Design of Aging-Resistant Wi-Fi Fingerprint-based Localization System with Continuous Active Learning

Youngsam Kim*, Soohyung Kim*

*Information Security Research Division, ETRI, Daejeon, Korea

kimzt@etri.re.kr, lifewsky@etri.re.kr

***Abstract*—Wi-Fi fingerprint-based localization systems are widely used for indoor localization as it only needs Wi-Fi network infrastructure that exists almost everywhere nowadays. However, it can be vulnerable to environmental change if Wi-Fi fingerprint-based localization system uses fixed Wi-Fi fingerprint database as training dataset and has no method for updating training dataset. In this paper, we propose AR-WFL system including update phase that can reflect environmental change periodically and prevent performance degradation. The proposed AR-WFL system is based on crowdsourcing and no dedicated annotator exists. In addition, we adopt active learning scheme with uncertainty selective sampling algorithm to maximize cost-efficiency of the update phase. We evaluate the performance of the update phase as location estimation accuracy using a dataset we collected for 2 months. It shows that average accuracy is increased by 1.83%p using update phase with uncertainty sampling algorithm compared with the system without an update phase.**

***Keyword*—Indoor localization, Wi-Fi fingerprint, Active learning, Selective sampling, Database update**

I. INTRODUCTION

Recently, context-awareness is noted in many areas like security and healthcare. People spend most of their time in the indoor environment [1] and indoor localization inevitably becoming one of the key technologies to build a context-awareness system. For about two decades, many indoor localization systems have been researched based on Wi-Fi fingerprint since Wi-Fi fingerprint-based localization(WFL) system has been introduced in [2]. According to [3], research topics can be divided in two aspects, one is for higher accuracy, the other is to efficiently deploy the system. In this paper, we focus on the latter aspect, especially the method that keeps Wi-Fi fingerprint database up-to-date.

The performance of WFL system mainly depends on Wi-Fi fingerprint database. However, Wi-Fi fingerprint database is fixed at the time it was firstly constructed in traditional systems. As a result, the database cannot reflect the environmental changes of Wi-Fi infrastructure, arrangement of furniture and so on. Thus, accuracy of location estimation decreased as time goes by. To address this issue, database update method that can reflect recent Wi-Fi fingerprint is very necessary. According to [3], Wi-Fi fingerprint database update method can be divided in two ways, using additional infrastructure and not using it. Additional infrastructure-based update methods may cost a lot because they need pre-installed infrastructure like sensor networks. Meanwhile, infrastructure-less update systems don't need additional infrastructure but need users' help to annotate the unlabeled Wi-Fi fingerprint data.

In this paper, we propose aging-resistant WFL (AR-WFL) system that uses crowdsourcing scheme. Designing the AR-WFL system with only Wi-Fi fingerprint is quite challenging for the following reason. To update Wi-Fi fingerprint database, the system needs labeled Wi-Fi fingerprint data from a human annotator. In crowdsourcing-based system, a label of Wi-Fi fingerprint should be obtained by querying to users who are not dedicated annotators. To reduce user intervention for labeling, the system needs to minimize the number of queries and it also prevent the performance degradation at the same time.

To address these challenge, we designed AR-WFL system based on selective sampling manager and active query manager. Selective sampling manager selects samples based on a pool-based selective sampling algorithm [4]. A pool accumulates Wi-Fi fingerprints from users' online query for location estimation. Active query manager gives a user a labeling query to obtain a label for a location. This manager monitors all online queries from users. When Wi-Fi fingerprint of online query is similar to one of the Wi-Fi fingerprint in the pool, the manager immediately queries a label to the user. If the user responds current his/her location, system can obtain reliable label. We describe these components in section 3, more detail.

We evaluate the performance of the proposed AR-WFL system using 2-month data. To evaluate aging-resistance characteristic, we show how far location estimation accuracy is

Manuscript received Aug. 3rd, 2017. This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No.2015-0-00168), and a follow-up of the invited journal to the accepted & presented paper of the 19th International Conference on Advanced Communication Technology (ICACT2017), and Grant ID is 20170071.

Youngsam Kim is with Electronics and Telecommunications Research Institute, Daejeon, Korea (e-mail: kimzt@etri.re.kr).

