Adapting Wi-Fi Samples to Environmental Changes Automatically

Takashi Sakaguchi
Ritsumeikan University
1-1-1 Noji-Higashi
Kusatsu, Shiga 5258577
JAPAN
gucci@ubi.cs.ritsumei.ac.jp

Masahiro Mochizuki
Ritsumeikan University
1-1-1 Noji-Higashi
Kusatsu, Shiga 5258577
JAPAN
moma@ubi.cs.ritsumei.ac.jp

Kazuya Murao
Ritsumeikan University
1-1-1 Noji-Higashi
Kusatsu, Shiga 5258577
JAPAN
murao@cs.ritsumei.ac.jp

Nobuhiko Nishio
Ritsumeikan University
1-1-1 Noji-Higashi
Kusatsu, Shiga 5258577
JAPAN
nishio@cs.ritsumei.ac.jp

UbiComp '14, September 13 - 17 2014, Seattle, WA, USA

ACM 978-1-4503-3047-3/14/09c$15.00.
http://dx.doi.org/10.1145/2638728.2641302

Abstract
In recent years, a positioning method which utilizes wireless LAN without using GPS has attracted attention. Especially, in the case of a method which combines absolute position with a Wi-Fi radio environment in advance, the cost of operation and management becomes enormous. Therefore, by sampling Wi-Fi radio information observed at points where users stay frequently or in the long-term, a method which automates to collect and update the Wi-Fi radio information has been proposed. In the case of a long-term operating, the positioning accuracy, however, decreases because this method does not perform well in maintaining and managing samples. It cannot adapt samples to environmental changes although Wi-Fi radio signals change in case of long-term operating. Accordingly, this paper proposes a new calculation formula for improving a positioning accuracy. The formula is calculated with the weight of each base station for avoidance of ill-behaving stations. In addition, this paper also proposes the automated management system with two steps. It adapts samples to changes of Wi-Fi radio signals and a user's behavior. As a result, a positioning accuracy of the new system is higher than existing one.

Author Keywords
Lifelog , Activity recognition, Indoor positioning, Wi-Fi based indoor positioning, Wi-Fi signal change

ACM Classification Keywords
I.5.1 [PATTERN RECOGNITION]: Models-Statistical.; K.6.4 [MANAGEMENT OF COMPUTING AND INFORMATION SYSTEMS]: System Management-Quality assurance.

Introduction
In recent years, a positioning method which utilizes wireless LAN has attracted attention because it is difficult to use GPS in indoor environment. Regarding methods utilizing wireless LAN, many researchers use the method utilizing the Wi-Fi fingerprinting. As for the method, it is necessary to collect Wi-Fi radio information at many points in advance. After that, a Wi-Fi radio environment map, which combines a Wi-Fi radio information collected at a point and absolute location information of the point, must be generated. A Wi-Fi radio information has BSSID which distinguishes each base station, RSSI from them, and so forth. Positioning is to match a Wi-Fi radio information observed in real time with this map. In addition, this method needs to recollect Wi-Fi radio information to update this map. Therefore, the cost of operation and management becomes enormous.

Accordingly, Nishio[1] proposed a method which automates to collect and update the Wi-Fi radio information to cut down the cost. This method assume that our daily indoor activity could be considered as transportation between frequently visited places. This method concentrate on frequently visited and long-term stay places, while neglects transitional states between them. The cost is cut down by positioning at these places. These places can be represented as Wi-Fi radio signal statistics when the accelerometer senses the user stops. Many fragmented Wi-Fi radio signal statistics are analyzed with similar radio signal statistics in order to make clusters. By matching these samples and Wi-Fi observation data, the user's position is estimated. In addition, samples are automatically managed with the $similarity$ called virtual distance. The number of samples are limited by setting a term of generating a sample. Also, samples are adapted to environmental changes by setting a term of updating a sample. In addition, a site concept that is a cluster of log periods each of which have same base stations in common at certain ratio is introduced to share ill-behaving base stations. In case of the operating for a long time, this method, however, does not perform properly to maintain and manage samples. It cannot adapt samples to environmental changes. There are two kinds of environmental changes. One is slow environmental change, such as adding, removing, or moving Wi-Fi base stations, a change of RSSI by a change of temperature and humidity, and a change of RSSI by obstacles or rearranging a room.The other is rapid environmental change such as a change of the point where users stay frequently or in the long-term by moving from the old desk to new one. Therefore, this paper proposes a new calculation formula to improve and maintain the positioning accuracy. Also, the automated management system with two steps for adapting samples to that two kind of environmental changes. One step is a preparatory step to maintain and manage preparatory samples. The other is a positioning step to select available samples for positioning. Samples are automatically adapted to the two kind of environmental changes by using this system.

