

## Research Article

# Efficient Sensor Localization Method with Classifying Environmental Sensor Data

Ae-cheoun Eun and Young-guk Ha

Department of Computer Science and Engineering, Konkuk University, Seoul 143-701, Republic of Korea

Correspondence should be addressed to Young-guk Ha, ygha@konkuk.ac.kr

Received 31 July 2012; Revised 15 October 2012; Accepted 30 October 2012

Academic Editor: Shan Lin

Copyright © 2012 A.-c. Eun and Y.-g. Ha. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sensor location estimation is important for many location-based systems in ubiquitous environments. Sensor location is usually determined using a global positioning system. For indoor localization, methods that use the received signal strength (RSS) of wireless sensors are used instead of a global positioning system because of the lack of availability of a global positioning system for indoor environments. However, there is a problem in determining sensor locations from the RSS: radio signal interference occurs because of the presence of indoor obstacles. To avoid this problem, we propose a novel localization method that uses environmental data recorded at each sensor location and a data classification technique to identify the location of sensor nodes. In this study, we used a wireless sensor node to collect data on various environmental parameters—temperature, humidity, sound, and light. We then extracted some features from the collected data and trained the location data classifier to identify the location of the wireless sensor node.

## 1. Introduction

Location-aware services are an important application of ubiquitous computing. Therefore, in wireless sensor networks (WSNs), localization has become an essential functionality. Essentially, the localization of a wireless sensor node is achieved by measuring the received signal strength (RSS) of wireless links between the target node and multiple reference nodes and using the theory that the signal strength of the wireless link between two wireless nodes decreases as the distance between them increases. Measured RSS data are used to determine the location of the target node in methods such as triangulation [1], a centroid method [2], or fingerprinting [3, 4]. However, such a method has some limitations when used in indoor environments owing to the reflection, loss, and distortion of signals because of the presence of indoor obstacles. In addition, the RSS between two sensor nodes for a given distance decreases with the battery capacity of the sensor nodes.

In this paper, we propose a novel localization method for sensor nodes in indoor wireless sensor network environments [5]. The method involves the classification of

environmental data, such as temperature, humidity, sound, and light, collected by the target nodes. To classify these environmental data according to the locations where they were recorded, we use a k-nearest neighbor (k-NN) classifier. In addition, we use a feature extraction method for the recognition through principal component analysis (PCA). We then perform localization experiments in an actual test environment to validate the proposed method.

The rest of this paper is organized as follows. In Section 2, the existing sensor localization methods and some problems that arise when using these methods in real-world applications are analyzed. In Section 3, we describe the design of the localization method proposed in this paper. In Section 4, the implementation of the method is explained and experimental results are discussed. Finally, in Section 5, the paper is summarized and future directions are given.

## 2. Related Work

2.1. Well-Known Localization Methods. Triangulation techniques include RSS indicator (RSSI) [6], time of arrival (ToA), time difference of arrival (TDoA), and angle of

arrival (AoA). RSSI measures the attenuation of the radio signal strength between a sender and a receiver. The power of the radio signal decreases exponentially with increasing distance, and the receiver can measure this attenuation and use it to estimate the distance from the sender. ToA [6–8] is based on the speed of radio wave propagation and the time that a radio signal takes to move between two objects. Combining these pieces of information allows a ToA system to estimate the distance between a sender and a receiver. TDoA [6, 9] measures the difference between arrival times. Beacon nodes are necessary to transmit both ultrasound and radio frequency (RF) signals simultaneously. A sensor measures the difference between the arrival times of the two signals and relays the range to the beacon node. Unlike the above techniques, which measure distance, AoA [10] techniques measure the angle at which a signal arrives. Angles can be combined with the estimated distance or other angle measurements to derive positions. AoA is an attractive method because of the simplicity of the subsequent calculations.

The use of triangulation methods for indoor environments is very problematic because they use the RSS; the drawback [11] of using the RSS has been described in Introduction. Thus, to avoid these problems, other methods should be used.

