Detecting Aged Deterioration of a Radio Base Station Map for Wi-Fi Positioning

Makiko Kawanaka
College of Information Science and Engineering, Ritsumeikan University
Shiga, Japan
maki@ubi.cs.ritsumei.ac.jp

Kazuya Murao
College of Information Science and Engineering, Ritsumeikan University
Shiga, Japan
murao@cs.ritsumei.ac.jp

Kohei Yamamoto
College of Information Science and Engineering, Ritsumeikan University
Shiga, Japan
moi@ubi.cs.ritsumei.ac.jp

Masahiro Mochizuki
Research Organization of Science and Technology, Ritsumeikan University
Shiga, Japan
moma@ubi.cs.ritsumei.ac.jp

Kota Tsubouchi
Yahoo Japan Corporation
Tokyo, Japan
ktsubouc@yahoo-corp.jp

Nobuhiko Nishio
College of Information Science and Engineering, Ritsumeikan University
Shiga, Japan
nishio@cs.ritsumei.ac.jp

ACM.
UbiComp/ISWC'17 Adjunct ,, September 11–15, 2017, Maui, HI, USA
ACM 978-1-4503-5190-4/17/09.
https://doi.org/10.1145/3123024.3124401

Abstract

Currently GPS is utilized for localization but is not reliable indoors or underground where its signals cannot reach. Instead, localization using Wi-Fi is investigated in the field of indoor positioning. Wi-Fi localization is performed using a radio model created by observed information collected beforehand; however, the model deteriorates over-time as Radio Base Stations (RBS) appear, disappear, and are relocated. As a result, localization accuracy decreases. Accuracy of positioning is affected by RBS and the number of RBS used for localization, thus, updating the model is required whenever deterioration occurs. There is no way to detect deterioration and necessity to update. This research investigates a method to detect such deterioration only by analyzing logs collected from navigation applications. Appearance and disappearance of RBS can be detected by analyzing observed date with its interval related to its average and standard deviation. Relocation of RBS can be detected by finding change-point of RBS's co-occurrence with other RBS. As results, it is demonstrated that RBS which possibly appears and/or disappears can be specified, if observed over certain days. It is also demonstrated that RBS which possibly be relocated can be specified by finding the best threshold.

Author Keywords
Indoor localization; Wi-Fi; Wi-Fi positioning; Aged deterioration

ACM Classification Keywords
C.2.m [COMPUTER-COMMUNICATION NETWORKS]: Miscellaneous

Introduction
Recently, mobile devices are widely used, which have a wide variety of sensors including accelerometers, a GPS receiver, and Wi-Fi and Bluetooth radio wave receiver. Along with this trend, location-based service provided with collected logs from mobile devices has been paid much attentions[*1*2*3]. GoogleNow provides news and timetable depending on user's location and tendency of searching Internet. GPS is currently used to localize users but it is not reliable indoors and underground where signals of GPS cannot be reached. Thus, many research were conducted to enable indoor localization using an accelerometer, gyroscope sensor, Wi-Fi strength, and Bluetooth strength[3][5][4][1]. Localization using Wi-Fi radio wave is attracting attention in the field of indoor localization from the view of its diffusion as infrastructures.
Wi-Fi localization can be performed using a model created from observed Wi-Fi strength collected beforehand at every observation point; however, the repeatability of the model deteriorates over time caused by appearance, disappearance, and relocation of RBS then it finally degrades the accuracy of localization. Generally, the more RBS used for localization, the better accuracy is. Thus, updating the model is required whenever deterioration occurs. Updating the model is high in cost since it is conducted manually. For now, there is no way to notice deterioration and when to update.
Previous methods did not take this deterioration into consideration when localization is conducted. Previous methods selected the RBS heuristically, for instance, they selected the RBS just looking at its Extended Service Set Identifier (ESSID) which appears to be highly reliable. In this research, we propose a method to detect deterioration of the model by analyzing logs collected from navigation application so that we can select the most appropriate RBS which should be used for localization. Following this method, RBS which was not used in a conventional method can be used and the accuracy of localization can be improved. By detecting appearance, disappearance, and relocation of RBS, over-time deterioration in the model created beforehand can be detected.