In the rest of this paper, problems and each of our two approaches are explained. In addition, The automated management system with two steps is evaluated. In the last section, a conclusion and directions for future research are described.

Problems

Nishio calculates the weight of each base station for avoidance of ill-behaving stations. A calculation formula of virtual distance is used, as follow. It is calculated by the Wi-Fi base stationfs MAC address and the strength of Wi-Fi signals. F_w indicates a feature vector of base stations currently observed while F_c indicates the feature vector already sampled. $|vector|$ means the number of base stations. $rssi_w$ indicates a rssi of each base station of F_w while $rssi_c$ indicates a rssi of each base station of F_c.

$$dist(F_w, F_c) = \sqrt{\frac{\sum(rssi_w - rssi_c)^2}{|F_w \cap F_c|} \times \frac{(|F_w| + |F_c| - |F_w \cap F_c|)}{|F_w \cap F_c|}} \quad (1)$$

The weight, however, cannot be reflected in this formula. A new calculation formula for virtual distance needs to accommodate the weight for improving the positioning accuracy.

Adding, removing or moving Wi-Fi base stations, a change of RSSI by temperature and humidity, and a change of RSSI by obstacles or rearranging a room are adverse influences for positioning. It is necessary to adapt samples to the slow environmental change for maintaining the positioning high accuracy. Hence, samples are updated with a Wi-Fi radio information currently observed for maintaining when a user visits places where a sample was generated. If the virtual distance satisfy the terms of updating a sample, a new Wi-Fi radio information must be accumulated to the sample. Therefore, the observation frequency of Wi-Fi base stations observed frequently in a Wi-Fi radio information accumulated to a sample is numerous. Wi-Fi base stations used for calculation of the virtual distance are limited for removing ill-behaving base stations. The calculation uses Wi-Fi base stations, the observation frequency of which are more than 40 percent of the most observation frequency of Wi-Fi base stations. Therefore, a new Wi-Fi base station added to a sample cannot be used for calculation. In addition, each Wi-Fi base station's RSSI used for calculation is the average value. If the data of sample increases, the RSSI is hard to change. It is difficult to adapt samples to slow environmental change if samples have long-term data. Therefore, it is necessary to limit the amount of data used for samples.

Positioning is worse affected by rapid environmental change such as a change of the point where users stay frequently or in the long-term by moving from an old desk to the new one. A new sample is generated if the virtual distance between the sample and others satisfies the terms of generating a sample. If the distance cannot satisfy the terms of generating a sample, a new sample cannot be generated. Therefore, in case a user visits a new place frequently and in the long-term, a new sample cannot be generated. For a higher precision positioning, samples have to be generated in accordance with change of places a user visits frequently or in the long-term.

Calculation Formula Accommodating the Weight for Virtual Distance

The existing calculation formula cannot reflect the weight. The weight can be reflected by using the observation frequency. There are two kinds of the observation frequency. One is the observed frequency of each Wi-Fi base station in a sample. The other is the observed frequency of each Wi-Fi base station of Wi-Fi radio signal statistics of a currently stay. A virtual distance is calculated via the following formula. F_w indicates a feature vector of base stations observed among each user's stay. F_{com} indicates a feature vector of base stations which are common to F_w and F_c. $weight_w$ indicates a weight using the observation frequency involved in a sample. $weight_c$ indicates a weight using the observation frequency involved in Wi-Fi radio signal statistics of a currently stay.

$$dist(F_w, F_c) = \sqrt{\frac{\sum weight_c(rssi_{F_w} - rssi_{F_c})^2}{\sum weight_c(F_{com})}} \times \left(\frac{\sum weight_w(F_{w-com}) + \sum weight_c(F_c)}{\sum weight_c(F_{com})} + \frac{|\sum weight_c(F_{com}) - \sum weight_w(F_{com})|}{\sum weight_c(F_{com})} \right) \quad (2)$$

$$weight = \frac{frequency}{maxFrequency}$$

The Automated Management System with two steps

Samples are automatically managed with two steps for adapting samples to the two kind of environmental changes. One step is a preparatory step to maintain and manage preparatory samples. The other is a positioning step to select available samples for positioning.