**2.2. RF Fingerprinting.** A fingerprinting [3, 4, 12] algorithm is usually the basis of a WLAN localization system. The proposed technique, based on the discriminant-adaptive neural network (DANN) [3] architecture, is implemented in a real-world WLAN environment, and realistic measurements of the signal strength are collected. This technique is used to extract useful information from available access points (APs) and transmit the information to the discriminative components (DCs). These components use this information for discriminating between different locations and rank it according to its quantity. Rank the locations according to the respective access point. The technique incrementally inserts DCs and recursively updates their weightings in the network until no further improvement is required. The network can accomplish learning intelligently using the information provided by the inserted DCs. Moreover, the weights of the input layer and the inserted components are determined using multiple discriminant analysis (MDA) [13] in order to maximize the useful information contained in the network. The RF fingerprinting technique also uses RSS values to determine the position of a sensor node. Thus, the problem explained in Section 2.1 is faced.

**2.3. eWatch System.** eWatch [14] is a wearable sensing, notifying, and computing platform that resembles a wrist-watch, a factor that renders it very accessible, instantly viewable, ideally located for sensors, and unobtrusive to its users. Information transfer from eWatch to a cellular phone or stationary computer occurs through wireless bluetooth communication.

eWatch senses light, motion, sound, and temperature and provides visual, sound, and tactile notification. It has ample processing capabilities and a multiday battery life, which allows realistic user studies. This paper describes the

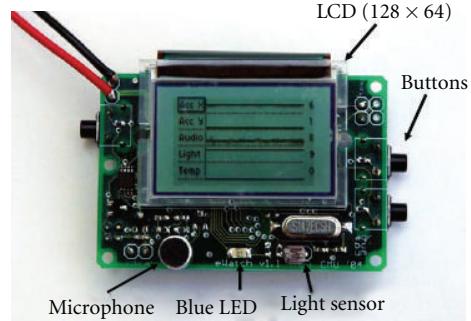


FIGURE 1: Top view of the eWatch board.

motivation for developing a wearable computing platform, a description of power-aware hardware and software architectures and demonstrates the identification and recognition of a set of frequently visited locations via online nearest-neighbor classification.

Figure 1 shows the board that was used for data collection and analysis in the eWatch project. eWatch finds a location using three environmental parameters: sound, temperature, and light. Note that the use of more parameters would increase the localization accuracy. In this paper, we discuss methods for measuring a user's location by using four parameters: sound, temperature, light, and humidity. In the present study, these sensing data were used in location-aware technology.

### 3. Design of the Proposed Method

In this section, we explain the design of the proposed system and describe the architecture and design concepts. In addition, details of the method for each module will be discussed.

**3.1. System Architecture.** Figure 2 shows the overall system architecture and data flow. The location data collection module (LDCM) periodically collects environmental data of each space and provides the data to the system. The environmental data of each space consists of temperature, humidity, light, and sound data.

The collected environmental data of each space is used for training the user location recognition module (ULRM). The location data feature extraction module (LFEM) provides a feature extraction function. This function is applied to the environmental data of the user location provided by the LDCM. The extracted features are input into the ULRM for the purpose of user location recognition. Primarily, feature extraction is used to decrease the amount of high-frequency data. In the LFEM, the data are converted from the format of the ULRM training module to the attribute-relation file format (ARFF) used by Weka [15]. Weka is a data mining tool. In addition, the LDCM module can sense the current environmental data communicated in the location test. Finally, the sensed and trained data will be used as test data to recognize a user's location.

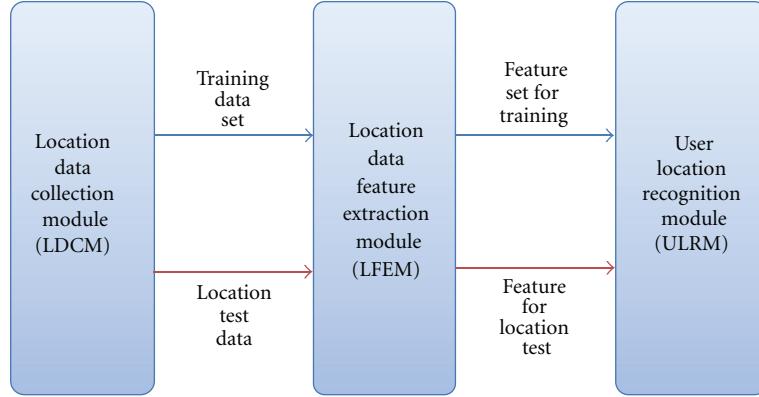


FIGURE 2: System architecture.

