Abstract—WLAN fingerprint-based positioning system is a viable solution for estimating the location of the mobile station. Many researchers have applied various machine learning techniques to the WLAN fingerprint-based positioning system to make a more accurate system. However, due to the noisy characteristics of the RF signal and lack of the study on environmental factors affecting propagation of the signals, the accuracy of previously suggested systems was highly dependent on environmental conditions. In this paper, we develop multi-classifier for WLAN fingerprint-based positioning system with a combining rule. According to the experiments of the multi-classifier performed in various environments, combining a multiple number of classifiers turned out to mitigate the environment-dependent characteristic of the classifiers. The performance of multi-classifier outperformed other single classifiers in all test environments; the average error distance and standard deviation of the error distance were improved by multi-classifier in all test environments.

Index Terms—Location-based service, WLAN, Fingerprint, Positioning system, Multiple Classifier system

I. INTRODUCTION

With the proliferation of smart phones, WLAN (Wireless Local Area Network) based positioning systems have become a main stream in Location-based Service (LBS). Compared with other technologies such as GPS [1], RFID [2], GSM [3], Ultrasonic [4], infrared-based system [5], etc., WLAN-based positioning systems have advantages in terms of coverage and costs. Most of researches on WLAN-based positioning systems use Received Signal Strength Indication (RSSI) from wireless network access points because RSSI, or called fingerprint, is easy to obtain using software, and one of the most relevant factors for positioning.

There are some studies of considering other factors such as Signal to Noise Ratio (SNR), Angle of Arrival (AOA), and Time of Arrival (TOA) for positioning systems. For example, Milos et al. [6] examined SNR as an additional input factor, and reported that considering both SNR and RSSI increases the performance of WLAN-based positioning system. R. Yamasaki et al. [7] reported that AOA and TOA are important factors in positioning. However, acquiring the factors including AOA, TOA, and SNR are not always possible in every wireless network interface card. As a result, RSSI has been adopted as a primary factor in WLAN-based positioning system.

In fact, utilizing the strengths of Radio Frequency (RF) signals for positioning is not a simple work. Due to the characteristics of RF signals like multipath fading and interference between signals, the signal strength severely changes depending on materials, position of doors and windows, width of the passage, the number of APs deployed, etc. Even if the fundamental parameters are known beforehand, deriving the path loss function of a WLAN signal is extremely complex. In this case, most of the WLAN fingerprint-based positioning system take statistical approaches [6].

Previously suggested statistical approaches applied various machine learning techniques to derive the position from measured fingerprints [2, 8-15]. The techniques usually comprise two phases: off-line and on-line phase. In the off-line phase, fingerprints are captured at various positions in the target place, and stored in a database called radio-map. In the on-line phase, the location of a fingerprint is estimated by comparing it with the stored fingerprints in the database.

The problem of the WLAN fingerprint-based positioning systems is that the performance of the system is too much dependent on environments; in other words, there is no general solution in WLAN fingerprint-based positioning system yet. Each system is designed to tackle different environments, and there is no analysis on the relation between algorithm and the test environments. One method may outperform other methods in an environment, but it may show inferior results in other environments. As an instance, Youssef et al. [12] suggested Joint-clustering technique, and they confirmed their proposed algorithm outperformed RADAR [2] in their evaluation. In Wilson et al.’s experiment [11], however, RADAR showed superior performance compared with Joint-clustering technique. This problem is also observed from our experiments.

In this paper, we introduce multi-classifier for WLAN fingerprint-based positioning system. We combined multiple classifiers to make an environment-independent classifier achieving stable and high estimation accuracy in diverse environments. The motivation of using a multiple number of classifiers is that the performance of classifier is severely dependent on the environments; therefore, if we can select the most accurate classifier for given situation, we can achieve the best performance in diverse environments. Thus, the key of utilizing a multiple number of classifiers is the ground of selecting one classifier among a set of classifiers. To combine a multiple number of classifiers, we used the Bayesian combination rule [22] and majority vote [23].
To prove the effect of combining classifiers, we evaluated the proposed system in three different environments. The evaluation results revealed that the multi-classifier outperforms single classifiers in terms of average error distance and standard deviation of the errors. This indicates that the proposed combining method is effective in mitigating the environment-sensitive characteristics of WLAN-based positioning system.