Related Work
Localization using Gaussian Mixture Model
Wi-Fi localization requires a model to be created beforehand. There are a variety of modeling methods such as fingerprint localization, weighted average localization, and Gaussian Mixture Model (GMM) localization. GMM is especially recognized as a model that requires less amount of data and computation. In order to improve the accuracy of localization, a very large scale database is needed. However, the amount of data can be shrunk when the database is expressed by GMM[2]. GMM is originally linear combination of multiple Gaussian distributions. Two dimensional GMM expresses every RBS. Each Gaussian distribution has average μ, variance-covariance matrix Σ, and mixing coefficient π_k as parameters. These parameters can be derived by Expectation-Maximization algorithm (EM algorithm). Then GMM can be created for localization by mod-

[1] <Foursquare https://ja.foursquare.com/>
(May. 24, 2017 last accessed)
[2] <GoogleNow https://www.google.com/intl/ja/landing/now>
(May. 24, 2017 last accessed)
[3] <Swarm https://ja.swarmapp.com/> (May. 24, 2017 last accessed)

Figure 1: An example of GMM of a RBS.

eling each RBS as shown in Figure 1. The map cannot be created for mobile hotspots since they just exist temporarily. This helps to eliminate such RBS which decreases the accuracy of localization.

Proposed Method

Expected environment

This research is expected to work on logs collected from navigation application that many people are using. We assume that the following RBS will be detected;

- Mobile hotspots
- Relocation of RBS
- Appearance of RBS
- Disappearance of RBS
- Intermittently-operated RBS

Mobile hotspots are RBS that people carry around including tethering of cell phones. When the person moves the RBS moves together. Therefore this kind of RBS decrease the accuracy of localization. They can be eliminated by analyzing the observed data over time and finding their movement. In addition, in the process of modeling GMM[2], they can be eliminated so this research does not target to detect mobile hotspots. Relocation of RBS other than hotspots also occurs, however, using the old model (e.g. GMM) decreases the accuracy. Thus, this research targets to detect this type of event. There can be disappearance of RBS. It causes deterioration too, thus, this research targets to detect this event. In addition, there can be appearance of RBS. Creating a RBS model with newly appeared RBS can improve the accuracy of localization, thus, this research targets to detect this type of event. Intermittently-operated RBS are the RBS which intermittently work depending on time and other conditions. This sort of RBS are observed on and off so that they are not suitable for localization. Therefore, this research aims to detect the appearance, disappearance, and relocation of RBS and identify the timing to update and/or remove the radio model of RBS.

Definition of RBS as observation targets

One scan output data is shown in Table 1, which are observed Basic Service Set Identifier (BSSID), ESSID, Received Signal Strength Indicator (RSSI), and the time when a scan is conducted. However, it is unknown where each data is scanned. According to Wi-Fi radio wave characteristics, RSSI showing more than -75[dBm] are in high proportion to its distance from the RBS while RSSI less than

RBS	BSSID	ESSID	RSSI
A	00:09:b4:70:1d:c7	00PASSPORT	-55
B	12:09:b4:70:15:f6	FREE_Wi-Fi	-80
C	b4:c7:99:16:07:34	Secure_Wi-Fi	-40

Table 1: Information of RBS captured for one scan.

showing -75[dBm] is in low proportion. Thus, this research targets the RBS showing RSSI more than -75[dBm].

Elimination of RBS showing local address
BSSID is changeable by users if it is a local address. Therefore, two different BSSID may be observed from a single RBS. Each BSSID consists of 48 bits and the 7th bit in the 1st octet is called G/L (Global/Local) bit as shown in Figure 2. If the G/L bit is 1, it is known that the BSSID is used as a local address in closed network. Thus, this research does not target such BSSID showing the local address.