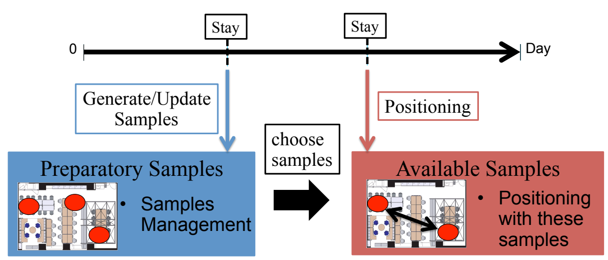

Figure 1: The Automated Management System with Two Steps

Figure 1 indicates our system concept. Preparatory samples are used for maintaining and management. Also, these can be used for detecting samples a user frequently visits. Available samples is important samples for the user. This system estimates user's position with available samples.

Firstly, preparatory step is described. There were two thresholds with the virtual distance. These were used as terms for generating and updating a sample. Additionally, a new threshold is added for generating a preparatory sample. Preparatory samples are generated for adapting a sample to the environmental changes. These are generated at as many places which a user visits frequently or in the long-term as possible. Also, these are updated with Wi-Fi radio signal statistics of a currently stay.

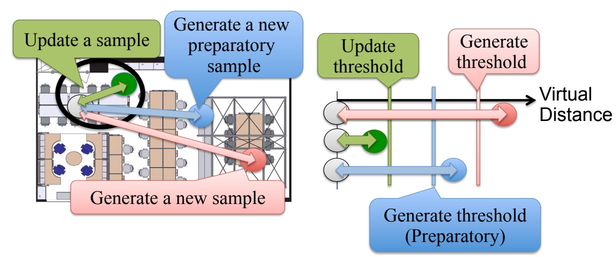

Figure 2: Relationships among Three Thresholds

Figure 2 describes the relationship between these thresholds. "Generate threshold" is set as a term for generating samples. Samples are generated if a virtual distance between Wi-Fi radio signal statistics and a sample is more than "Generate threshold". The distance between each sample is enough to use it for positioning. In addition, "Update threshold" is set as a term for updating samples. Samples are updated if a virtual distance between Wi-Fi radio signal statistics and a sample is less than "Update threshold". Furthermore, "Generate threshold(preparatory)" is set as a new term for generating preparatory samples. This threshold is less than "Generate threshold" and more than "Update threshold". Preparatory samples are generated at as many places a user visits frequently or in the long-term as possible if a virtual distance between Wi-Fi radio signal statistics and a sample is more than "Generate threshold(preparatory)". Preparatory samples are not sufficient to use for positioning. For the preparatory step, "Generate threshold(preparatory)" and "Update threshold" are used to maintain and manage preparatory samples. The amount of data used for samples , however, is needed to limit because it is difficult to adapt samples to slow environmental change in case of long-term data. Therefore, samples consist of the data within the limits of fixed intervals called window size. Figure 3 shows the window. This window slides if the amount of data is more than the window size. Specifically, in case of that, new Wi-Fi radio signal statistics are added to a sample after the oldest Wi-Fi radio signal statistics of the data accumulated to the sample were removed.