Soohyung Kim is with Electronics and Telecommunications Research Institute, Daejeon, Korea (corresponding author to provide phone: +82-42-860-1520; e-mail: lifewsky@etri.re.kr).

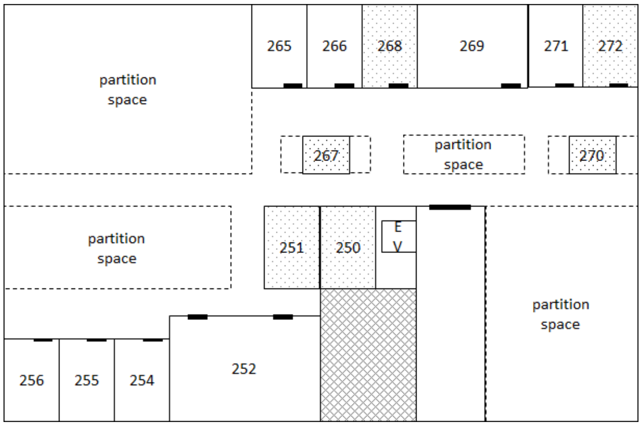

Fig. 1. The floor plan of a target building

increased when using our AR-WFL system compared with the system uses no update method. The result is that when using active learning with uncertainty sampling algorithm for kNN (Nearest Neighbor), the average location estimation accuracy is increased by 1.83%p.

The rest of this paper consists of as follows. In section 2, we describe our dataset, kNN classifier to be used in performance evaluation, and preliminary experimental result about aging problem. The proposed AR-WFL system and selective sampling algorithms are describes in section 3 and 4, respectively. Section 5 shows the result of performance evaluation about the AR-WFL system. Related works are described in section 6 and our conclusion is drawn in Section 7.

II. Preliminaries

A. Dataset

Wi-Fi fingerprint data was collected in a building as depicted in figure 1. We developed an android application that uses the getScanResult method in WiFiManager class. The device we used for data collection is Samsung Galaxy S4.

The office has 14 conference rooms but only 8 rooms are accessible. Room 254, 255, 256, 265, 266, 271 are about 3x5 meters, room 269 is about 6x5 meters, and room 252 is about 8x6 meters. The period of data collection is from August 29th to October 28th in 2016 and we collected the data two or three times a week. Thus total number of collected days is 19 (DS0829~DS1028). To cover the multiple reference points in a room, data was collected in a center, along the wall, and some arbitrary points in a room. The average number of samples is 405 per each day and 53 per each room.

B. kNN classifier

The kNN classifier has been widely used for WFL systems since it utilized in [2]. It is a simple classifier but the performance is comparable, especially when the number training samples is not enough for statistic modeling. The kNN classifier can be expressed like following equation.

$$y_p = \underset{y \in Y}{argmax} \sum_{x_i \in \phi(x)} I(y_i = y) \quad (1)$$

where $\phi(x)$ is a set of k neighbors closest to x. It selects k closest samples called nearest neighbors and classify the test sample x into a class y_p that is majority in $\phi(x)$. Minkovski or mahalanobis distance metric is generally used to figure out the nearest neighbors. We used kNN classifier with euclidean distance for the proposed AR-WFL system.

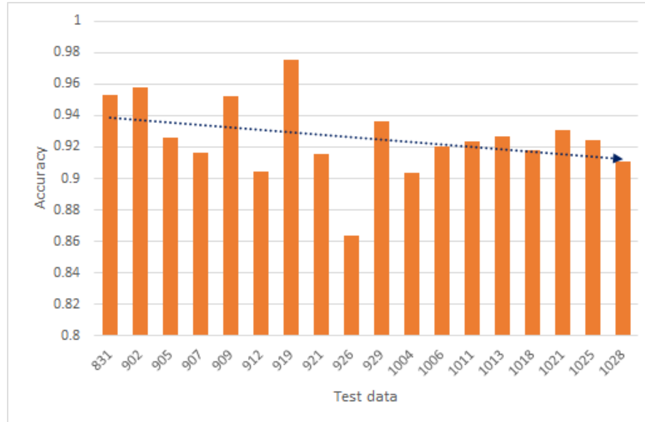

Fig. 2. Change of positioning accuracy over time. DS0829 is used for training dataset.