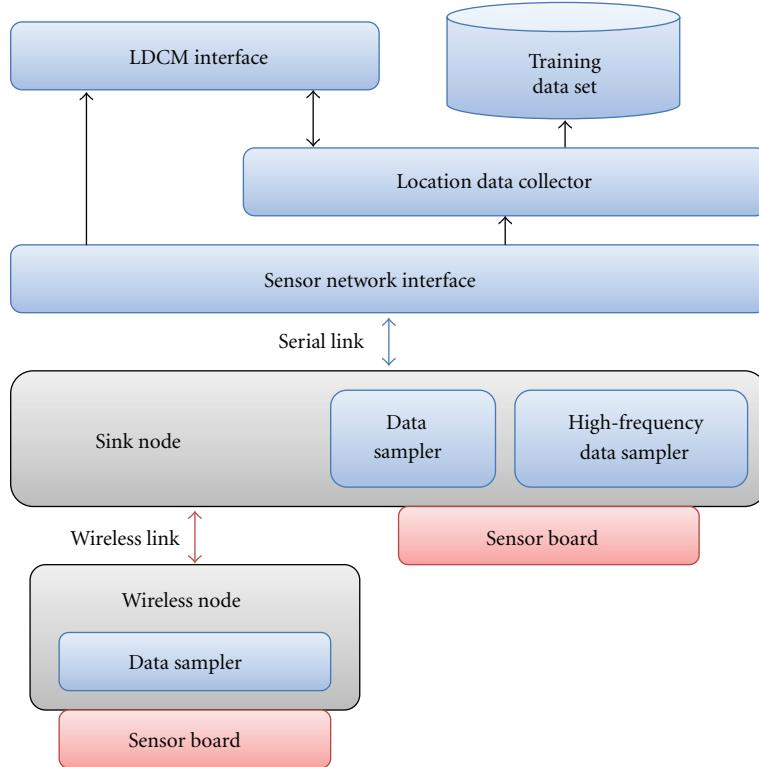


FIGURE 3: LDCM.

In addition, the LFEM uses a different extraction method for each feature. It uses PCA for feature extraction. In PCA, the number of principal components is less than or equal to the number of original variables. The ULRM uses a set of trained data for recognizing location. In this section, we discuss the data format for data training and that of the collected data. In addition, the ULRM shows the location recognition results based on real-time data extracted from the LFEM module.

**3.2. LDCM.** This section describes the elements of the LDCM. Figure 3 shows the structure of the LDCM. This module periodically senses and collects the environmental data of each space and provides it to the system. These data

are then used for recognizing the user location. The WSN [16] consists of a wireless sensor node and sink nodes. A Hmote2420 sensor, which can sense temperature, humidity, light, and sound, is used in the sensor board.

The wireless sensor node loads data from the data sampler program and sensor board. Thus, the sensor nodes can acquire environmental data from the sensor board. While the data (temperature, humidity, light, and sound data) are being sent, the WSN can also send the data to the sink node through a wireless link by using a sampler program. The wireless link operates in the half-duplex transmission mode. The sink node delivers sensor data to the base station and the sensor network interface through a serial link. The sink node can also acquire environmental data directly from

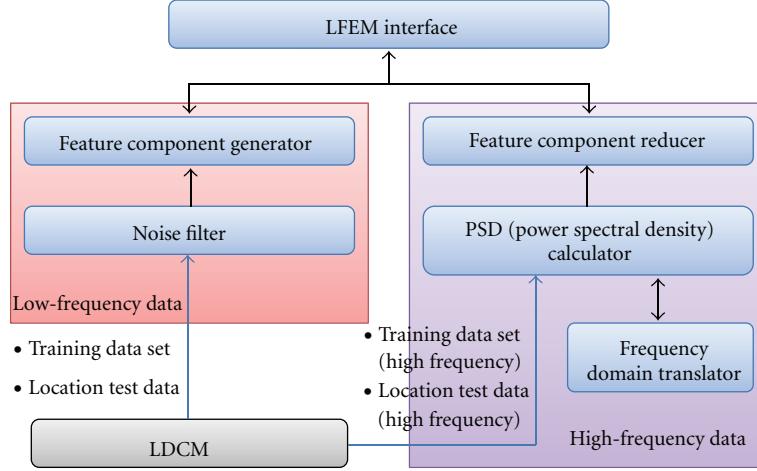


FIGURE 4: LFEM.

the installed data sampler and sensor board, but not through the wireless sensor node. The sink node has a high-frequency data sampler for sampling high-frequency data effectively. Two types of samplers, a high-frequency sampler and a low-frequency sampler, are used because of the very large amount of processing required for high-frequency data.