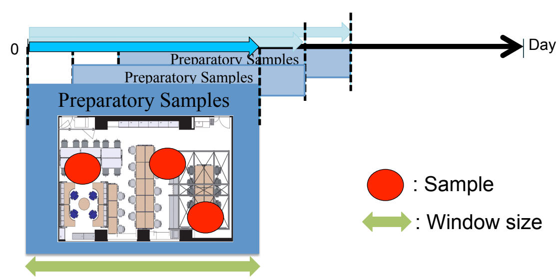

Figure 3: Preparatory Samples

Secondly, positioning step is described. For the positioning step, "Generate threshold" and "Update threshold" are used to select available samples from preparatory samples for positioning. Available samples are selected from preparatory samples to be adapted to rapid environmental change. It is not sufficient for positioning to use preparatory samples because some distance between each of preparatory samples are less than "Generate threshold". Therefore, some samples which are sufficient for positioning have to be selected from preparatory samples. Each sum of time of staying on each preparatory sample can be used for selecting available samples. It is possible to select available samples which generated at places a user frequently visits by selecting available samples from the longest sum of time of preparatory samples. It is sufficient for positioning to use available samples because all virtual distance between each of available samples are more than "Generate threshold". Available samples, however, do not change frequently according to rapid environmental change. Accordingly, available samples are re-selected every fixed interval called refresh interval. Also, Figure 4 shows the refresh interval. This interval is used for calculation of each sum of time of staying on each preparatory sample. It is possible to re-select available samples quickly to use this interval if a places a user visits frequently or in the long-term change.

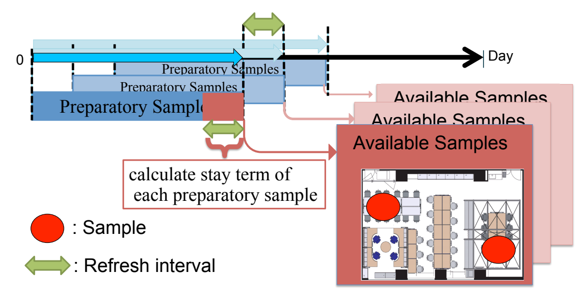

Figure 4: Available Samples

A user, however, has the possibility to visits palaces except where available samples are generated. Therefore, available samples are merged for positioning such these places. The label which is a name to define each sample is used for merging. There are some preparatory samples not to be selected as available samples. Each virtual distance from these samples to available samples is calculated to detect the nearest sample. This virtual distance is calculated with Equation 1 because each weight of Wi-Fi base stations is different in each sample. Figure 5 shows the merging process. The two samples are the nearest are merged. The merged label is generated from these two labels. These samples are renamed with the merged label.

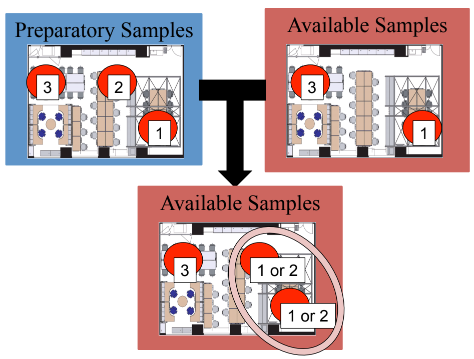

Figure 5: Merge of Samples

Finally, the merged available samples are used for positioning. The distance between a Wi-Fi radio signal statistics which a user observes currently and these samples are calculated. The virtual distance is calculated with Equation 2 in real time. Position is detected by the available sample when the distance value is the smallest. Furthermore, preparatory and available samples are possessed by each site. Sites can not only share ill-behavior base stations but also manage the samples limited number. The samples limited number are easily managed. Equation 1 is used for generating a site.

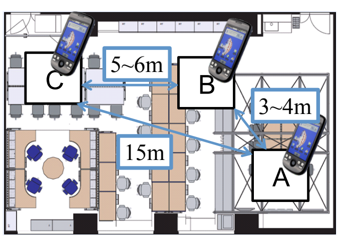

Figure 6: Fixed Observation Points

Evaluation

The automated management system with two steps is evaluated by a simulation of human behavior. An experiment is carried out in our laboratory. A fixed observation was conducted at three places for two weeks. The three places are described in Figure 6.

