There are two approaches of using Wifi for indoor localization: fingerprint-based and model-based. The fingerprint-based method builds a “radio map” of the building at offline, which records signal strengths (RSSI: Received Signal Strength Index) received from nearby access points at known locations called “reference points”. At online, a client at an unknown location collects signal strengths from nearby access points and sends them to the location server, which possesses the radio map. The server compares the signal strength vector collected at user location with fingerprints of reference points in order to estimate user location.

Communication between client and server can take place through any wireless network interface, such as 3G, LTE and also WiFi [14].

The model-based method, on the other hand, does not use radio map but relies on indoor path-loss model in order to estimate user location. Signal strengths from different access points are converted into distances using a path-loss model such as log-distance model, and the user location is computed using triangulation. Generally, a pure model-based method which does not rely on a radio map shows very low accuracy, due to the fading characteristics of wireless channel. Walls, objects, and people moving around all distort the signal and the received signal strength deviates from the model.

When using fingerprint-based method to estimate location, the major question to answer is: how do we determine the user location? A widely used answer to this question is to select the “nearest neighbor (NN)”, which is the closest reference point derived from the signal strength vectors. If the reference points are sparsely located, selecting  $k$  nearest neighbors (kNN) and averaging their locations can help improving location accuracy. The next question is: who is the nearest neighbor? A typical answer to this question is the reference point that has shortest Euclidean distance in signal space.

These two questions are enough if signals from all access points are “stable”, which is unfortunately not true. An access point may be active when building the map, but turned off when a user tries to estimate his location. It could be the other way around. Sometimes beacons may not be received from some access points during scanning, even if they are active. There are access points carried around by users, such as WiBro-Wifi converter or devices with Wifi tethering enabled. With high probability these access points will not be active and placed at the same location at both the time of radio map generation and the location estimation.

The problem with these unstable access points is that they can lead to serious estimation errors when their signal strengths are fed into the standard estimation algorithm using Euclidean distance in signal space. The signal strength difference caused by unstable access points could be large and thus dominate the location estimation, as observed in the experiments (described in the next section). Also, estimation algorithm using Euclidean distance does not differentiate between strong signals and weak signals, although it is logical to give high priority to strong signals.

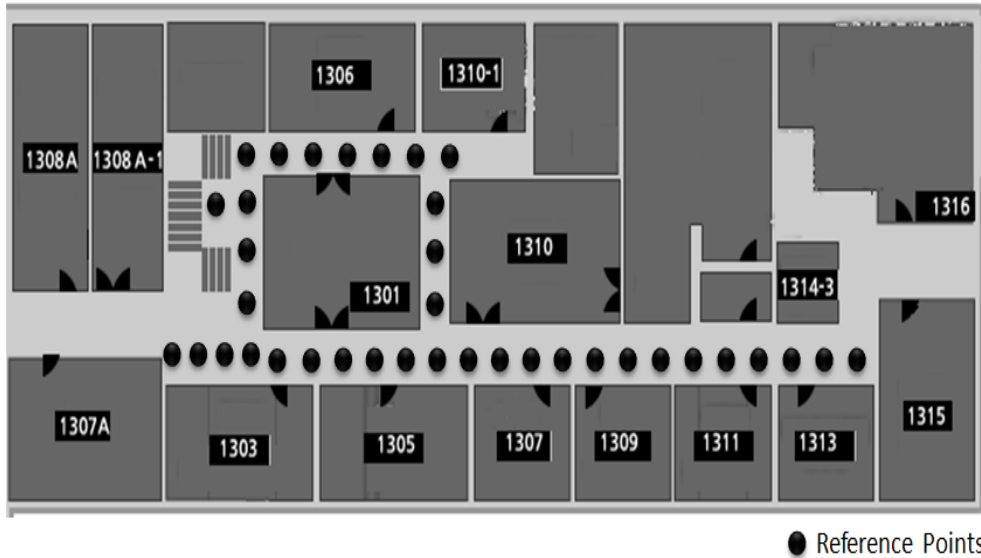
We need a location estimation algorithm that penalizes unstable access points, and emphasizes strong signals. A new estimation algorithm is designed under these criteria, and is shown to provide higher accuracy compared to the original Euclidean distance-based algorithm. In Section 2, we analyze location estimation algorithms through experiments and discuss observations made from the experiments. In Section 3, we propose a new estimation algorithm and evaluate its performance in comparison with other algorithms. In Section 4, we discuss related works regarding indoor localization. Finally, we conclude with remarks on the future work in Section 5.