The sensor network interface links the sensor network to a base station. The hardware interface, such as USB or RS-232, uses a common serial link. On the other hand, the software interface has a device driver and a system application programming interface (API) for processing data received from the serial link. The location data collector saves environmental data in the data file of the training set.

This training set is created after the data file is given as the input to the LFEM, and it is used by the LFEM for training the ULRM with the feature extraction process. The LDCM interface provides an API, which can be used to obtain environmental data at the user's location. In the next section, the data extraction method will be explained.

**3.3. LFEM.** In our system, the LFEM performs data extraction. The structure of the module is shown in Figure 4. The extraction method used in the LFEM depends on the type of environmental data used. We perform noise filtering for low-frequency data and determine the power spectral density (PSD) for high-frequency data. Therefore, the collection of low-frequency data, such as temperature and humidity, involves noise filtering. Noise filtering helps distinguish between usable data and unusable data. Thus, our module acquires only usable data. However, high-frequency data, such as sound and light, are not subjected to noise filtering.

For collecting high-frequency data, the PSD should be used. Sound data and the top five principal component data are then extracted through frequency domain conversion. These real-time data are provided as input to the LFEM interface. They are used for feature extraction in the ULRM during user localization. The LFEM then creates a feature component on the basis of these data.

**3.4. ULRM.** Figure 5 shows the ULRM. The module is based on the space recognition features generated by the LFEM for training. This module also provides a user interface with an application level. The location data classifier classifies the current user's location features. To perform this task, the location data classifier is trained on a set of environmental data. The ULRM input is processed using the user location recognizer classification based on the received environmental data to provide an output. The ULRM performs a location test and training using the location data classifier.

In the first recognition test, the feature data can be sent to the location data classifier through the user location recognizer. The recognizer uses k-NN as the location data classifier. The k-NN classification was developed in view of the need for performing discriminant analysis when reliable parametric estimates of probability densities are not available. This classifier is traditionally based on the Euclidean distance between a test sample and specified training samples. k-NN is an algorithm for measuring the distance between bound objects from the value of  $K$ , which is the Euclidean distance. Finally, the result is returned to the user location recognizer through the ULRM interface and is displayed on the recognizer. ULRMs transfer training data from the user location trainer to the location data classifier. Finally, the data are displayed on the ULRM interface.

**3.5. Location Feature Extraction and Recognition Procedure.** Figure 6 shows location feature extraction and recognition procedure. The LDCM can sense environmental data and transfer them to the base station. The base station has the LFEM and the ULRM. The upper part of Figure 6 shows a method for feature extraction, which is the function of the LFEM (see Section 3.3).

The LFEM can extract features. For example, assume that we apply PCA to the collected sound and light data. The data are then analyzed using PSD. In spectrum analysis, PSD of data whose analysis element is limitless is used. Fourier transform is used to express limitless data as power per hertz. This representation is often simply called the

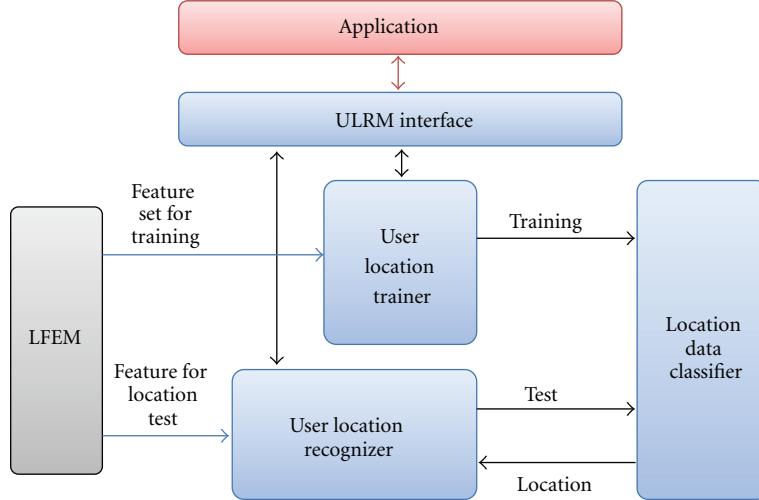


FIGURE 5: ULRM.

