Scalable and Consistent Radio Map Management Scheme for Participatory Sensing-based Wi-Fi Fingerprinting

Yungeun Kim, Yohan Chon, Hojung Cha
Yonsei University
Seoul, Korea
{ygkim,yohan,hjcha}@cs.yonsei.ac.kr

MyungIn Ji, Sangjoon Park
Electronics and Telecommunications Research Institute
Daejeon, Korea
{myungin,sangjoon}@etri.re.kr

ABSTRACT
Although Wi-Fi fingerprinting is a promising solution for indoor localization, its widespread use is limited due to the necessity of time-consuming site survey. Recently, active research has been conducted to reduce site-survey cost with participatory sensing. However, most of the previous schemes focused on construction of radio map, but not on radio map management regarding scalability and consistency issues. Radio map construction should handle a large amount of data collected from many people over a long period in limited space. It should also guarantee consistent accuracy regardless of time and user context, such as device type and direction. In this paper, we describe a radio map management scheme, based on fingerprint clustering technique, which considers both the scalability and consistency issues. The proposed scheme clusters the fingerprints collected at different times with different types of devices and picks a cluster head as a representative fingerprint. We validate the feasibility of the proposed scheme with real experiments in an office environment.

Categories and Subject Descriptors
C.3 [**Special-purpose and Application-based Systems**]: Microprocessor/microcomputer applications; H.4.m [**Information Systems Applications**]: Miscellaneous

General Terms
Management.

Keywords
Participatory sensing, Wi-Fi fingerprinting, Clustering fingerprints.

1. INTRODUCTION
Recently, diverse location-based services (LBS) such as mobile marketing, life-logging, and social networking have been actively emerging in the market. Locating mobile users is an essential function for these types of LBS. With widespread use of smartphones, LBS should now cover indoor spaces where people spend a significant amount of time in daily life [1]. However, unlike in outdoor space, localization in indoor space is challenging because GPS is unavailable in indoor environments. Several indoor LBS solutions have been developed using various sensors, such as UWB [2], GSM [3], and Wi-Fi [4]. The Wi-Fi fingerprinting (WF) technique has been extensively used for indoor localization because meter-level accuracy is achieved by exploiting existing access points (APs) without additional infrastructure. However, WF requires the expense of constructing a radio map that contains fingerprints in the target area. In addition, locational accuracy is degraded when Received Signal Strength (RSS) changes according to time, device type, and user direction.

Participatory sensing could be a good solution for these problems, since a radio map is constructed and updated continuously by casual users. However, radio map construction with participatory sensing introduces new challenges: coverage, scalability, and consistency issues. The coverage of a radio map must be rapidly extended, and the radio map must handle a large number of fingerprints collected from casual users over a long period with restricted storage. The radio map should also guarantee consistent accuracy regardless of time and user context. Several WF systems that construct radio maps with participatory sensing have been proposed [5,6], but they focused on coverage issue without considering scalability and consistency problems.

In this paper, we propose a participatory sensing-based radio map management scheme that considers both scalability and consistency. Scalability and consistency of a radio map are in a trade-off relationship; if the radio map contains multiple fingerprints for a single location, consistency increases but scalability decreases. The proposed scheme clusters the fingerprints collected at different times from different users in the same location and picks a cluster head as the representative fingerprint for that location. We validate the feasibility of the proposed scheme with real experiments in an office building.

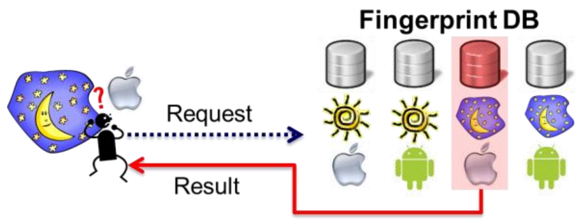
The paper is structured as follows. Section 2 presents related work. Section 3 analyzes the scalability and consistency issues of radio maps in preliminary experiments. Section 4 describes our clustering-based radio map management scheme for resolving scalability and consistency issues. Section 5 presents our evaluation of experiments conducted in an office building. Section 6 presents conclusions and discusses future work.

2. RELATED WORK

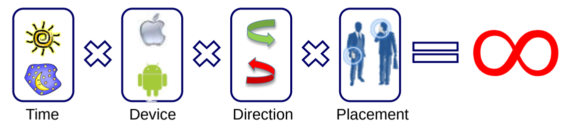
RADAR [4] was the first WF system, and the scheme requires training and localization phases. In the training phase, a radio map is constructed by collecting the fingerprints of signal strengths of APs at each location. Location is then determined by comparing observed signal strengths with the collected fingerprints. The RADAR-like system typically requires a high cost of radio map construction in the training phase.