## **2. Analysis of Location Estimation Algorithms**

### **2.1. Initial experiment**

The experiments took place in the Engineering building at Hallym University. The floor plan of the experiment area is shown in Figure 1. 36 reference points (marked as the black dots in the figure) were selected along the corridor, approximately 1.5 meters apart. We have used an HTC Magic smartphone for radio map generation as well as location estimation. In the experiment region, signals were received from a total of 26 access points.

The Wifi scan returns the RSSI of neighboring access points approximately every 1 second. (It has been observed that other smartphones have different scan periods.) To generate a radio map, we have recorded 50 RSSI values from each access point, in every reference point. The RSSI values were recorded in dBm unit. If no signal is received from a particular AP, it is treated as -99dBm, which is a value lower than minimum observed RSSI. The radio map was saved in a server.



**Figure 1. The floor map of Seongho Hall where the experiment took place. The black dots are the reference points**

Once the radio map was generated, we have tested location estimation at 144 different places in the region. None of the test locations are exactly the same with reference point locations. The location estimation procedure is as follows.

First, 50 RSSI values of an access point were averaged into a single a value. Thus each reference point is represented by a signal strength vector  $s^{(i)} = (s_1^{(i)}, s_2^{(i)}, \dots, s_k^{(i)})$ , where  $k$  is the number of access points. Second, RSSI values measured at the test location are also averaged and made into a vector  $t = (t_1, t_2, \dots, t_k)$ . Third, the Euclidean signal distance from the test location is calculated as follows:

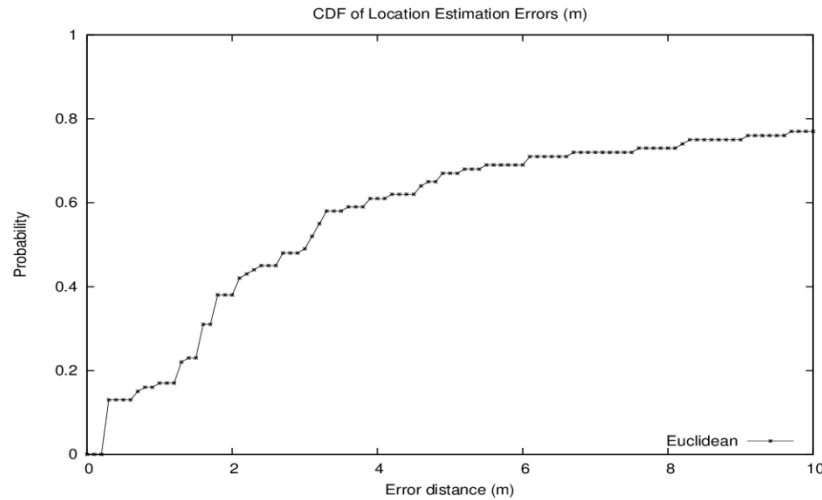
$$l^{(i)} = \sqrt{(s_1^{(i)} - t_1)^2 + (s_2^{(i)} - t_2)^2 + \dots + (s_k^{(i)} - t_k)^2}$$

Finally, the reference point which has the minimum Euclidean signal distance is selected as the estimated location.

Estimating location based on the Euclidean signal distance is one of the most widely used location estimation algorithm [1]. Model-based estimation algorithms exist, but they are not shown to produce higher accuracy than Euclidean-distance based method. The Horus system [2] uses a probabilistic approach, which they claim to achieve a higher accuracy than deterministic method. This may be possible, but it requires a large number samples per access point at each point in order to make a distribution.