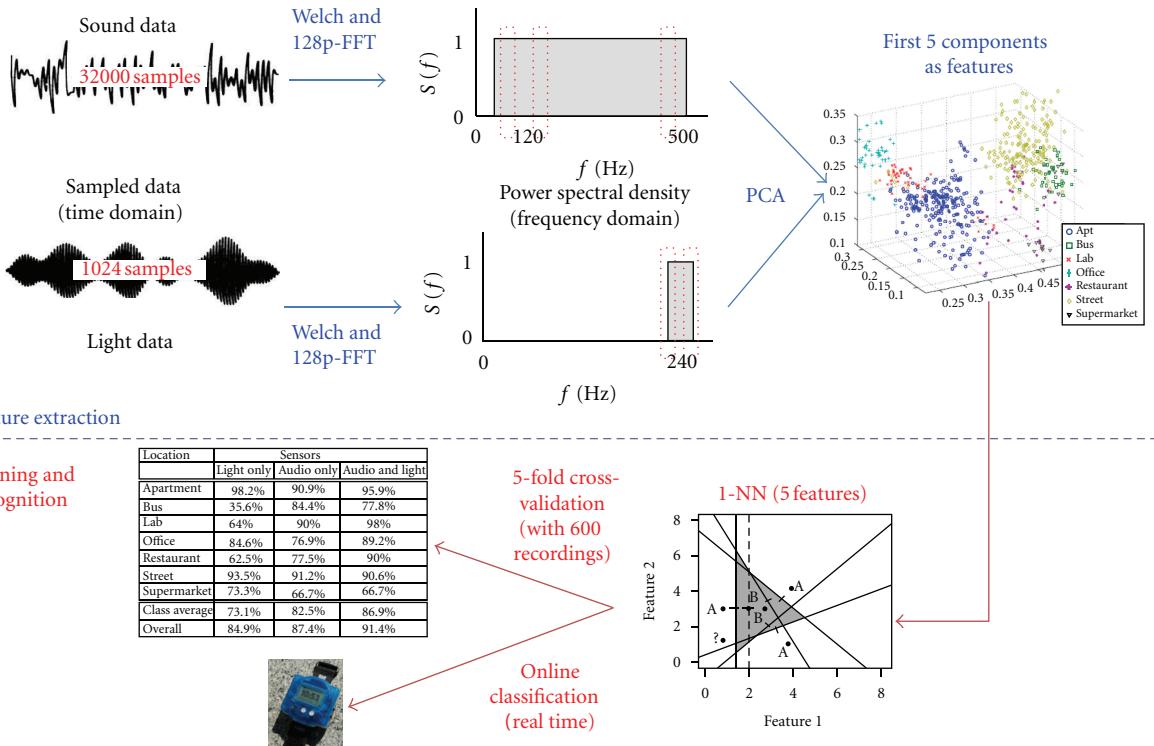


FIGURE 6: Location feature extraction and recognition procedure.

power spectrum of the data. Intuitively, the spectral density measures the frequency content of a stochastic process and helps identify periodicities. Thus, different extraction methods are applied to different types of data. In addition, PCA is applied to data for high-speed analysis. PCA is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that

the first principal component has the largest possible variance, and each succeeding component has the highest variance possible under the constraint that it is orthogonal to the preceding components. The principal components are guaranteed to be independent only if the dataset is jointly and normally distributed. PCA is sensitive to the relative scaling of the original variables. We perform PCA on and partial characteristics from the sound and light data.

The lower part of Figure 6 shows the method used for location recognition, which is the function of the ULRM. The ULRM either recognizes a user location or trains

```

@relation usn

@attribute temp numeric
@attribute hum numeric
@attribute light numeric
@attribute sound numeric
@attribute class {lobby, lab, toilet, cafeteria, bank, bookstore}

@data
#temp value, #hum value, #light value, #sound value, #class name

```

FIGURE 7: ARFF format of dataset file.