To reduce training costs, several crowdsourcing-based approaches have been proposed. Park et al. [6] proposed an organic indoor location system to replace the training phase with participatory sensing, in which the radio map was organically constructed by collecting data from individual users. The system used a Voronoi diagram-based method to enlarge the coverage of the radio map efficiently, and it also used a clustering-based method to filter out erroneous user input automatically. Park et al. [7] enhanced the system to overcome the device-diversity issue that causes signal strength to be different according to device type. Kim et al. [5] proposed a radio map construction method that used opportunistic sensing, in which individual smartphone users not only collected signal strength but also estimated their location automatically with Pedestrian Dead Reckoning (PDR). This system collected data without user intervention. The quality of this radio map was, however, lower than those with the participatory sensing-based method because of drift error in PDR. These systems all used the crowdsourcing technique. However, they all focused on increasing coverage of the radio map, whereas our work focused on a scheme of scalable and consistent radio map management.

Other types of approaches have also been proposed to reduce the cost for radio map construction. In RADAR, a model-based approach for radio map construction was proposed to eliminate the training phase [4]. This method simulated the signal strengths of APs at each location by using a log distance path loss model. The model required coordinates of APs and walls to calculate the signal strength accurately. Wang et al. [8] proposed a system that reduced the training cost in multi-floor buildings. The system generated a radio map of each floor incrementally, and it used a characteristic pertinent to signal: the locations of maximum strength of a certain AP overlap in 2D-space regardless of the floor level. Chintalapudi et al. [9] proposed the EZ system, which constructed a radio map with signal strength information from unknown locations. This system calculated locations and other unknown factors by using constraints in the physics of signal propagation.



(a) Consistency issue: user location should be estimated consistently regardless of environmental factors or user contexts.



(b) Scalability issue: radio map should handle a large number of fingerprints collected from many users over a long period in limited space.

Figure 1. Consistency and scalability of radio map constructed by participatory sensing.

3. CONSISTENCY VS. SCALABILITY OF RADIO MAP

The accuracy of a radio map should be consistent regardless of time, user, and environmental changes. Several works [7, 10] have clarified the RSS variance problem, in which RSSs in the same location can be different according to time and environmental changes, user direction, and device type. This RSS variance problem results in degradation in accuracy of Wi-Fi fingerprinting. One solution would be participatory sensing-based radio map construction, in which the radio map includes RSS fingerprints that are collected at different times with different devices, as illustrated in Figure 1(a). However, we should consider the scalability issue when a radio map is constructed with participatory sensing (Figure 1(b)). A radio map should efficiently handle a large number of fingerprints collected from many users over a long period in limited space.

To analyze the relationship between scalability and consistency, we conducted preliminary experiments in an office building. We collected 6,832 scan results at 23 locations with 6 different devices during single day, in the office environment shown in Figure 2. During collection, we also differentiated the direction as facing north or south. We used one-half of the scan results for constructing a radio map and the other half for a localization test. Figure 3 shows the changes in precision of localization and size of the radio map when the number of fingerprints per location in the map increased. In the case of a single fingerprint per

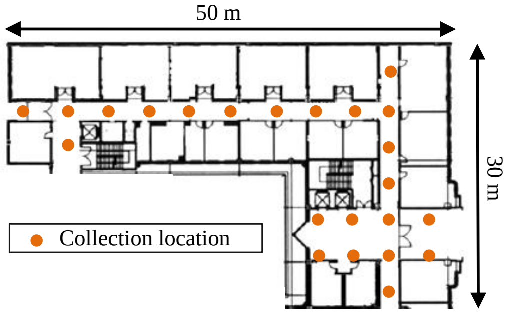


Figure 2. Layout of the experiment site.