The remainder of this paper is organized as follows. The overview on WLAN fingerprint-based positioning is given in section II. We introduce multi-classifier for WLAN fingerprint-based positioning system in section III. Section IV describes the experimental setup and the results. In section V, we summarize our work and present the future work.

II. RELATED WORK

The location estimation using WLAN fingerprint often refers to the machine learning problem because estimating the signal propagation is a very complex task. In this reason, various machine learning techniques have been applied.

The RADAR system developed by Bahl et al. [2] is the most representative WLAN fingerprint-based system. In this system, the authors use Pentium-based PCs as access points and laptop computers as mobile device. The system uses nearest neighbor heuristics and triangulation methods to infer a location of user. It maintains a radio map which charts the signal strength received from different access points at selected locations. Each signal-strength measurement is then compared against the radio map and the positions of the best matches are averaged to give the location estimation. Roos et al. [10] proposed the probability-based system which uses the received signal strength samples to create the probability distributions of signal strength for some known locations. Once an input instance is given, it matches to these probability distributions to find out the location of the mobile device with the highest probability. The histogram method suggested by Castro et al. [24] is another example of the probability-based system. Instead of using Gaussian distribution, it derives the distribution of signal strength from the learning data. In addition, adaptive neural networks [13], decision tree [14-15], and support vector machine [16] are popular on WLAN-based positioning system; Kushki et al. [8] suggested the kernelized distance calculation algorithm to infer the location of measured RSSI.

Recently, researchers focused on compensating the characteristics of RF signal. Berna et al. [17] suggested the system using the database considering the unstable factor such as open/close doors and humidity changing. They utilized sensors to capture current status of the environment. Yin [15] introduced the learning approach using the database temporally updated depend on the current environment. Moraes [18] investigated the dynamic RSS mapping architecture. Wilson Yeung et al. [11] suggested using RSSI transmitted from mobile device as an additional input. Basically, they have two types of database: RSSI transmitted by APs and RSSI transmitted by mobile device. In the on-line phase, the system inferences multiple results from databases, and provides the final decision using combining method.

Some researches [12, 20] has tackled the issue of how to reduce the computational overhead; because of the client devices are usually small, self-maintained devices depend on limited power. Youssef et al. [12] developed a joint-clustering technique to group locations in order to reduce the computational cost of the system. In this method, a cluster is defined as a set of locations sharing the same set of access points. The location determination process is as follow: for a given RSSI data, the strongest access points are used to determine one cluster to search within for the most probable location. Chen et al. [20] suggest a method which selects the most discriminative APs in order to minimize the used AP numbers in the positioning system. This approach selects an appropriate subset of the existing features to the problem of computational complexity. Reducing the number of APs is referred to the dimension reduction in a signal space, which reduces the computational overheads required on the mobile device.

The problem of WLAN fingerprint-based positioning system is that the performance of the system is too much dependent on the environment. One system may outperform the other methods in an environment; it may show inferior performance in other environments. To solve this problem, we suggest multi-classifier for WLAN fingerprint-based positioning system. Our proposed system utilizes multiple numbers of classifiers to make more accurate results. The overview of our proposed multi-classifier is given in the next section.

III. PROPOSED METHOD

We utilize multiple numbers of classifiers using different algorithms to build environment-independent classifier. Combining multiple numbers of classifiers to create a strong classifier has been a well-established research area in the pattern recognition, so called Multiple Classifier System (MCS) [19]. When it comes to the term combining, it indicates selecting the most trustable prediction result attained from classifiers.