C. Aging of Wi-Fi fingerprint database

Most of WFL systems are supervised learning-based system. Thus, Wi-Fi fingerprint (WF) database should be constructed before system deployment as a training dataset. As expanding Wi-Fi infrastructure, private APs or Wi-Fi hotspots on smartphone can be included in Wi-Fi fingerprint. Additionally, arrangement of furniture of target building can also be changed. WF database cannot accept these environmental changes without a database update mechanism. In our preliminary experiment, we observed that accuracy of location estimation is decreased over time like figure 2. We define this phenomenon as the aging of WF database.

III. System Architecture

To resolve the aging problem of WF database mentioned in section 2, the database must be updated periodically. As with the initial database construction, labeled Wi-Fi fingerprints are needed to update WF database. However, updating WF database leads to an increase in labor cost due to a dedicated annotator. To address the issue, we designed crowdsourcing-based AR-WFL system using continuous active learning.

Continuous active learning is defined as periodically updating the WF database according to the active learning scheme [4]. Active learning is a method to query a label to a human by selective sampling algorithm. Selective sampling algorithm which is core of active learning selects the target samples that will be queried to a human annotator.

We designed AR-WFL system based on continuous active learning which consists of offline phase, online phase, and update phase as depicted in figure 3. Both online phase and offline phase is similar to traditional WFL system except for storing collection time of Wi-Fi fingerprint t. Update phase consists of three components, Selective Sampling Pool (SS Pool), Selective Sampling Manager (SSM), and Active Query Manager (AQM).

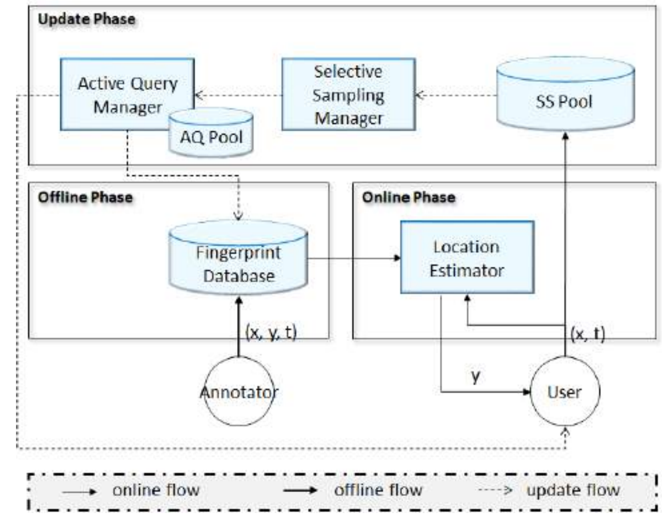

Fig. 3. Proposed aging-resistant WFL system architecture.

SS Pool stores Wi-Fi fingerprints collected from users via crowdsourcing. The Wi-Fi fingerprints collected from the user are unlabeled data and recorded together with time *t*.

SSM analyzes the data accumulated in the SS Pool during the period τ and samples the data to request labeling to the users. The period τ can be changed according to the service environment and it is set as 1 day in this paper. The SSM selects only parts of the data in the SS Pool and passes them to the AQM. The AQM eventually requests labeling for those selected data to the users. Selective sampling strategy of SSM is very important because it affects both user convenience and update performance. In this paper, we propose an uncertainty sampling algorithm for kNN. We will describe the proposed algorithm in detail in section 4.

AQM queries the labels of Wi-Fi fingerprints selected by the SSM to the users. However, AQM cannot get accurate location labels because Wi-Fi fingerprints are difficult for human to see and understand their meaning like figure 4. To address this issue, AQM stores the selected Wi-Fi fingerprints in the AQ Pool and monitors the Wi-Fi fingerprints queried from users. When a queried Wi-Fi fingerprint of a user is highly similar to one of the Wi-Fi fingerprints in the AQ Pool, the AQM requests labeling for that fingerprint to the user. If the user immediately responds to his or her location as a label, AQM can get an accurate label and update the Wi-Fi fingerprint database.