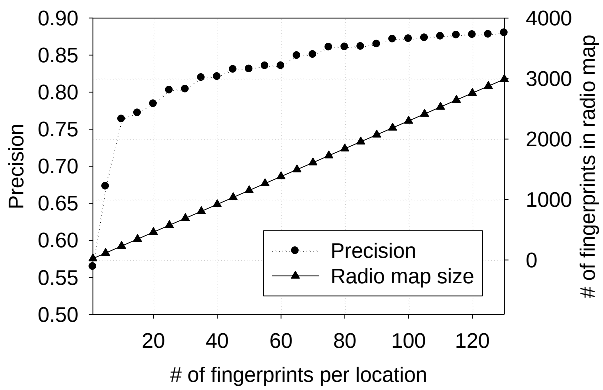


Figure 3. Trade-off relationship between consistency and scalability of radio map: more fingerprints are required to achieve higher precision.

location, the number of fingerprints in the radio map was 23 and the precision was only 0.54. With 130 fingerprints per location, the number of fingerprints was 2,990 and the precision was 0.88. Precision increased with the number of fingerprints, meaning that the consistency of the radio map improved when it stored more fingerprints per location. The size of radio map increased linearly with the number of fingerprints per location. Overall, the experiment showed that scalability and consistency of a radio map was a trade-off relationship. This led us to develop a management scheme that considered both scalability and consistency issues.

4. CLUSTERING-BASED RADIO MAP MANAGEMENT

As shown in Figure 3, improvements in precision slow down when the number of fingerprints per location exceeds 10. This means that a radio map constructed with participatory sensing produces many redundant fingerprints that have similar signal characteristics. In order to solve consistency and scalability issues simultaneously, a radio map should store only representative fingerprints. This means that redundant fingerprints must first be identified. This can be done from user contexts such as device type and user direction, but privacy issues arise. Moreover, the fingerprints collected from different user contexts are not always guaranteed to have different signal characteristics. Alternatively, clustering can be a good solution for finding similar fingerprints and choosing an appropriate representative.

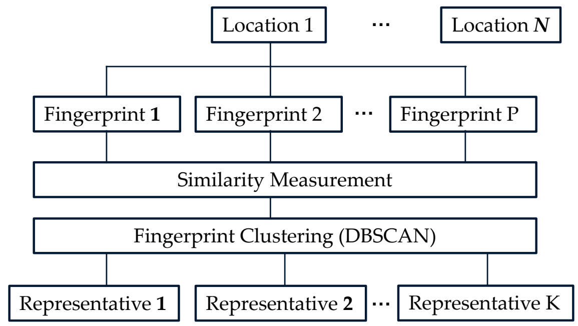


Figure 4. The clustering-based radio map management scheme.

We propose the following clustering-based radio map management scheme. Figure 4 illustrates the overall process. For each location, there are multiple fingerprints $\{F_1, F_2, F_3, \dots, F_P\}$ collected from different users over a period. F_i in Equation (1) contains the signal strengths from APs at location i.

$$F_i = \{rss^1, rss^2, rss^3, \dots, rss^M\} \quad (1)$$

where M represents the number of APs in the target area. To cluster fingerprints, similarity among them should be calculated. Equation (2) shows that similarity table T_s is built by comparing fingerprints with each other.

$$T_s = \begin{bmatrix} S_{(1,1)} & S_{(2,1)} & \dots & S_{(P-1,1)} & S_{(P,1)} \\ S_{(1,2)} & S_{(2,2)} & \dots & S_{(P-12)} & S_{(P,2)} \\ \dots & \dots & \dots & \dots & \dots \\ S_{(1,P-1)} & S_{(2,P-1)} & \dots & S_{(P-1,P-1)} & S_{(P,P-1)} \\ S_{(1,P)} & S_{(2,P)} & \dots & S_{(P-1,P)} & S_{(P,P)} \end{bmatrix} \quad (2)$$

Here, $S_{(a,b)}$ represents the similarity between F_a and F_b calculated with the Tanimoto coefficient, which yields the Jaccard coefficient of given vectors. The clustering algorithm is then performed based on the similarity table. We used DBSCAN [11], which is one of the density-based clustering algorithms. DBSCAN does not dictate the number of clusters and filters noisy objects. Note that other clustering algorithms can be used for the scheme, but algorithm choice is out of scope of this study. DBSCAN requires two parameters for clustering: ε and $MinPts$. ε represents the density-reachable distance between two points, and $MinPts$ represents the minimum number of points required to form a cluster. Since we estimate not the distance but the similarity between two fingerprints, we define that two fingerprints, F_a and F_b, are density-reachable in the case of $S_{(a,b)} > \varepsilon$.