We have already verified that positioning is sufficient if the distance between samples is more than 6 meters. The distance between "A" and "C" is more than 6 meters. Therefore, the samples generated at "A" and "C" are sufficient for positioning. Although, The distance between "A" and "B" is less than 6 meters. The samples generated at "A" and "B" are not sufficient. The samples generated at "B" and "C" are also the same. In the simulation, a user visits each place at 5-minute interval. In addition, a Wi-Fi radio signal statistics are collected for 5 minutes. After 5 minutes, a user visits other or same place randomly. A Wi-Fi radio signal statistics are also collected for 5 minutes. The simulation is carried out by repeating these steps. This simulation represents the human behavior like a user repeats stay for 5 minute and transfer. A sample is generated and updated with Wi-Fi radio signal statistics collected at each place for 5 minutes. Also, positioning is conducted every 5 minutes. Some patterns exists by changing a transition probability that a user visits after. These patterns are described in Table 1.

Table 1: Simulation Patterns

Pattern Name	Description
PatternAC	A transition probability to "A" and "C" is 40 % each, a transition probability to "B" is 20 %
PatternB	A transition probability to "B" is 60 %, a transition probability to "A" and "C" is 20 % each
PatternACtoB	patternAC is carried out at first, patternB is after a fixed intarval

Adapting sample to slow environmental change is evaluated with PatternAC and PatternB. A user often visits "A" and "C" in PatternAC. Two samples are selected as available samples because the samples generated at "A" and "C" are sufficient for positioning. Slow environmental change is evaluated with PatternAC in case the number of available samples is two . A user often visits "B" in PatternB. A sample is selected as available samples because the samples generated at "A" and "B" are not sufficient. The samples generated at "B" and "C" are also the same. Slow environmental change is evaluated with PatternB in case one sample is generated. Adapting sample to rapid environmental change is evaluated with PatternACtoB. In PatternACtoB, a user changes the point where he stays frequently or in the long-term. A user stays frequently or in the long-term at "A" and "C" at first. A user changes the places where he stays frequently or in the long-term to "B" after fixed interval.

"Window Size", "Refresh Interval", "Generate threshold", "Generate threshold(preparatory)", and "Update threshold" must be set before carrying out the experiment. "Window Size" is set for one week. The shorter it is set, more easily the system adapts the samples to environmental changes. The longer it is set, more stable the samples are against environmental change. "Refresh Interval" is set for one day to re-select available samples quickly. "Generate threshold" is set to 16. The threshold is the virtual distance between "A" and "C". Also, the virtual distance is sufficient for positioning. "Generate threshold(preparatory)" is set to 11. The threshold is less than the virtual distance between "A" and "C". Also, it is less than the virtual distance between "B" and "C". Therefore, three samples are generated with the threshold. "Update threshold" is set to 6. We have verified updating process with various thresholds. As a result, we set to 6 for the threshold.

The Precision and the Recall are calculated by the following formulas for the evaluation.

$$\text{Precision} = \frac{\text{Number of correct system output}}{\text{Number of all positioning}} (\%) \quad (3)$$

$$\text{Recall} = \frac{\text{Number of correct positioning}}{\text{Number of all positioning}} (\%) \quad (4)$$

Precision indicates what is correct for the system. "Update threshold" is used for positioning. The distance between Wi-Fi radio signal statistics which a user observes in currently and samples are calculated. A sample is selected if the distance is the smallest. The sample is selected as an output if the distance is smaller than "Update threshold". Correct behaviors for the system are that outputting an available sample if a user visits a place the sample is generated, and not selecting an available sample if a user does not visit places samples are generated. The Recall indicates the correct for positioning. Correct behaviors for positioning are that outputting an available sample if a user visits a place the sample is generated. It is wrong for positioning not to output a sample.

First, the evaluation of PatternAC is described. Adapting sample to slow environmental change is evaluated in case of two samples are selected as available samples. The automated management system with two steps is compared with the system with one step. There are two versions of the automated management system with two steps. One is the system which is conducted with a merging process. The other is the system which is not. Figure 7 indicates the Precision and the Recall every one day. Samples are generated at "A" and "C" in case of the system with one step. Preparatory samples are generated at "A", "B", and "C". Two samples generated at "A" and "C" are selected as available samples. 80% of positioning is conducted at places where samples that are selected as available samples are generated. On the other hand, 20% of positioning is conducted at places where samples that are not selected as available samples are generated. The Precision of both the system with one step and with two steps is about 80% on the 9th day because these system output sample is generated at "A" when a user visits "B". The Precision of two steps system is improved slightly on the 13th and 14th day. The reason why the Precision is improved is samples of the two steps system can be maintained adequately. The Recall of both systems is about 80% because the Recall decreases when a user visits "B". In addition, the Precision and the Recall of the two steps system which conducts that the system merges available samples is about 100% in the whole period because the system could merge samples generated at "A" and "B".