The overall flow of the update phase can be represented by the Continuous Active Learning (CAL) algorithm. The inputs of the CAL algorithm are the unlabeled Wi-Fi fingerprint data $\widehat{X_i}$ (i>=2) accumulated in the SS Pool, the model θ_1 learned through the Wi-Fi fingerprint database X_1 constructed in the offline phase. At first, CAL algorithm selects $\widehat{X_i^*} (\subset \widehat{X_i})$ through selective sampling algorithm. Second, AQM makes active queries to the users to get labels for $\widehat{X_i^*}$. After getting labels for $\widehat{X_i^*}$, labeled Wi-Fi fingerprints X_i^* is created. Finally, the system can add the labeled fingerprint data to an existing database and create a re-learned model θ_i based on it. Through the CAL algorithm, the Wi-Fi environment changed every period τ can be reflected to the database.

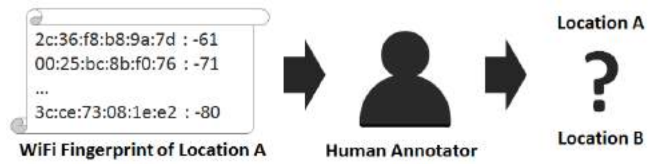

Fig. 4. A problem of labeling Wi-Fi fingerprint by a human annotator

[Continuous Active Learning]
Input : $\widehat{X_i}, \theta_1$
Output : θ_i
Algorithm :
1. $\widehat{X_i^*} = Selective\ Sampling(\widehat{X_i})$
2. $\widehat{y_i} = Active\ Query(\widehat{X_i^*})$
3. $X_i^* = (\widehat{X_i^*}, \widehat{y_i})$
4. $\theta_i = L(X_1 (\cup_{j=2}^{i} X_j^*))$
5. Return θ_i

On the other hand, the update phase may consist of a Naïve Incremental Learning (NIL) algorithm. The NIL algorithm first classifies unlabeled Wi-Fi fingerprint data $\widehat{X_i} (i \geq 2)$ based on the current model θ_{i-1} and obtains corresponding labels $\widehat{y_i}$. The process of updating an existing database using labeled fingerprints is the same as the CAL. Although the NIL algorithm has the advantage of updating the database without the users' help, it is difficult to expect higher performance than the CAL algorithm because it depends on classification result based on WF database. As a result, the labels of NIL is likely to bound to WF database and NIL would only adapt minor change of environments. Performance evaluation for NIL is described in section 5.

[Naïve Incremental Learning]
Input : $\widehat{X_i}, \theta_1$
Output : θ_i
Algorithm :
1. $\widehat{y_i} = C_{\theta_{i-1}}(\widehat{X_i})$
2. $X_i = (\widehat{X_i}, \widehat{y_i})$
3. $\theta_i = L(X_1 (\cup_{j=2}^{i} X_j))$
4. Return θ_i

IV. Proposed Selective Sampling Algorithm

The selective sampling algorithm has been used mainly in image recognition, image retrieval, and natural language processing [6-8]. However, it is the first time to apply the selective sampling algorithm to indoor localization systems based on kNN classifier to the best of our knowledge.

In the proposed AR-WFL system, the core operation of the SSM component is processed through the selective sampling algorithm. The output of selective sampling algorithm is used to query the labels to the users. To minimize the labeling cost, the number of queries must be minimized, but for aging resistant system, the more labeled data collected in recent environments could lead to the better the performance. Therefore, we should find a selective sampling algorithm that can achieve maximal performance with lowest number of queries.

In this section, we describe uncertainty sampling for kNN algorithm newly designed for kNN classifier. This is extended version of previous uncertainty sampling for nearest neighbor (NN) referred to [8].

Uncertainty Sampling for NN (US_NN) algorithm calculates the uncertainty of samples and preferentially selects samples with high uncertainty. Uncertainty can be calculated based on least confident, margin, or entropy metric [5]. However, these uncertainty metrics are difficult to apply to the nearest neighbor (NN) classifier, which is a deterministic classifier since they are based on stochastic modeling techniques. The uncertainty metric for NN classifier has been defined in [7] as followings.

$$x_H^* = \underset{x}{argmax}\left[-\sum_{0 \le i < |L|} P_\theta(y_i|x) log P_\theta(y_i|x)\right],$$

$$P(y|x) = \frac{a_y}{\sum_{0 \le i < |Y|} d_{y_i}}, \quad (2)$$

$$d_y = \min_{s \in S}(\delta(s,x)).$$

where Y is a set of all classes, δ is a distance algorithm, and S is a set of all samples of class y. The function (2) is an entropy-based uncertainty function and does not use Bayesian probabilistic likelihood, but use distance-based likelihood. Likelihood $P(y|x)$ in function (2) means ratio of distance between a sample x and corresponding nearest neighbor in class y to sum of distances $d_{y_i}(0 \le i < |Y|)$. The d_{y_i} is a distance between a sample x and nearest neighbor in class y_i.