After clustering is complete, fingerprints are divided into K groups, and each group is represented as $\{F_1, F_2, F_3, \dots, F_{P_k}\}$ where P_k is the number of fingerprints in group k. Then, a representative fingerprint for each group is selected with Equation (3).

$$RF_i^{\,k} = \arg\max_{F_i} \sum_{j=1}^{P_k} S_{(i,j)} \tag{3}$$

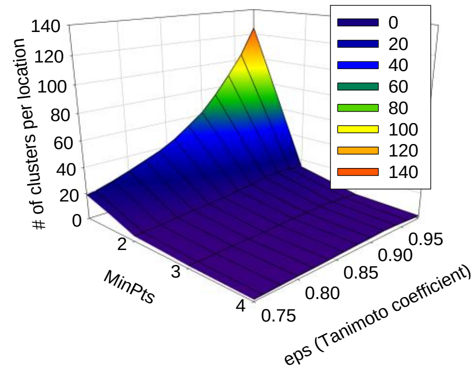
The fingerprint that has the largest sum of similarity with other fingerprints in the group is selected as the representative. After the representative fingerprints are selected, a radio map is constructed as $\{LF_1, LF_2, \dots, LF_N\}$. LF_i contains the coordinate information and representative fingerprints of location i as shown in Equation (4).

$$LF_i = \{x, y, RF_i^{\,1}, RF_i^{\,2}, \dots, RF_i^{\,K_i}\} \tag{4}$$

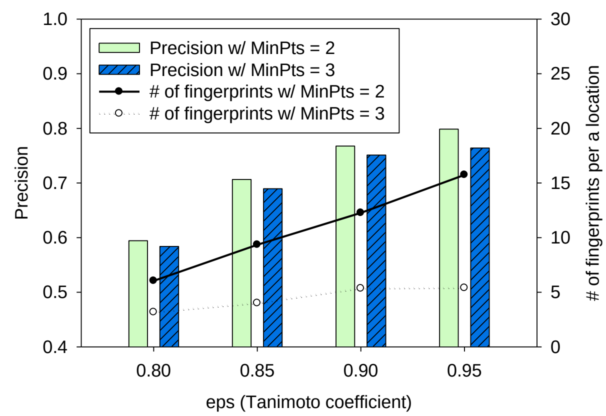
Fingerprint clustering reduces the size of the radio map during map update with participatory sensing. However, it does not guarantee scalability completely, since the size of a radio map gradually increases when many update processes occur over a long period. To overcome this problem, we removed meaningless fingerprints from the radio map during localization. These fingerprints can be removed by finding the Least Recently Used (LRU) fingerprints, which is motivated from conventional cache algorithms.

5. EVALUATION

We evaluated performance of the proposed scheme with the data used in Section 3. In the experiment, the radio map was constructed by the clustering-based method using the training data. We first investigated the effect of two parameters of DBSCAN. Figure 5 (a) shows that the performance of DBSCAN changes depending on these parameters. Clustering with $MinPts$ larger than 4 was not meaningful, since it rarely creates clusters due to the high threshold. As ε increases, DBSCAN creates more clusters. Based on this result, we created radio maps with eight different settings: $MinPts$ = {2, 3}, and ε = {0.8, 0.85, 0.9, 0.95}. Then, we evaluated the performance of each setting by performing localization with the test data. Figure 5(b) shows the number of clusters and the performance of radio maps created with different settings. ε smaller than 0.85 produces fewer clusters per location, and the precision is lower than other settings. This is because low ε tends to form a group with fingerprints that have different signal characteristics. In the case of $MinPts$ = 2, precision is not much higher, whereas the number of fingerprints per location is much greater than that with $MinPts$ = 3. Consequently, DBSCAN with $MinPts$ = 3, and ε = 0.95 shows the best consistency and scalability in the experiment. Optimal parameters could be different depending on the environment. However, the optimal parameters can be found by performing 2-fold cross-validation with data collected by participatory sensing.



(a) Clustering performance.



(b) Localization performance.

Figure 5. Effect of parameters on DBSCAN.

To investigate improvement provided by the proposed scheme, we compared it with other schemes that store a fixed number of fingerprints per location. Figure 6 shows that precision increased from 0.54 to 0.79 as the number of fingerprints per location increased from 1 to 20. However, the size of the radio map also increased, from 23 fingerprints to 460. The proposed scheme achieved precision of 0.76 with 124 fingerprints. Compared to the results from using 5 fingerprints per location, precision improved 17%, while the size of the radio map increased by only 7% with the proposed scheme. Results showed that the proposed clustering scheme is feasible for efficient radio map management with participatory sensing.