At least two reasons justify the necessity of combining multiple classifiers [21]. First, there are a number of classification algorithms available developed from different theories and methodologies for almost any one of the current pattern recognition application areas. Usually, for a specific
application problem, each of these classifiers could reach a certain degree of success, but maybe none of them is totally perfect or at least one of them is not so good as expected in practical application. Second, for a specific recognition problem, there are often many types of features which could be used to represent and recognize patterns. These features are also represented in various diversified forms and it is rather hard to lump them together for one single classifier to make a decision. As a result, multiple classifiers are needed to deal with the different features. It also results in a general problem how to combine those classifiers with different features to yield the improved performance.

The location estimation using WLAN fingerprint often refers to the classification problem because of the noisy characteristics of the RF signals. Many algorithms have been proposed based on the different machine learning techniques; and, none of them achieve the best performance in diverse environments. At this point, we realized that utilizing multiple numbers of classifiers can be a solution for general solution in WLAN fingerprint-based positioning system.

In this work, we combine the Bayesian combination rule [22] and majority vote [23] to our multi-classifier. The Bayesian combination rule weights the decisions of classifiers based on the basis. Usually, the basis is given a form of matrix called confusion matrix. The confusion matrix is constructed by cross-validation with learning data in the off-line phase. The majority vote is a simple algorithm, which is based on the basis. Usually, the basis is given a form of matrix called confusion matrix.

Figure 1 illustrates the idea of our proposed system. In the off-line phase, fingerprints are collected over the target environment as learning data. The fingerprint is a collection of pair-wise data containing MAC address of an access point and its signal strength. Usually, in one fingerprint, there are multiple tuples of this pair-wise data such as \{AP1, BSSI1\}, \{AP2, BSSI2\}, \{AP3, BSSI3\} ... . After attaching the collected location label to the fingerprint, the database stores the labeled-fingerprint data.

After collecting learning data, each classifier C constructs their own confusion matrix M (Figure 2) using cross-validation with the learning data. The confusion matrix would be used as indicator of its classifier. If there are L possible locations in positioning system, M is a L x L matrix in which the entry M_{ij} denotes the number of instants collected in location i, that is assigned as location j by the classifier.

From the matrix M, the total number of data collected in location i can be obtained as a row sum \(\sum_{j=1}^{K} M_{ij}\), and the total number of data assigned to location j can be obtained as a column sum \(\sum_{i=1}^{K} M_{ij}\). When there are K classifiers, there would be K confusion matrices \(M^{(k)}\), 1 \(\leq k \leq K\).

In the on-line phase, for the measured fingerprint x, the positioning results gained by K classifiers are \(C_k(x) = j_k\) (1 \(\leq k \leq K\)) and \(j_k\) can be any location of the L possible locations. The probability that the decision made by the classifier \(C_i\) is correct can be measured as follow:

\[
\varphi(j_k) = P(x \in j_k | C_1(x) = j_1, ..., C_K(x) = j_K)
\]

The equation (1) is called belief function, and the value of this function is called belief value. Assuming that all classifiers are independent, and applying Bayes’ theorem to the equation (1), the belief function \(\varphi(j_k)\) can be reformulated as:

\[
\varphi(j_k) = \prod_{i=1}^{K} \frac{P(x \in j_k \cap C_i(x) = j_i)}{P(C_i(x) = j_i)}
\]

(2)

The denominator and numerator in the equation (2) can be calculated with the confusion matrix M. The denominator indicates the probability that classifier \(C_i\) assigned unknown fingerprint x as \(j_i\). This can be presented as follow:

\[
P(C_i(x) = j_i) = \sum_{j=1}^{K} M_{ij}
\]

(3)

The numerator in the equation (2) means the probability of classifier \(C_i\) assigning the unknown fingerprint x collected in \(j_i\) to \(j_k\). This term is simply described as below:

\[
P(x \in j_k \cap C_i(x) = j_i) = \frac{M_{jk/i}}{\sum_{j=1}^{K} M_{ij}}
\]

(4)

After applying equation (3) and (4) to the equation (2), the equation (2) can be reformulated as:

\[
\varphi(j_k) = \prod_{i=1}^{K} \frac{M_{jk/i}}{\sum_{j=1}^{K} M_{ij}}
\]

If more than a half of estimation of the classifiers point a specific location, the location would be selected as the final result. Otherwise, the belief value of each prediction is cal-
calculated, and the location with the highest belief value would be the final result. In case there are many locations with the same highest belief value, the multi-classifier system determines the middle point of those locations as the final result.