Uncertainty Sampling for kNN (US_kNN) algorithm is similar to the US_NN algorithm except that it considers k nearest neighbors. Since the US_NN algorithm considers only closest samples for each class, the effect of outliers can be significant. We designed US_k-NN algorithm to overcome the problem of US_NN algorithm.

The US_k-NN algorithm is basically similar to the function (2). The difference is that US_kNN algorithm considers all class candidates that can be classified as class y in kNN classifier to calculate the distance d_y. The kNN classifier classifies the sample x as a class y, if the majority of the labels of k closest samples with x are y. The function $I(Y, k, y)$ in (3) outputs a set of all class candidates, Y_C.

$$I(Y, k, y) = \{(i_0, \ldots, i_{k-1}) \mid i_{0,\ldots,\lceil \frac{k}{2} \rceil - 1} = y,\ i_{\lceil \frac{k}{2} \rceil, \ldots, k-1} = I_H(Y, \overline{k})\} \quad (3)$$

where Y is a set of all classes, $\bar{k} = 1 - \lceil \frac{k}{2} \rceil$, and $I_H(Y, \bar{k})$ is an extractor that outputs ${}_{|Y|}H_{\bar{k}}$ combinations with repetition from a set Y-$\{y\}$.

Now, we can calculate d_y through the Distance of kNN Candidates (DKC) algorithm including function (3). The d_y is the average of d_{KC} which is the average of k distances $\{d_{KC(y_0)}, \ldots, d_{KC(y_{k-1})}\}$. For each distance $d_{KC(y_i)}$, it represents the distance between the sample x and the closest sample x' with one of the label y_i in Y_C. Because one or more labels in Y_C are the same, we need the function (4) that is able to compute the distance between the sample x and the α-th closest sample with label y.

$$kMinDist(x, y, \alpha) = \min_{\alpha, s \in S}(\delta(s, x)) \quad (4)$$

where δ is a distance algorithm, S is a set of all samples in a class y, and $min_\alpha(\cdot)$ is the function that outputs α-th minimum value.

After getting d_{KC} with the number of $|Y_C|$, we can get d_y by averaging out the all d_{KC} for kNN candidates.

[DKC algorithm]

```
Input : (x, y, k, Y)
Output : d_y
Algo:
Y_c ← I(Y, k, y)
sum_y = 0
For {y_0,...,y_k-1} in Y_C:
   For i = 0 to k-1:
      C(y_i) = 0   //initialization
   sum_KC = 0
   For y_i in {y_0,...,y_k-1} :
      sum_KC = sum_KC + kMinDist(x, y_i, C(y_i))
      C(y_i) = C(y_i) + 1
   d_KC = sum_KC / k
   sum_y = sum_y + d_KC
d_y = d_KC /|Y_C|
return d_y
```

V. EMPIRICAL EVALUATION

In this section, we evaluate the performance of the proposed AR-WFL system using the dataset described in section 2. Two update algorithms are evaluated in detail; Naive Incremental Learning (NIL) and Active Learning with uncertainty sampling for kNN (AL_US_kNN). We used the first dataset, DS0829, as a Wi-Fi fingerprint database and others are used for test datasets. Classification algorithm is kNN and scikit-learn library in python is used for evaluation. We evaluate the performance by comparing accuracy of room-level classification.

$$accuracy = \frac{U_{r \in R} correct(S_r)}{|S|} \quad (5)$$

where S is all test samples, S_r is a subset of S whose samples are labeled with a room r, and $correct(S_r)$ is a subset of S_r whose samples are predicted as a room r.

The proposed continuous active learning system is a crowdsourcing based system and aims to improve the performance while minimizing the number of active queries. Thus, we checked the change of performance varying the number of samples selected through the selective sampling algorithms. The number of samples selected from the selective sampling algorithm is varied from 10 to 200 in units of 10. We limited the number of samples to 200 because the minimum number of samples in our collected datasets is 200.