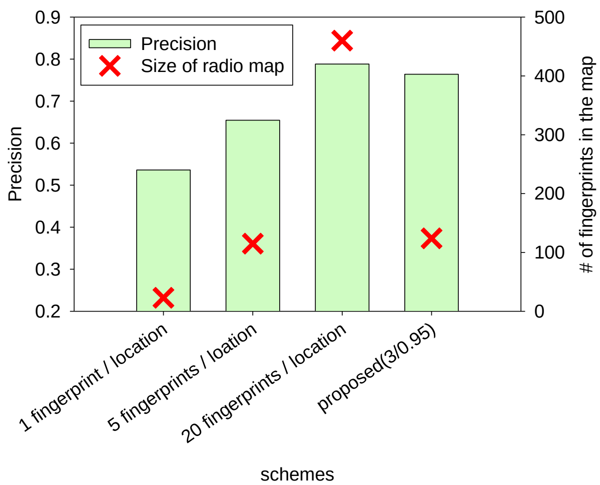


Figure 6. Performance improvement.

6. CONCLUSION

In this paper, we focused on the scalability and consistency issues that should be considered for radio map construction with participatory sensing. With our experiment, we clarified that scalability and consistency are in a trade-off relationship. We also proposed a clustering-based radio map management scheme that solves consistency and scalability issues simultaneously. The scheme uses a clustering algorithm to filter redundant fingerprints and keep only representative ones. We validated that the proposed scheme achieves precision of 0.76 with an average of 5.4 fingerprints per location in a scenario in which data is collected with 6 devices and 2 directions. We expect that the proposed scheme will realize greater improvement when collected fingerprints have more dynamic signal characteristics. As part of our future work, we plan to validate the proposed system with large-scale experiments in which participatory sensing is performed by a large number of people over a long period, using social network applications such as FourSquare and LifeMap [1].

7. ACKNOWLEDGMENTS

This research was supported by the “Development of 5m accuracy LBS Platform supporting AR service” project funded by the Korea Communications Commission

8. REFERENCES

[1] Chon, Y., and Cha, H. 2011. LifeMap: Smartphone-Based Context Provider for Location-Based Services. IEEE Pervasive Computing, vol. 10, no.2, pp. 58–67.

[2] Zhang, C., Kuhn, M., Merkl, B., Fathy, A. E., and Mahfouz, M. 2006. Accurate UWB indoor localization system utilizing time difference of arrival approach. IEEE Radio Wireless Symp., pp. 515.

[3] Otsason, V., Varshavsky, A., LaMarca, A., and de Lara, E. 2005. Accurate GSM indoor localization. UbiComp 2005, Lecture Notes, Computer Science, Springer-Varlag, vol. 3660, pp.141.

[4] Bahl, P., and Padmanabhan, V. N. 2000. RADAR: An In-building RF-based User Location and Tracking System. In Proc., 19th IEEE Int. Comput. Commun. (InfoCom), vol. 2, pp. 775–784.

[5] Kim, Y., Chon, Y., and Cha, H. 2012. Smartphone-Based Collaborative and Autonomous Radio Fingerprinting,” IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 42, no. 1, pp. 112–122.

[6] Park, J., Charrow, B., Curtis, D., Battat, J., Minkov, E., Hicks, J., Teller, S., and Ledlie, J. 2010. Growing an Organic Indoor Location System. In Proc. of the 8th Int. Conference on Mobile Systems, Applications, and Services, pp. 271–284.

[7] Park, J., Curtis, D., Teller, S., Ledlie, J. et al., 2011. Implications of Device Diversity for Organic Localization. In Proc. 30th IEEE Int. Comp. Commun. (InfoCom), Shanghai, China, pp. 3182–3190.

[8] Wang, H., Zheng, V., Zhao, J., and Yang, Q. 2010. Indoor localization in multi-floor environments with reduced effort. In 8th IEEE International Conference on Pervasive Computing and Communications (PerCom 2010), pp. 244–252, Mar. 29–Apr. 2.

[9] Chintalapudi, K., Padmanabha Iyer, A., and Padmanabhan, V. N. 2010. Indoor localization without the pain. In Proceedings of the ACM MobiCom, New York, NY, USA, pp. 173–184.

[10] Kim, Y., Shin, H., and Cha, H. 2012. Smartphone-based Wi-fi pedestrian-tracking system tolerating the RSS variance problem. In 10th IEEE International Conference on Pervasive Computing and Communications (PerCom 2012), pp. 11–19, Mar. 19–23.

[11] Ester, M., Kriegel, H.-E, Sander, J., Xu, X. 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining, Portland, OR, USA, AAAI Press, pp. 226–231.