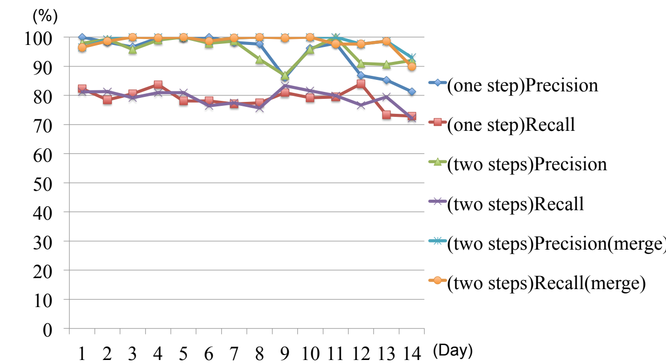

Figure 7: Precision and Recall of PatternAC

Second, the evaluation with PatternB is described. Adapting sample to slow environmental change is evaluated in case a sample is selected as an available sample. The systems to be compared are the same used before. Figure 8 indicates the Precision and the Recall every one day. Samples are generated at "B" in case of the system with one step. Preparatory samples are generated at "A", "B", and "C". A sample generated at "B" is selected as an available sample. 60% of positioning is conducted at places where samples that are selected as available samples are generated On the other hand, 40% of positioning is conducted at places where samples that are not selected as available samples are generated. The Precision of both the system with one step and with two steps is about 80% on the 5th day because these system output sample which is generated at "B" when user visits "A". The Precision of both systems are almost the same in whole period. The Recall of both systems about 60% because the Recall decreases when a user visits "A" and "C". In addition, the Precision and the Recall of the two steps system which conducts with a merging process is about 100% in whole period because the system could merge samples generated at "A", "B", and "C".

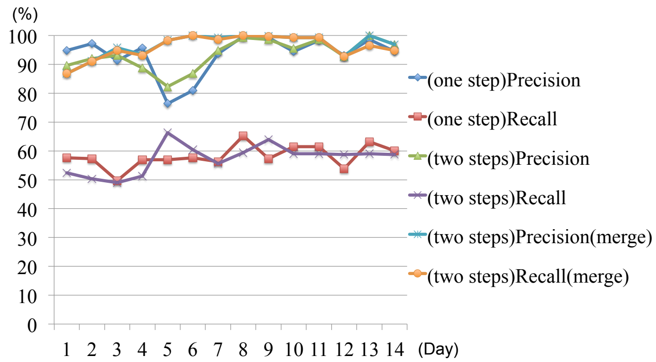

Figure 8: Precision and Recall of PatternB

Finally, the evaluation with PatternACtoB is described. Adapting sample to rapid environmental change is evaluated. the systems to be compared are the same used before. A user frequently visits "A" and "C" at first. Although, he frequently visits "B" after a fixed interval. The pattern change is conducted on the 7th day. Figure 9 indicates the Precision and the Recall every one day. Samples are generated at "A" and "C" in case of the system with one step. Preparatory samples are generated at "A", "B", and "C". Preparatory samples generated at "A" and "C" is selected as available samples in PatternAC and PatternB. A preparatory sample generated at "B", however, is selected as an available sample from 9th day. The Precision of the system with one step decreases on the 10th, 13th, and 14th day because the system outputs a sample which is generated at "A" when a user visits "B". The Recall of both of the system with one step and with the two steps decreases to about 40 % from the 8th day. After changing pattern, 40% of positioning is conducted at places where samples selected as available samples are generated. On the other hand, 60% of positioning is conducted at places where a sample that are not selected as available samples are generated. Hence, the Recall decreased. The Recall of the system with two steps, however, is improved to 60% from 9th day. The system with two steps can adapt samples to rapid environmental change because samples which are important for a user are selected according to a pattern changes. In addition, the Precision and the Recall of the two steps system which conducts with a merging process is about 100% in whole period. It is exceedingly high positioning accuracy in all patterns. The range of positioning, however, is rougher than positioning which does not conduct with a merging process. Therefore, the positioning accuracy and the range of positioning are a trade-off relationship.