Figure 5 shows the positioning accuracy according to the number of samples of active query for AL_US_kNN algorithm. The y-axis of the chart represents the average accuracy for the all test data from DS0831 to DS1028. There are 3 highest accuracies in the chart; (60, 94.38%), (170, 94.41%), and (200,

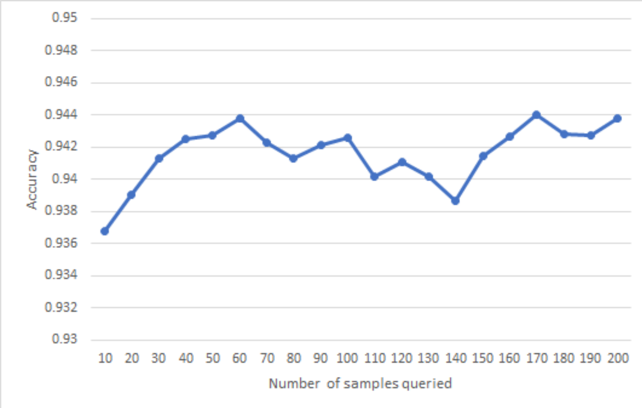

Fig. 5. Positioning accuracy of AL_US_kNN according to the number of active queries.

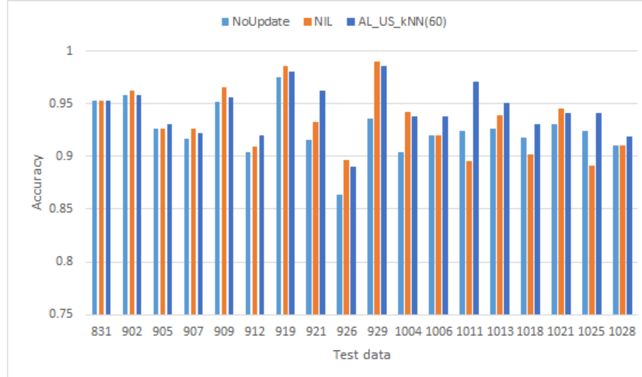

Fig. 6. Comparison of positioning accuracy for each test data with *NoUpdate*, *NIL*, and *AL_US_kNN*. The number of samples actively queried of *AL_US_kNN* is 60.

94.38%). Although maximum accuracy is when the number of samples queried is 170, we must consider the usability which means user intervention. The lower queries, the better usability so it would be more reasonable that we choose the number of queries as 60.

For performance comparison, we also evaluated the NIL and the case that update module is not be applied called NoUpdate. As shown in figure 6, for NoUpdate, the maximum accuracy is 97.53% for DS0919 and the average accuracy is 92.55%. For NIL, the maximum accuracy is 99.04% for DS0929 and the average accuracy is 93.30%. For AL_US_kNN, the maximum accuracy is 98.53% for DS0929 and the average accuracy is 94.38%.

NIL, the easiest way to think about database updates, is a fully automated method that does not require user participation. When the NIL technique was applied, the average accuracy increased by 0.75%p compared to NoUpdate. This shows that the NIL method can mitigate the performance degradation due to aging of the Wi-Fi fingerprint database.

Active learning method, AL_US_kNN, achieved the higher average accuracy than NIL. When the AL_US_kNN technique was applied, the average accuracy increased by 1.83%p compared to NoUpdate when the number of samples queried 60. It demonstrates that crowdsourcing based AR-WFL system using selective sampling algorithm can cope with the aging problem more efficiently in terms of accuracy than NIL method.

VI. Related Works

Wi-Fi fingerprint-based indoor localization technology has been continuously studied for about 20 years, starting with RADAR [2]. RADAR is the first supervised learning based WFL system that collects signal strength from 3 base stations, builds WF database, and estimates the location with kNN classifier.

Since RADAR, various WFL systems have been studied. According to [3], the study of WFL systems can be divided into two categories, one for high accuracy and the other for efficient deployment. For higher accuracy, [9] analyzed the wireless channel characteristics with temporal variation and spatial variation, [10-13] uses feature selection methods based on constancy, strength, or coverage of public Wi-Fi APs, and [14] analyzed probability distribution of Wi-Fi fingerprint database called IGDG (Improved Double-peak Gaussian Distribution).