TABLE 1: Implementation environments.

Operating system	(i) Location recognition system: Windows Vista (ii) Wireless sensor node: TinyOS
Programming language	(i) Location recognition system: Java (ii) Wireless sensor node: nesC
Software tools	(i) Location data feature extraction: MATLAB (ii) Location data classification: Weka
Hardware	(i) Location recognition system: Intel 2.0 Hz PC (ii) Wireless sensor node: Hmote2420

user location data. The training element uses  $K$ -fold cross-validation and k-NN methods.

## 4. Implementation and Experiments

**4.1. Implementation Environments.** Various software and hardware tools are used in our system. Table 1 shows the implementation environments. The operating system used for the location recognition system, which is coded in Java, is Microsoft Windows Vista. The wireless sensor is developed using TinyOS. We created a wireless sensor node using Hmote2420 and nesC. We used nesC in the TinyOS environment in order to use the Hmote2420 wireless network system. The operating systems and programming tools are described in the software section, while the hardware specifications of the sensor and the computer are presented in the hardware section.

The Hmote2420 sensor and TinyOS were used in the LCDM. Hmote2420 was used to collect environmental data and information at the base station. TinyOS was used to deliver the collected data into base station. In addition to the LFEM, we used a computer, a sensor node, a Java platform, and MATLAB to extract features from the collected data. The ULRM used the Java platform to show the recognized user's position, which was determined from the collected features. In addition, the k-NN algorithm was used for location recognition.

Table 2 shows the information related to sampling of environmental data. These sampled data were extracted using MATLAB, which was also used to convert the data to the ARFF format used by Weka.



FIGURE 8: Environments considered in the experiments.

**4.2. Environmental Dataset Generation.** The format of environmental datasets used in this study was ARFF. Temperature, humidity, light, and sound data were used to build training datasets, as explained in Section 4.2. The reason why we have used light, sound, temperature, and humidity is that they are the main physical parameters that characterize a place.

Feature extraction from a dataset involves different processes, depending on the sampling rate of the dataset (see Figure 4). High-frequency data, such as light and sound data, may lead to the training and classification process being slow, because the size of the dataset is too large. Therefore, to reduce the number of feature components, PCA was used to extract the most representative feature components for each location. Before the feature extraction procedure, high-frequency environmental datasets are transformed into the frequency domain using FFT.

On the other hand, environmental data sampled at a low frequency, such as temperature and humidity data, can be directly used as representative features for each location. Therefore, PCA need not be performed on these datasets. Figure 7 shows the format of ARFF training dataset files.

**4.3. Experimental Method.** In our experiments, data were collected from different places in Konkuk University (Figure 8): a laboratory, a toilet, the lobby of the New Millennium Hall, a bank, a bookstore, and a cafeteria (the last three are located in the student union building). The experiments are explained below.

First, we collected 100 datasets from each place by using the sensor. A total of 600 datasets were collected from

TABLE 2: Environmental data collection methods.

	Rec/sec	Sampling rate (Hz)	Duration (sec)	samples/rec	Type of sampler
Temperature	5/10	1	4	4	Std.
Humidity	5/10	1	4	4	Std.
Light	5/10	2048	0.5	1024	High Freq.
Sound	5/10	8000	4	32000	High Freq.

TABLE 3: Offline localization experimental results.

Test data	Classified					
	Lobby	Laboratory	Toilet	Cafeteria	Bank	Bookstore
Lobby	<b>91</b>	0	1	0	8	0
Laboratory	0	<b>99</b>	0	0	0	1
Toilet	1	0	<b>94</b>	0	5	0
Cafeteria	0	0	0	<b>99</b>	1	0
Bank	4	0	2	2	<b>92</b>	0
Bookstore	0	3	0	0	0	<b>97</b>

the six locations. Second, the collected data were classified into high- and low-frequency data. The classified data were extracted using the feature extraction method of MATLAB. The extracted data were then converted into formats compatible with Weka. Next, ten more datasets were collected at the same time and at the same locations. Finally, our system used the collected data to recognize user locations.

**4.4. Results and Discussion.** After training the localization classifier, we collected 10 additional feature datasets from different places at each location to test the classifier. The sensor's location was then identified using the 10 datasets.