How to address two BSSID which has the same G/L bit from 1.1st to 5.5th octet
The sequence of bits from the 1.1st to 5.5th octet is vendor ' s and model code. Therefore, if different BSSID show the same value from the 1.1st to 5.5th octet, they can be handled as the same BSSID obtained from the same RBS. We actually tried plotting the location of such BSSID predicted by GMM. As shown in Figure 3, it was observed that three sets of two different BSSID but with the same G/L bit from the 1.1st to 5.5th octet were plotted almost in the same area.

Detection of appearance and disappearance of RBS
In order to detect the appearance and disappearance of RBS over certain period of time, we analyze observed dates and its interval regarding average and standard deviation by comparing the data with the non-observed period

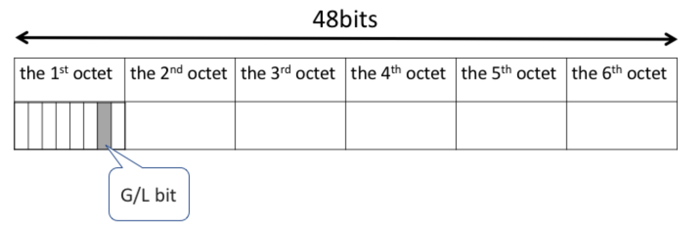

Figure 2: G/L bit.

as shown in Figure 4. The first and the last observed date of each RBS during the objected term are specified. The RBS is observed on-and-off during the period. For each RBS, the average μ and standard deviation σ of the days between each observation within the term are calculated. Thus, around 70% of RBS are supposed to be observed within before and after $\mu + \sigma$ days from the last observed date. Therefore it can be judged that RBS disappears if the number of days between the last observed date and the end date of objected term is more than $\mu + \sigma$ days. In the same way, it can be judged that RBS appears if the number of days between the first observed date and the beginning date of objected term is more than $\mu + \sigma$ days. Other than above, if the number of days between the last observed date and the end date of objected term and the number of days between the first observed date and the beginning date of objected term are less than or equal to $\mu + \sigma$ days each, the RBS can be considered to exist throughout the targeted term.

Detection of relocation of RBS
Relocation of RBS can be identified by analyzing changes of co-occurrence of observed RBS. First, a RBS named A is vectorized by N dimension. N represents the number of observed RBS. Then, every scan including RBS (A) is

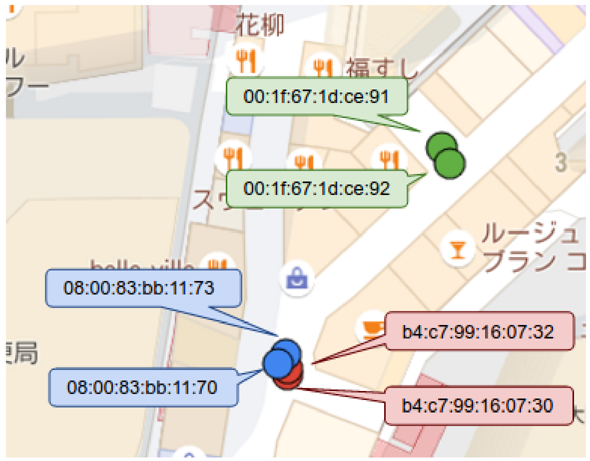

Figure 3: Location of three sets of two different BSSID but with the same G/L bit from the 1.1st to 5.5th octet of BSSID predicted by GMM.

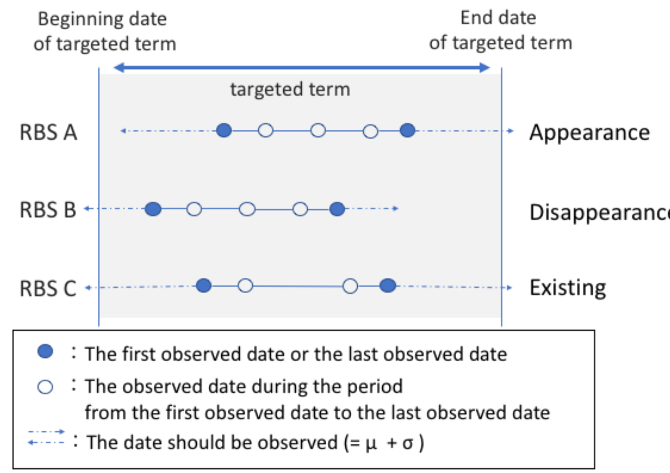

Figure 4: Chronological analysis of RBS.