## 2.2. Analysis of the Result

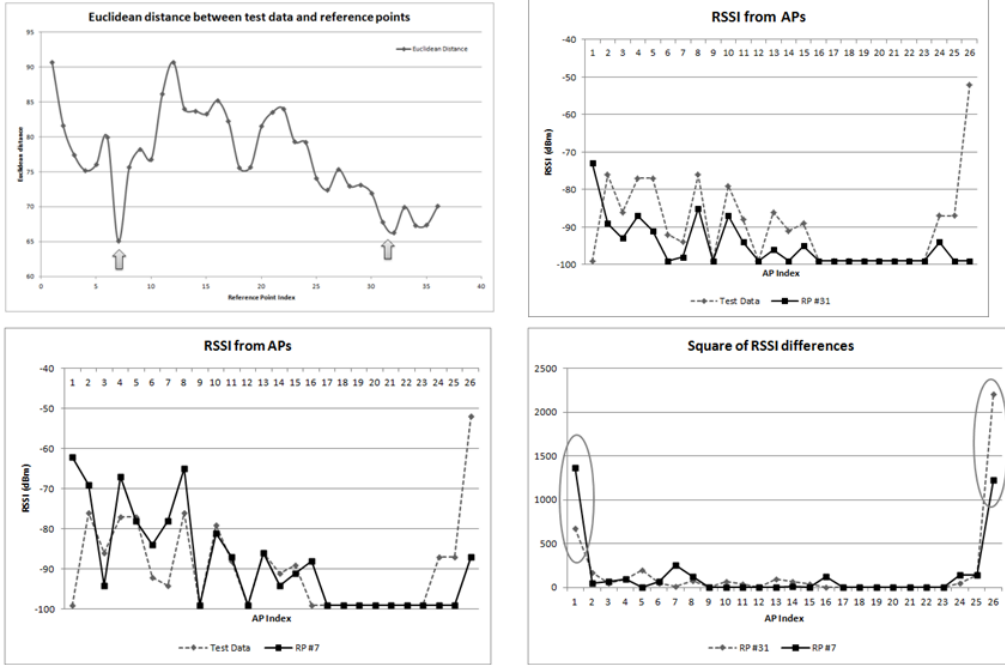
Figure 2 is the result of the first experiment. The median error distance is 3.10m, meaning that for half of the test locations the estimation algorithm returned with an error larger than 3 meters. The 80% error was 10.6 meters, and 90% error was 13.6 meters.



**Figure 2. CDF of error distance for Euclidean distance-based location estimation**

The question is: is the location error caused from signal variability or erroneous estimation algorithm? To find the answer, we analyze a particular test location which showed a large estimation error. Figure 3(a) shows the Euclidean distance between test data and 36 reference points. The X-axis is the index of reference points. The test location was somewhere between reference point #31 and #32, but the estimated location is #7, which results in a location error of 18.97 meters. Even with  $k$  nearest neighbor method, the location error will be still high. While other reference points near RP #7 shows much greater distance from the test location, RP #7 strangely shows low distance. Why is that?

Figure 3(b), 3(c), and 3(d) reveals a clue to answering the question. In the figure, RSSI differences of AP #1 and AP #26 dominates the result. At test location, the RSSI of AP #26 is -52 dBm which is very high. The RSSI of AP #26 is -87dBm and -99dBm (no signal), for RP #31 and RP #7, respectively. Because of their large difference, their square value dominates the computation of the Euclidean distance. However, can we say that 35dB difference is a strong evidence that RP #7 is closer to the test location than RP #31, which did not receive signal from AP #26? Rather, it is more logical to conclude that both reference points are far from the test location. Since no reference point has similar RSSI value with the test location for the particular AP, we can suspect that the AP has moved since the time of radio map generation, and thus the fingerprints of this AP should not be considered significantly.