For example, there are three classifiers, $a$, $b$, and $c$, and there are three possible locations, location 1, location 2 and location 3. After the off-line phase, the confusion matrices as follow:

$$M^{(a)} = \begin{pmatrix} 18 & 4 & 7 \\ 2 & 12 & 3 \\ 0 & 4 & 10 \end{pmatrix}$$

$$M^{(b)} = \begin{pmatrix} 12 & 6 & 6 \\ 3 & 9 & 3 \\ 2 & 5 & 11 \end{pmatrix}$$

$$M^{(c)} = \begin{pmatrix} 14 & 2 & 2 \\ 4 & 11 & 5 \\ 2 & 7 & 13 \end{pmatrix}$$

If classifier $a$, $b$, and $c$ assigned the unknown instance $x$ to location 1, location 2, and location 3 respectively, the belief value of predictions can be calculated as follow:

$$\phi(a) = \frac{18}{29} \times \frac{3}{15} \times \frac{2}{22} = \frac{108}{9570}$$

$$\phi(b) = \frac{4}{29} \times \frac{9}{15} \times \frac{7}{22} = \frac{252}{9570}$$

$$\phi(c) = \frac{7}{29} \times \frac{3}{15} \times \frac{13}{22} = \frac{273}{9570}$$

The multi-classifier assigns location 3 to the unknown instance $x$, because $j_c$, the prediction of the classifier $c$, has the highest belief value.

IV. EVALUATION

A. Experimental Setup

The performances of WLAN-based positioning systems depend on the environment where the evaluation is conducted. In this reason, we evaluated the proposed multiclassifier in three different environments; table 1 briefly illustrates the test environments.

The testbed 1 is an office environment; the dimension of the corridor in the office is $3 \times 60$ m. The location of the office is the 3rd floor of the faculty building of the KAIST-ICC in Daejeon, South Korea. In the corridor, we collected 100 samples of Fingerprint from 60 different locations. Each location is 1 meter far from each other. The testbed 2 is another office environment, and the dimension of the corridor is $4 \times 45$ m. The office is located at the 2nd floor of the Truth building of the KAIST-ICC. We collected 100 samples of Fingerprint from 45 different locations. Each location is 1 meter far from each other. The testbed 3 is a large and empty space inside the building located at the 1st floor in the Lecture building of the KAIST-ICC. The dimension of the space is $15 \times 15$ m. In the testbed 3, we collected 100 samples of Fingerprint from 25 locations. Each location is 3 meter far from each other on a $5 \times 5$ lattice.

Figure 3. Average error distance versus number of AP used for positioning in a) Testbed 1, b) Testbed 2, and c) Testbed 3 respectively.
B. Result

From the results, we can observe that none of the single classifier outperformed others in all three test environments. These results indicate that the performance of WLAN fingerprint-based positioning system is sensitive to the environments and the multi-classifier is turned out to be effective in mitigating such characteristics of WLAN signals.

Figure 3 reports the average error distance with respect to different number of APs. From figure 3-(a) and (b), the performances of classifiers are quite different in proportion to the environments. Although testbed 1 and testbed 2 are similar indoor environments, the performances in testbed 1 were better than in testbed 2. Especially, the average error distance of $k$-NN classifier in testbed 1 was 1.21 meter when 15 APs used for positioning, while it was 4.6m in testbed 2. In case of the histogram classifier, the average error distances were 1.9 and 2.7 meter with 15 APs in testbed 1 and testbed 2, respectively. With the same condition, the Naïve Bayesian classifier’s average error distance in the testbed 1 and 2 were 1.25 and 2.47 meter, respectively.