vectorized as $\vec{A_{tn}}$ at the time t, as shown in Equation (1).

$$\begin{aligned} \vec{A_{t1}} &= (C_{1_{AA}}, C_{1_{AB}}, C_{1_{AC}}, \cdots, C_{1_{AZ}}) \\ &\vdots \\ \vec{A_{tn}} &= (C_{n_{AA}}, C_{n_{AB}}, C_{n_{AC}}, \cdots, C_{n_{AZ}}) \end{aligned} \tag{1}$$

Each dimension of one scan indicates the combination of RBS (A) with other observed RBS (A-Z) and holds binary value whether each of other RBS (A-Z) was observed with (1) or not (0). RBS (A) has 0 as the combination ($C_{n_{AA}}$) with RBS (A) itself. Then, by calculating the balance (the center of gravity) $\vec{g_A}$ of all the vectors A_{ti} (0 $\leq$ i $\leq$ n) and the variance by Equation (2), it can be visualized if the RBS (A) is stably observed. Followed by, by applying the same process to other RBS (B-Z), every scanned data is vectorized and the each variance is calculated. If the variance shows more than a certain value, it can be considered that relocation occurs. As shown above, if relocation occurs, the directions of the vectors are changed since co-observed RBS are changed. Thus, relocation can be detected.

$$\sigma_A^2 = \frac{1}{n}\sum_{n=1}^{n} |\vec{A_{ti}} - \vec{g_A}|^2 \tag{2}$$

Evaluation

Purpose of the experiment

As explained in the precious chapter, this research aims to observe existing RBS and to detect the appearance, disappearance, and relocation of RBS using the data collected from navigation application. It also targets to predict the RBS which are expected to be observed even in the next observation. In addition, we aim to know how the RBS which were installed by carriers would be observed differently compared to other RBS.

Experimental environment

The navigation application which our laboratory developed and released for smartphones named "Umechika-navi"*[4] has been collecting Wi-Fi data (location where each data is scanned is unknown) in the Umeda underground area in Osaka city, the second biggest and structurally complicated city in Japan. This application can collect observed RBS ' s Wi-Fi information including observed BSSID, ESSID, RSSI, scanned time from the whole area of the underground. Usage of this application does not put a heavy burden on a battery. We used the logs collected for 123 days from Sep. 1, 2016 to Jan. 1, 2017. The total number of scan was 21,085. During the term, the number of RBS which was observed more than or equal to once was 6,505. In order to apply the method explained in the precious chapter, the RBS were supposed to be observed more than or equal to twice, thus, 5,719 data were used for the experiment. In addition, the maximum number of days that one RBS was observed was 39 days and the maximum number of times that one RBS was observed was 1,965 times.

The whole program is written in Java8. We used mongoDB to store all the scanned data in the database. We run the program on the PC whose OS is Windows 8.1 Pro, processors are Intel (R) Core (TM) i7- 4770CPU @ 3.40GHz, and RAM is 24.0GB.

Results of detection of appearance and disappearance

From the observed data for more than certain days (2-32days), we derived the number of RBS which were judged as appearance, disappearance, existing condition. The results including the ratio are summarized in Table 2, 3, 4, and 5.

Figure 5 shows each status (appearance, disappearance, and existing) regarding observed days. It took around 40 minutes to calculate each. The ratio of disappearance decreases and the ratio of existing increases before the total observed days are eight. This means that observed data contains noises such as hotspots. The ratio of disappearance increases and the ratio of existing decreases after the total observed days are 16.