For efficient deployment, researches which are to reduce the site-survey for Wi-Fi fingerprint database [15-19], and keeping WF database up-to-date [20-23] have been conducted. Wi-Fi fingerprint database is constructed and updated based on the data collected by a person who walks around the building. This causes an increase of labor costs and a decrease of efficiency to build and manage the WFL system. Therefore, these studies aim to design an automated system that minimizes human intervention.

For keeping database up-to-date, [20] proposed LeManCoR system based on manifold co-regularization, [21] proposed the WFL system with dead reckoning that can minimize the reference points to be labeled by a human annotator. [22] proposed a crowdsourcing-based system, which assumes that users have a navigation solution called T-PN. T-PN operates based on inertial sensors and provides LLH (latitude, longing, height) information. The system automatically updates the Wi-Fi fingerprint database if it detects the difference between user's current Wi-Fi fingerprint and that of the database although LLH is the same within thresholds. This system has the advantage of enabling database update without user's active participation, but it could be a limitation that system depends on T-PN solution. [23] proposed the automated system to update Wi-Fi fingerprint database by continuously tracking changes in signal strength using wireless sensor network (WSN). This system can automatically collect the labeled Wi-Fi fingerprint data based on WSN which sensors are pre-installed in fixed locations. Thus it has the advantage of instantly reflecting the change of Wi-Fi environment but disadvantage of being required the additional infra, WSN.

The active learning technique used in this paper has been studied in the field of machine learning. According to [4], active learning techniques have been studied for probabilistic classifiers and non-probabilistic classifiers. In particular, [7] proposed a selective sampling algorithm for NN classifier, which is a non-probabilistic classifier.

Active learning techniques are designed to solve the problem that many unlabeled data exist but obtaining labels of them are expensive. Since the WFL system is difficult to acquire labeled data, it is considered that the active learning scheme is applicable to the WFL system. There are no previous works to apply the active learning scheme to the WFL system to the best of our knowledge.

VII. Conclusion

In this paper, we proposed AR-WFL system with continuous active learning to solve the aging problem in WFL system. The AR-WFL system is a crowdsourcing-based system that does not assume a dedicated annotator. We applied selective sampling algorithms to minimize user participation for updating Wi-Fi fingerprint database. For the performance evaluation of the AR-WFL system, 19 datasets were collected for 2 months in a target building with 8 rooms. Based on the collected data, uncertainty sampling for k-NN algorithm were implemented to compare the accuracy of location estimation. As a result, it showed that average accuracy is increase by 1.83%p compared with NoUpdate.

Acknowledgment

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No.2015-0-00168, Development of Universal Authentication Platform Technology with Context-Aware Multi-Factor Authentication and Digital Signature)

References

[1] U.S. Environmental Protection Agency Green Building Workgroup, "Buildings and their impact on the environment: A statistical summary," 2009.

[2] Paramvir Bahl and Venkata N. Padmanabhan, "RADAR: An In-Building RF-based User Location and Tracking System," *Tech. Rep. MSR-TR-00-12*, Microsoft Research, Feb. 2000.

[3] Suining He and S.-H. Gary Chan, "Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 466-490, First quarter 2016.

[4] Burr Settles, "Active Learning Literature Survey," *Computer Sciences Technical Report 1648*, University of Wisconsin-Madison, Jan. 2010.

[5] Bur Settles, "An Analysis of Active Learning Strategies for Sequence Labeling Tasks," *Proceedings of the Conference on Empirical Mehtods in Natural Language Processing*," pp.1070-1079, Oct. 2009.

[6] Prateek Jain and Ashish Kapoor, "Active Learning for Large Multi-class Problems," *IEEE Conference on Computer Vision and Pattern Recognition*, pp.762-769, 2009.

[7] Michael Lindenbaum, Shaul Markovitch, and Dmitry Rusakov, "Selective Sampling for Neareset Neighbor Classifiers," *Machine Learning*, 54, pp.125-152, 2004.

[8] Martina Hasenjäger and Helge Ritter, "Active Learning with Local Models," *Natural Processing Letters*, pp.107-117, 1998.

[9] Moustafa Youssef and Ashok Agrawala, "The Horus WLAN Location Determination System," *MobiSys'05*, pp.205-218, June 2005.