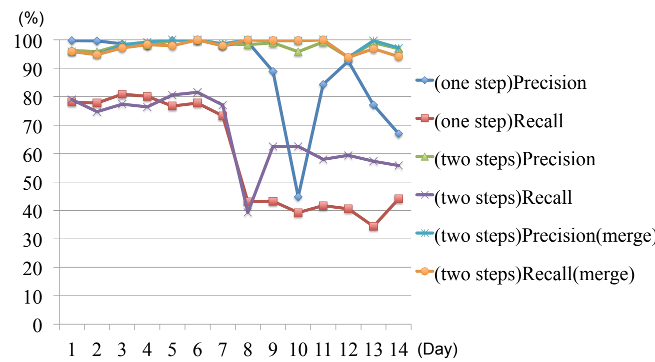

Figure 9: Precision and Recall of PatternACtoB

RelatedWork

There are many researches to address the problem is that positioning accuracy decreases because of environmental change. Two kinds of these research is introduces.

Indoor positioning methods using absolute position
The method must combines a Wi-Fi radio information and absolute location information at first. Pan[2] proposed a method to update a database of Wi-Fi finger printing with Manifold co-Regularization. Yin[3] proposed a method the relationship between a Wi-Fi radio signal from a user's device and one from devices were set in the environment is predicted. These method, however, need to combine a Wi-Fi radio information collected at a point and absolute location information of the point. Therefore, the cost of the maintenance and management is numerous.

Indoor positioning methods not using absolute position
The method don't have to combine a Wi-Fi radio information and absolute location information. Jiang[4] proposed a method automates to correct and update a Wi-Fi radio information with samples fragmented by accelerometer. Positioning is conducted with Bayesian estimation. The range of positing, however, is limited to a room. In addition, a user has to visit five times at a place and a sample has to be consisted of a Wi-Fi radio signal information for over twenty minutes. Therefore, positioning can not be conducted in smaller range and at places a user does not visit frequently.

Conclusion

This paper proposes a new calculation formula and the automated management system with two steps for improving positioning accuracy. The present result suggested that the system can adapt samples to environmental changes. Furthermore, it is exceedingly high positioning accuracy if the two steps system which conducts with a merging process although the positioning accuracy and the range of positioning is trade-off relationship. We examined the system in our laboratory. Positioning accuracy, however, decreases if the number of Wi-Fi base stations is a few. We have to examine the system in outdoor, house, and so forth in the future.

Acknowledgements

Part of this research is supported by SCOPE (#132307011) (Strategic Information and Communications R&D Promotion Program) from the Ministry of Internal Affairs and Communications(MIC), Japan.

References

[1] Nishio Nobuhiko, Fukuzaki Yuuki, and Azumi Takuya. Detecting wi-fi base station behavior inappropriate for positioning method in participatory sensing logs. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, UbiComp '13 Adjunct, pages 665–672. ACM, 2013.

[2] Sinno Jialin Pan, James T. Kwok, Qiang Yang, and Jeffrey Junfeng Pan. Adaptive localization in a dynamic wifi environment through multi-view learning. In *Proceedings of the 22Nd National Conference on Artificial Intelligence - Volume 2*, AAAI'07, pages 1108–1113. AAAI Press, 2007.

[3] Yin Jie, Yang Qiang, and Lionel Ni. Adaptive temporal radio maps for indoor location estimation. In *Proceedings of the Third IEEE International Conference on Pervasive Computing and Communications*, PERCOM '05, pages 85–94. IEEE Computer Society, 2005.

[4] Jiang Yifei, Pan Xin, Li Kun, Lv Qin, Dick Robert P., Hannigan Michael, and Shang Li. Ariel: Automatic wi-fi based room fingerprinting for indoor localization. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 441–450. ACM, 2012.