All the RBS are predicted as existing after the total observed days are 32. The ratio of existing decreases temporarily but it starts to increase again as the total observed days increases. Thus, it is concluded that eight (the observed days) is the change-point since hotspots are eliminated and the amount of data is not too limited. Hence, it is considered that there are appearance and disappearance of the targeted RBS which were observed for more than eight days.

Results of detection of relocation

We calculated the variance on each of 5,719 RBS which was observed more than or equal to twice. It took 20 hours to complete the calculation. The vertical axis of Figure 6 represents the variance and the horizontal axis represents the number of times observed. In Figure 7, the vertical axis shows the variance and the horizontal axis shows the observed days. Both graphs show each of variance converges to five as the number of scanning times or days increase.

We investigated how the variance will change when relocation occurs. One scan observed nine RBS and 11 scans are conducted one day on average. Theoretical value is calculated when relocation occurs assuming nine RBS can be observed by one scan. In addition, we assumed that the timing of relocation is the one-second, one-third, one-forth, and one-fifth of the whole term. The vertical and horizon-

[4]<Umechika-navi http://www.umechikanavi.jp/>
(Dec. 2, 2016 last accessed)

	Number	Ratio
Appearance	62	10%
Disappearance	125	21%
Existing	412	69%
Total	599	100%

Table 2: The number of RBS and the ratio of RBS observed more than four days and predicted as appearance, disappearance, and existing.

	Number	Ratio
Appearance	16	5%
Disappearance	55	19%
Existing	222	76%
Total	293	100%

Table 3: The number of RBS and the ratio of RBS observed more than eight days and predicted as appearance, disappearance, and existing.

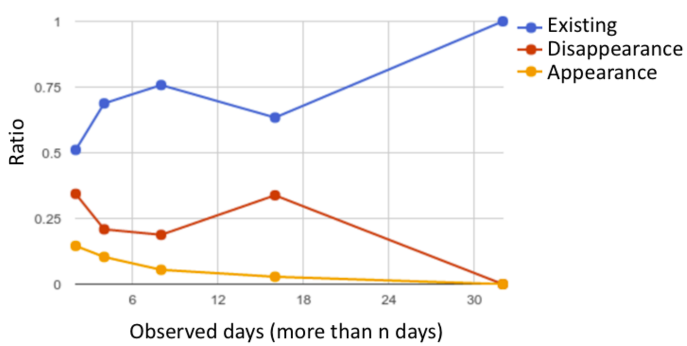

Figure 5: The ratio of each status(appearance, disappearance, and existing) regarding certain n days each.

	Number	Ratio
Appearance	2	3%
Disappearance	24	34%
Existing	45	63%
Total	71	100%

Table 4: The number of RBS and the ratio of RBS observed more than 16 days and predicted as appearance, disappearance, and existing.

	Number	Ratio
Appearance	0	0%
Disappearance	0	0%
Existing	2	100%
Total	2	100%

Table 5: The number of RBS and the ratio of RBS observed more than 32 days and predicted as appearance, disappearance, and existing.

tal axis in Figure 8 shows the variance and the number of times observed, respectively. The vertical and horizontal axis in Figure 9 represents the variance and observed days, respectively. Both graphs show each of variance converges to five as the number of times or days increase. If relocation occurs in the middle of the objected term, the variance becomes a little big. However, it is considered that there are a fluctuation in variance among not relocated RBS since temporal RBS like hotspots are also co-scanned. Thus, by choosing two RBS which were in existing condition throughout the whole term and simulating one RBS which is relocated to the other RBS in one-second, one-third, one-forth, and one-fifth of the whole term, the variance are calculated considering the fluctuation. The vertical axis of Figure 10 represents the variance and the horizontal axis represents