**Figure 3. (a) Analysis of a particular case. The actual location is somewhere between RP #31 and #32, but the estimation algorithm returns RP #7 as the closest neighbor. (b) RSSI of APs for test location and RP #31. (c) RSSI of APs for test location and RP #7. (d) Score of each AP for RP #31 and #7**

### 3. Proposed Scheme and Evaluation

We propose new schemes to find the nearest reference point using Wifi signal strengths. The idea is to make the algorithm so that the impact of large differences in RSSIs is reduced. There are two approaches that serve this goal. The first scheme (Scheme-1) is to filter out large differences so that if RSSI difference is larger than a certain threshold, the distance measure is no longer increased. The distance measure using this scheme can be described as Figure 4(a), and can be written as follows.

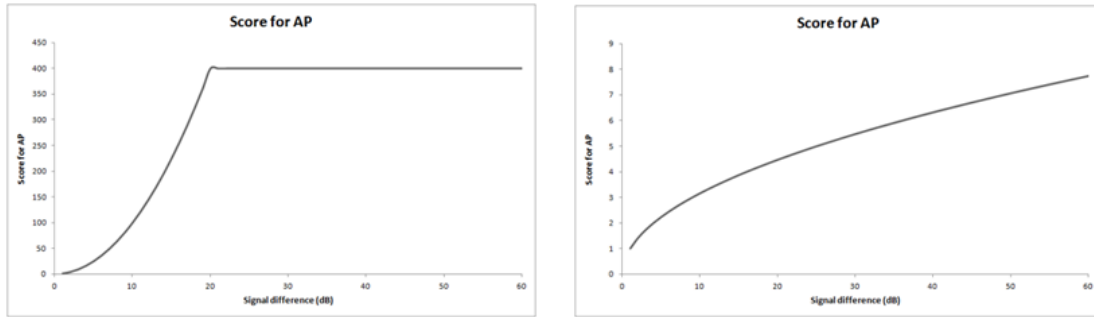
$$l^{(i)} = \sqrt{c_1^{(i)^2} + c_2^{(i)^2} + \dots + c_k^{(i)^2}$$

$$c_j^{(i)} = s_j^{(i)} - t_j \quad \text{if } |s_j^{(i)} - t_j| \leq D_{th}$$

$$c_j^{(i)} = D_{th} \quad \text{if } |s_j^{(i)} - t_j| > D_{th}$$

The second scheme (Scheme-2) is to use a power coefficient that is less than 1, instead of 2 as in Euclidean distance. This will reduce the impact of large RSSI differences, and increase the importance of exact match. The scheme is described in Figure 4(b), and can be written as follows.

$$l^{(i)} = c_1^{(i)\alpha} + c_2^{(i)\alpha} + \dots + c_k^{(i)\alpha}, \quad 0 < \alpha < 1$$



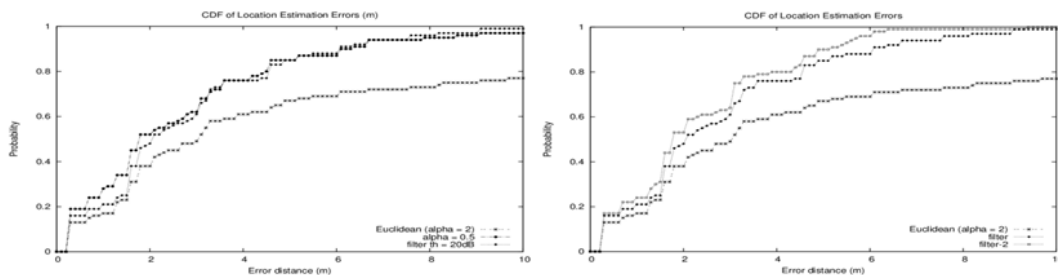
**Figure 4. Alternatives to Euclidean distance algorithm. (a) Scheme-1: If the RSSI difference is higher than a certain threshold, stop increasing score. (b) Scheme-2: Use a power coefficient less than 1**

The performance results for these two schemes are shown in Figure 5(a). As shown in the figure, the two schemes outperform Euclidean distance-based scheme significantly, whereas their performances are comparable. The problem with Scheme-1 is how to decide a proper threshold, whereas the problem with Scheme-2 is that small differences of RSSI, which is very likely to occur due to fading, has a large impact on estimation. In our experiment, the error distance of Scheme-1 was slightly smaller than that of Scheme-2.