Compared with other classifiers, the multi-classifier showed the improved results. In the testbed 1 and 2, the average error distances of the multi-classifier with 15 APs were 1.1 and 2.3 meter, respectively. In the testbed 3, the accuracies of all classifiers are extremely poor than the results in the other testbeds. This result supports that WLAN fingerprint-based positioning system shows better performance in office environment compared against the hall environment which has a few attenuation factors. From the result on depicted in the figure 3, the multi-classifier mitigate the environment-dependent characteristic of single classifier, and there are marginal improvements of the performance in terms of the average error distance.

Table 2 illustrates the performance summary of classifiers. The standard deviation of error of multi-classifier in the testbed 1 was 3.6 meter, while this of $k$-NN, Histogram, and Bayesian were 5.8, 3.8, and 3.9 meter respectively. In the testbed 2, the standard deviations of error of all classifiers were lower than the values in the testbed 1. The standard deviation of $k$-NN, Histogram and Bayesian were 3.0, 2.5, and 1.8 meter respectively. The standard deviation of error of the $k$-NN, histogram, and Bayesian classifier in testbed 3 were 4.5, 4.6, and 5.1 meter, respectively. These results confirm the standard deviation error of the WLAN fingerprint-based positioning system is also dependent on the environment. Proposed multi-classifier outperformed others in all testbeds in terms of the standard deviation of error. In testbed1, 2, and 3, the standard deviations of error of the multi-classifiers were 3.6, 1.6, and 4.5, respectively, which are higher or equivalent performance with others.

From the results, we confirmed that multi-classifier mitigated the environment-dependent characteristic of single classifier, and the performance of multi-classifier was higher than others in all environments. Even if the performance was improved in marginal, we observed that combining a number of classifiers is one of the promising approaches to make environment-independent WLAN fingerprint-based positioning system.

V. SUMMARY AND FUTURE WORK

In this paper, we presented multi-classifier for WLAN fingerprint-based positioning system to mitigate the effects of environmental factors, and construct environment-independent classifier. We developed a combining method to utilize multiple numbers of classifiers. The purpose of combining multiple numbers of classifiers is error-correction; for example, if a single classifier predicted wrong, the other classifiers correct it. In other words, the classifiers in multi-classifier complement each other.

We evaluated multi-classifier in three different environments with various environmental factors: number of APs, the width of corridor, the materials, etc. The multi-classifier was constructed with three classifiers which are $k$-NN (with $k=3$), Bayesian, and Histogram classifiers. The experimental result indicates that multi-classifier shows consistent performance in diverse environments while other classifiers’ performances were various. The performance of multi-classifier tends to follow the performance of the classifier showing the best performance among single classifiers. This means that the classifiers in multi-classifier complement each other, therefore the error was corrected.

For the next step, we are going to investigate more efficient combining rule. In this work, we mixed the Bayesian combining rule and majority vote; however, the performance enhancement was too marginal. Considering the complexity overhead of using multiple numbers of classifiers, the multi-classifier are not effective in cost-effective aspects.

Finding the best combination of classifiers is another direction of our work. We tested only three classifiers, and two of them take similar approaches; fingerprint is the only feature for positioning. There are numbers of system considering various aspects of WLAN signals, which used additional features. We are going to implement and evaluate the multi-classifier with various types of classifiers.

<table>
<thead>
<tr>
<th>Testbed 1</th>
<th>Testbed 2</th>
<th>Testbed 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$-NN</td>
<td>Histogram</td>
<td>Bayesian</td>
</tr>
<tr>
<td>Average (m)</td>
<td>4.0</td>
<td>2.6</td>
</tr>
<tr>
<td>Std.Dev (m)</td>
<td>5.3</td>
<td>3.8</td>
</tr>
<tr>
<td>Max (m)</td>
<td>43</td>
<td>29</td>
</tr>
<tr>
<td>Min (m)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>90th Percentile (m)</td>
<td>12.0</td>
<td>7.0</td>
</tr>
</tbody>
</table>
REFERENCES