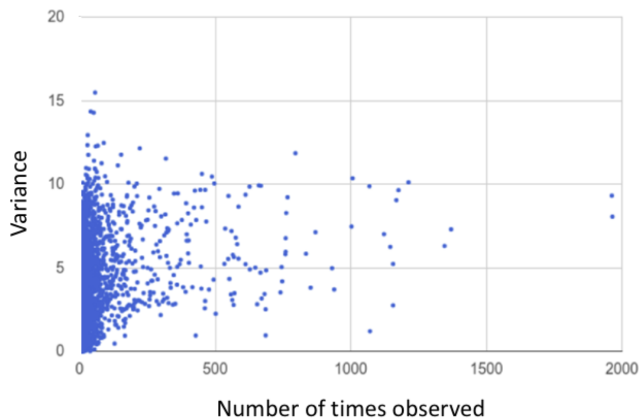

Figure 6: The variance and the number of times observed, calculated for each of 5,719 RBS.

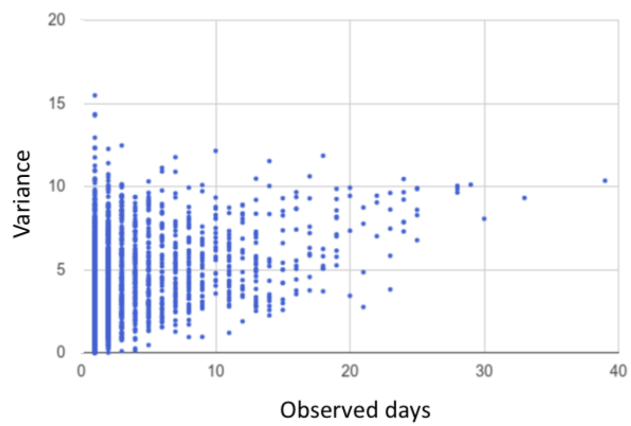

Figure 7: The variance and observed days, calculated for each of 5,719 RBS.

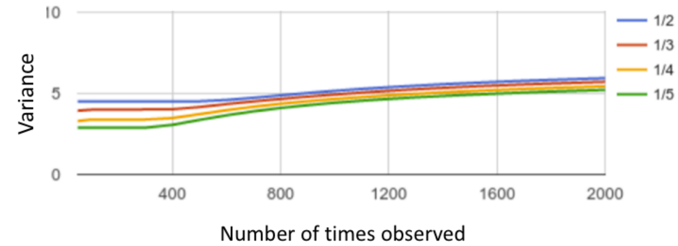

Figure 8: The variance and the number of times observed assuming nine RBS can be observed by one scan (theoretical values).

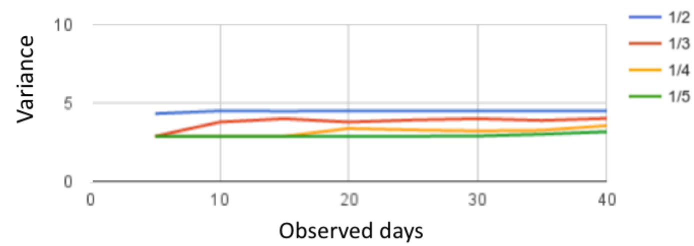

Figure 9: The variance and observed days assuming nine RBS can be observed by one scan (theoretical values).

the number of times observed. In Figure 11, the vertical axis shows the variance and the horizontal axis shows the observed days If relocation occurs early in the objected term, the variance gets a little small. If relocation occurs in the middle of the objected term, the variance gets a little big. However, variance converges to around nine regardless of the times or days. Therefore, if the variance is more than nine, the RBS can be considered to be relocated. Figure 12 shows the variance calculated from the 195 RBS

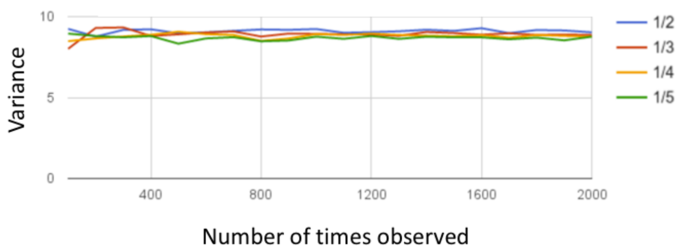

Figure 10: The variance and the number of times observed considering a fluctuation caused by co-scanned and not relocated RBS (e.g. hotspots).