During the case analysis, we observed that the signal strength pattern is similar between the test location and its nearest reference point, but the RSSIs are shifted. (The case analysis is omitted due to space limit.) This is expected to be caused from various reasons, such as user orientation, user height, how the user holds the smartphone, *etc.* The RSSI shift leads to wrong location estimation, and sometimes results in very high error distance. In order to compensate for this effect, we shift the RSSIs of test data inside a certain range, and choose the minimum signal distance in this range. Specifically, we compute the signal distance as follows. (We use this RSSI shift mechanism on top of Scheme-1.)

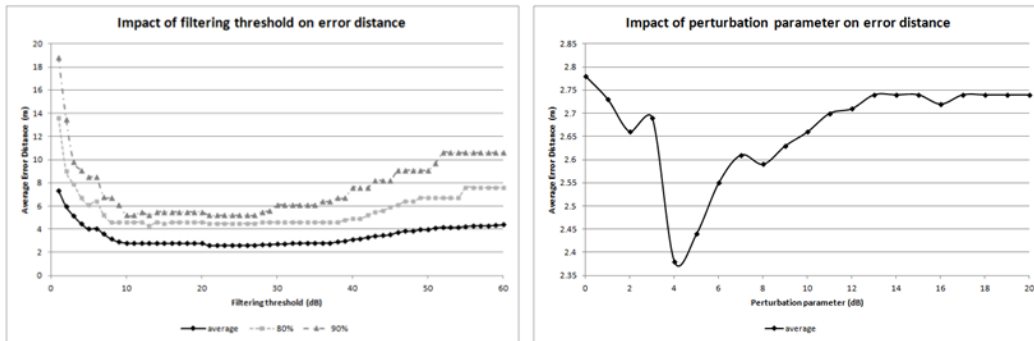
$$l^{(i)} = \sqrt{(c_1^{(i)} + d)^2 + (c_2^{(i)} + d)^2 + \dots + (c_k^{(i)} + d)^2}, \quad d_{min} \leq d \leq d_{max}$$

The performance of RSSI shift mechanism is shown in Figure 5(b). The average error distance is reduced by 15% using this mechanism. With Scheme-1 and RSSI shift mechanism combined together, the average error distance is reduced by 57%, and 90th-percentile error distance is reduced by 64%.



**Figure 5. (a) CDF of error distance for Euclidean distance-based algorithm, Scheme-1 and Scheme-2. (b) CDF of error distance for Scheme-1 with RSSI shift mechanism**

Finally, we studied the impact of parameters on average error distance. The impact of filter threshold is shown in Figure 6(a). The average error distance increases if the threshold is too small or too large, but is consistent somewhere between 10dB and 30dB. The optimal threshold will depend on the density of access points and reference points, but since there is no sharp optimal point, the environmental effect is expected to be small as long as reference points are not located too sparsely. The impact of RSSI shift range is shown in Figure 6(b). In this experiment setting, a range of (-4, 4) produce the best result. In general, 3 to 10dB shift range provide improved results, whereas choosing a very large shift range can be harmful.



**Figure 6. (a) Impact of filter threshold on average error distance. (b) Impact of RSSI shift range on average error distance**

#### 4. Related Work

Research on indoor localization has been active for more than a decade [13,15,16]. The Active Badge system [4] is one of the earliest localization systems that uses Infrared (IR) transmissions. Rooms in a building are equipped with IR sensors, which detect signals transmitted from badges worn by users. The Active Bat system [5] uses ultrasound signals to infer user location. A major drawback with these systems is the large infrastructure cost, because specialized sensors need to be deployed in the building, and users need to carry (or wear) special devices too.