[10] Arsham Farshad, Jiwei Li, Mahesh K. Marina, and Francisco J. Garcia, "A Microscopic Look at WiFi Fingerprinting for Indoor Mobile Phone Localization in Diverse Environments," *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN2013)*, Oct. 2013.

[11] Jungmin So, Joo-Yub Lee, Cheal-Hwan Yoon, and Hyunjae Park, "An Improved Location Estimation Method for Wifi Fingerprint-based Indoor Localization," *International Journal of Software Engineering and Its Applications*, vol. 7, no. 3, pp. 77-86, May 2013.

[12] Pei Jiang, Yunzhou Zhang, Wenyan Fu, Huiyu Liu, and Xiaolin Su, "Indoor Mobile Localization Based on Wi-Fi Fingerprint's Important Access Point," *International Journal of Distributed Sensor Networks*, vol. 2015, pp. 1-8, 2015.

[13] Youngsam Kim and Soohyung Kim, "Rethinking of Feature Selection Methods for Room-Level Localization Using Public APs," *19th International Conference on Advanced Communication Technology (ICACT2017)*, pp. 24-28, Feb. 2017.

[14] Lina Chen, Binghao Li, Kai Zhao, Chris Rizos, and Zhengqi Zheng, "An Improved Algorithm to Generate a Wi-Fi Fingerprint Database for Indoor Positioning," *Sensors2013*, 13, pp. 11085-11096, Aug. 2013.

[15] Anshul Rai, Krishna Kant Chintalapudi, Venkata N. Padmanabhan, and Rijurekha Sen, "Zee: Zero-Effort Crowdsourcing for Indoor Localization," *MobiCom'12*, pp. 293-304, Aug. 2012.

[16] He Wang, Souvik Sen, Ahmed Elgohary, Moustafa Farid, Moustafa Youssef, and Romit Roy Choudhury, "No Need to War-Drive: Unsupervised Indoor Localization," *MobiSys'12*, pp. 197-210, June 2012.

[17] Ahmad Abadleh, Sangyup Han, Soon J. Hyun, Ben Lee, and Myungchul Kim, "ILPS: Indoor Localization using Physical Maps and Smartphones Sensors," *15th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, June 2014.

[18] Chensu Wu, Zheng Yang, Yunhao Liu, and Wei Xi, "WILL: Wireless Indoor Localization without Site Survey," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 4, pp. 839-848, Apr. 2013.

[19] Chensu Wu, Zheng Yang, and Yunhao Liu, "Smartphones Based Crowdsourcing for Indoor Localization," *IEEE Transactions on Mobile Computing*, vol.14, no.2, pp. 444-457 , Feb. 2015.

[20] Sinno Jialin Pan, James T. Kwok, Qiang Yang, and Jeffrey Junfeng Pan, "Adaptive Localization in a Dynamic WiFi Environment Through Multi-view Learning," *AAAI'07*, vol. 2, pp. 1108-1113, July 2007.

[21] Le T. Nguyen and Joy Zhang, "Wi-Fi fingerprinting through Active Learning using Smartphones," *UbiComp'13*, pp.969-976, Sep. 2013.

[22] Yuan Zhuang, Zainab Syed, You Li, and Naser El-Sheimy, "Evaluation of Two WiFi Positioning Systems Based on Autonomous Crowdsourcing of Handheld Devices for Indoor Navigation," *IEEE Transactions on Mobile Computing*, vol. 15, no. 8, pp. 1982-1995, Aug. 2016.

[23] Walter Balzano, Aniello Murano, and Fabio Vitale, "WiFACT – Wireless Fingerprinting Automated Continuous Training," *30th International Conference on Advanced Information Networking and Applications Workshops(WANIA)*, pp. 75-80, Mar. 2016.

Youngsam Kim received a B.S. (2009) degree in computer engineering from Chungbuk National University and a M.S. (2011) in information security engineering from the University of Science and Technology in South Korea. In 2011, he joined future internet research team in NIMS as a researcher Currently, he is a researcher at the Electronics and Telecommunications Research Institute. His research interests include context-aware authentication, machine learning, and security protocols.

Soohyung Kim received the B.S. and M.S. degrees in computer science from Yonsei University, Seoul, Korea, in 1996 and 1998. He received the Ph.D. degree in computer science from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea in 2016. He is currently a Director of Information Security Research Division in Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea. His research interests include payment system, biometrics, identity management, network and system security.