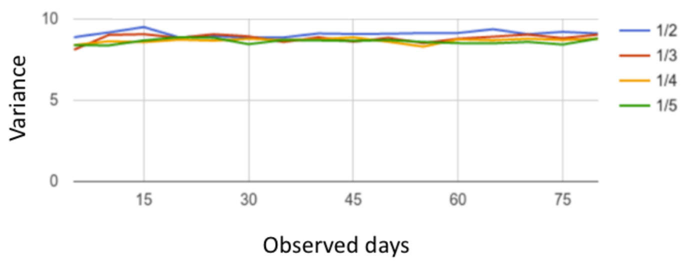

Figure 11: The variance and observed days considering a fluctuation caused by co-scanned and not relocated RBS (e.g. hotspots).

which are thought to be carriers using the above mentioned method. The vertical and horizontal axis in Figure 12 shows the variance and observed days, respectively. As the variance of most of the RBS, is less than nine, it can be considered that the RBS whose variance is more than nine could be relocated.

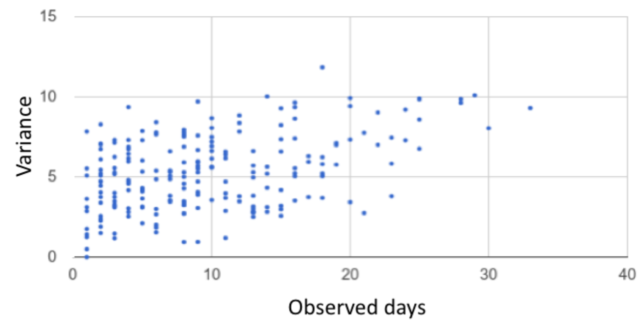

Figure 12: The variance and observed days considering RBS that carriers installed.

Conclusion

This research investigates a method to detect appearance, disappearance, and relocation of RBS only by using logs collected from a navigation application. The accuracy of Wi-Fi positioning is strongly affected by RBS and the number of RBS used in localization. Thus, to detect such events highly affect the over-time deterioration of the radio wave model. Experimental results show that appearance and disappearance can be detected by defining the lower limit of the observed days while relocation can be detected by comparing the theoretical value of variance with the value considering the fluctuation of co-occurrence. In addition, relocation can be detected even among RBS that carriers installed.

The minimum unit used in this research is "day" so intermittent operations of RBS cannot be detected. In the future, we will work on detecting the RBS that operating intermittently by analyzing timewise dimensional change of the RBS whose co-occurred RBS is anomaly compared to the previous observed data.

REFERENCES

1. Chen Feng, Wain Sy Anthea Au, Shahrokh Valaee, and Zhenhui Tan. 2012. Received-Signal-Strength-Based Indoor Positioning Using Compressive Sensing. *IEEE Transactions on Mobile Computing* 11, 12 (2012), 1983–1993.
2. Katsuhiko Kaji and Nobuo Kawaguchi. 2012. Design and Implementation of WiFi Indoor Localization based on Gaussian Mixture Model and Particle Filter. *Indoor Positioning and Indoor Navigation (IPIN)* (2012).
3. Wonho Kang and Youngnam Han. 2015. SmartPDR: Smartphone-Based Pedestrian Dead Reckoning for Indoor Localization. *IEEE Sensors Journal* 15, 5 (2015), 2906–2916.
4. Ling Pei, Ruizhi Chen, Jingbin Liu, Heidi Kuusniemi, Tomi Tenhunen, and Yuwei Chen. 2010. Using Inquiry-based Bluetooth RSSI Probability Distributions for Indoor Positioning. *Journal of Global Positioning Systems* 9, 2 (2010), 122–130.
5. Chouchang Yang and Huai rong Shao. 2015. WiFi-based indoor positioning. *IEEE Communications Magazine* 53, 3 (2015), 150–157.