The RADAR system [1] is the first RF-based localization system that uses offline-generated fingerprints. At calibration phase, RSSI fingerprints are recorded at reference points and stored in a server. At positioning phase, the client trying to locate its position scans channel for beacons sent from access points, and send their RSSI values to the server. The server runs the estimation algorithm to find the nearest neighbor or  $k$  nearest neighbors, and finally returns the estimated location to the client. The RADAR system uses Euclidean distance of signal strengths in order to determine the nearest neighbor. King et al. developed the COMPASS system [3], which considers orientation of the user measuring signals. Orientation is significant because human body absorbs wireless signals. If the signal passes through human body before reaching the destination, it can cause serious attenuation which can result in estimation error. Thus, in the COMPASS system, multiple radio maps are constructed, one for each direction that the user is facing. At online phase, the user equipped with a digital compass sends its orientation along with the RSSI vector to the server. The server selects which radio map to use according to the user's orientation. This system improves localization accuracy at the cost of more calibration effort. Youssef et al. developed the Horus system [2], which combines various techniques in order to improve accuracy. The major component is probability-based estimation, which computes probability of each

reference point being the nearest neighbor. The authors claim that this probabilistic approach achieves higher accuracy compared to the deterministic approach used in systems such as RADAR. The drawback of this approach is that a large number of samples are needed for each access points, in order to create distribution. This increases amount of time consumed for calibration, as well as computation time and storage.

Systems have been proposed that use means other than Wifi signals as location indicators. Chung, *et al.*, [6] proposes a system which uses geo-magnetism for localization. The observation is that in a building, the magnetic strength which distorts the output of digital compass varies spatially, but is static temporally. Thus, it could be used as information for localization. Tarzia, *et al.*, proposes Batphone [9] which uses acoustic background spectrum as location indicator, and Chung, *et al.*, [6] proposed to use FM radio signals as location indicators. The common characteristic of these signals is that they vary over space, and they are static over time. Any other signals that have this characteristic can be used as location indicators. However, these signals themselves do not produce high accuracy compared to Wifi signals. Rather, they could be used with the Wifi signals to further improve localization accuracy.

Recent trend in indoor localization is focused in removing the cost for calibration. The EZ system [11] uses model-based location estimation plus location fixes using GPS when the user is near doors or windows. The server collects RSSI vectors from users, and runs a genetic algorithm using physical constraints to locate the user. The UnLoc system [12] uses motion tracking using sensors included in the smartphones, in order to locate the user. The high error of motion tracking is corrected whenever a user walks through the “landmarks”, which are locations in the building that can pinpoint the user location, such as elevators, escalators, and stairs. The Zee system [10] builds the radio map automatically by crowdsourcing. When users walk around with their smartphones, their locations are estimated from motion tracking, and the estimated location and RSSI vector that the location are sent to the server to build the radio map. Once the radio map is established, it can help reducing the localization errors of caused from motion tracking.

Although there are numerous approaches and systems for indoor localization, a good estimation algorithm is essential in any fingerprint-based localization system. Automating the calibration phase can significantly improve the practical value of indoor localization system, but the localization error tends to become higher. The proposed algorithm does not depend on a specific system component, and can help improve localization accuracy of fingerprint-based indoor localization systems.

## 5. Conclusion and Future Work

Location estimation algorithm is at the heart of an indoor localization system. A good estimation algorithm should not only produce low-error answers, but also be resilient to real-life events such as unstable access points. We observe that Euclidean distance based nearest neighbor selection, which is one of the most widely used methods, is prone to error when unstable access points are present. Also, the Euclidean distance method does not emphasize strong signals over weak signals, which does not match well with propagation characteristics. We propose a location estimation algorithm that penalizes unstable access points, as well as signifies strong signals. The performance evaluation from real measurements shows that the proposed algorithm reduces localization errors compared to the original Euclidean distance-based algorithm.



As a future work, we plan to design an algorithm that differentiates access points by their contribution to localization. Some access points have high variations, whereas other have lower variations. Some access points are placed together closely, or even frequently a single access point acts a multiple access points by using multiple SSIDs. Since these access points can skew the estimated location, signal strengths from these access points should be weighed according to their contribution. Computing weights for access points based on analysis of reference point fingerprints can further increase localization accuracy.

## Acknowledgements

This work is supported by Hallym University Research Fund 2011 (HRF-2011-029(II)).

## References

- [1] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system", Proceedings of IEEE INFOCOM, (2000), pp. 775-784.
- [2] M. Youssef and A. Agrawala, "The Horus location determination system", Springer Wireless Networks, (2008).
- [3] T. King, S. Kopf, T. Haenselmann, C. Lubberger and W. Effelsberg, "COMPASS: A probabilistic indoor positioning system based on 802.11 and Digital Compasses", Proceedings of WINTECH, (2006).
- [4] R. Want, A. Hopper, V. Falcao and J. Gibbons, "The Active Badge Location System", ACM Transactions on Information Systems, (1992), pp. 91-102.
- [5] Ward, Jones, and A. Hopper, "A New Location Technique for the Active Office", IEEE Personal Communications, vol. 4, no. 5 (1997), pp. 42-44.
- [6] J. Chung, M. Donahoe, C. Schmandt and I. Kim, "Indoor Location Sensing using Geo-Magnetism", Proceedings of ACM MobiSys, (2011).
- [7] C. Laoudias, M. P. Michaelides and C. G. Panayiotou, "Fault Detection and Mitigation in WLAN RSS Fingerprint-based Positioning", Proceedings of Int. Conf. on Indoor Positioning and Indoor Navigation, (2011).
- [8] Y. Chen, D. Lymberopoulos, J. Liu and B. Priyath, "FM-based Indoor Localization", Proceedings of ACM MobiSys, (2012).
- [9] S. Tarzia, P. Dinda, R. Dick and G. Memik, "Indoor Localization without Infrastructure using the Acoustic Background Spectrum", Proceedings of ACM MobiSys, (2011).
- [10] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan and R. Sen, "Zee: Zero-Effort Crowdsourcing for Indoor Localization", Proceedings of ACM MobiCom, (2012).
- [11] K. Chintalapudi, A. P. Iyer and V. N. Padmanabhan, "Indoor Localization Without the Pain", Proceedings of ACM MobiCom, (2010).
- [12] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef and R. R. Choudhury, "No Need to War-Drive: Unsupervised Indoor Localization", Proceedings of ACM MobiSys (2012).
- [13] H. Liu, H. Darabi, P. Banerjee and J. Liu, "Survey of Wireless Indoor Positioning Techniques and Systems", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 37, no. 6, (2007), pp. 1067-1080.
- [14] H. Jung, S. Park, J. Lee and Y. Ko, "Efficient Data Deduplication System Considering File Modification Pattern", International Journal of Security and Its Applications (IJSIA), vol. 6, no. 2, (2012).
- [15] G. Kang, T. Perennou and M. Diaz, "Barycentric Location Estimation for Wireless Network Indoors Localization", International Journal of Hybrid Information Technology (IJHIT), vol. 2, no. 1, (2009).
- [16] C. Chen, D. Wang and Y. Huang, "A Novel Method for Unstable-signal Sensor Localization in Smart Home Environments", In International Journal of Smart Home (IJSIH), vol. 2, no. 3, (2008).

## Authors



**Jungmin So** received his B.S. degree in computer engineering from Seoul National University in 2001, and Ph.D. degree in Computer Science from University of Illinois at Urbana-Champaign in 2006. He is currently an assistant professor in Department of Computer Engineering, Hallym University. His research interests include wireless networking and mobile computing.



**Joo-Yub Lee** received B.S. in Computer Engineering from Hallym University in 2013. He is currently a Graduate student in Department of Computer Engineering, Hallym University, pursuing a Master's degree. His research interests include wireless networking and large-scale systems.



**Cheal-Hwan Yoon** graduated from Sokcho high school in 2006. He is currently an undergraduate student in Department of Computer Engineering, Hallym University. His research interests include wireless networking and mobile computing.



**Hyunjae Park** graduated from Boseong high school in 2009. He is currently an undergraduate student in Department of Computer Engineering, Hallym University. His research interests include wireless networking and mobile computing.