The average localization accuracy ( $A_{ave}$ ) was calculated with formula (1), where  $T_l$  denotes the set of all the datasets collected at location  $l$ ,  $TC_l$  is a correctly classified dataset for location  $l$  ( $TC_l \subset T_l$ ), and  $L$  is the number of locations considered in the localization experiments:

$$A_{ave} = \frac{\sum_{1 \leq l \leq L} |TC_l| / |T_l|}{L}. \quad (1)$$

Table 3 shows the confusion matrix for the test results. The 3-NN classification method with 20-fold cross-validation was used in the experiments. As shown in the matrix, the average localization accuracy was about 95.3%. This table shows that the highest levels of recognition were achieved for the laboratory and cafeteria.

In the table, the correct location data are shown in bold font. High localization accuracy is achieved for the laboratory and cafeteria data because of the correct classification of features. This implies that a high localization accuracy will be obtained in places where the features are well separated. Errors in recognition occasionally occur in the case of the lobby and bank. This implies that these two environments are similar in temperature, humidity, light, and sound.

Table 4 shows the real-time localization accuracy. In an experiment, the average localization accuracy of real-time location recognition was 82.2%. The highest localization accuracy was achieved for the toilet environment. On the other hand, the bookstore showed the lowest localization

TABLE 4: Real-time localization experimental results.

Location	localization accuracy
Laboratory	76.7%
Lobby	83.3%
Toilet	100%
Cafeteria	86.7%
Bank	93.3%
Bookstore	53.3%
Average	82.2%

accuracy because the indoor light data for it are similar to those for the lobby.

The classifier confused the bookstore with the lobby. This occurred because both the locations have similar light and temperature conditions. However, in the case of the toilet, because of the high humidity, the recognition results showed high localization accuracy. Finally, we can improve the localization performance of our system further by using additional types of environmental data, especially for environments with similar conditions with regard to temperature, humidity, light, and sound.

## 5. Conclusion

In this paper, we have proposed a novel location recognition method for wireless sensor nodes. The method involves the classification of environmental data features using the k-NN localization data classifier. We performed localization experiments in an actual test environment by using the proposed method. The experimental results indicated high localization accuracy. In a real-time recognition experiment, the localization accuracy was found to be 82.2%. This value indicates that environmental data can be used for the purpose of location recognition. It also shows the importance of environmental data recognition in location recognition. Our future research will focus on combining the proposed

location recognition method and other localization methods, such as RSS pattern recognition methods. Furthermore, we intend using a modified version of PCA [17] and k-NN for location feature extraction and in the classification procedures of the proposed method to improve the overall localization performance.

## Acknowledgment

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), which is funded by the Ministry of Education, Science and Technology (Grant no. 2012006817).

## References

- [1] Z. Xia and C. Chen, "A localization scheme with mobile beacon for wireless sensor networks," in *Proceedings of the 6th International Conference on ITS Telecommunications (ITST '06)*, pp. 1017–1020, June 2006.
- [2] C. H. Lim, Y. Wan, B. P. Ng, and C. M. S. See, "A real-time indoor WiFi localization system utilizing smart antennas," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, pp. 618–622, 2007.
- [3] S. H. Fang and T. N. Lin, "Indoor location system based on discriminant-adaptive neural network in IEEE 802.11 environments," *IEEE Transactions on Neural Networks*, vol. 19, no. 11, pp. 1973–1978, 2008.
- [4] J. Yim, "Comparison between RSSI-based and TOF-based indoor positioning methods," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 7, no. 2, 2012.
- [5] Y.-g. Ha, "Dynamic integration of zigbee home networks into home gateways using OSGI service registry," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 2, pp. 470–476, 2009.
- [6] C. C. Chen, D. C. Wang, and Y. M. Huang, "A novel method for unstable-signal sensor localization in smart home environments," *International Journal of Smart Home*, vol. 2, no. 3, 2008.
- [7] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM '00)*, pp. 775–784, March 2000.
- [8] A. Savvides, C. C. Han, and M. B. Strivastava, "Dynamic fine-grained localization in ad-hoc networks of sensors," in *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking*, pp. 166–179, July 2001.
- [9] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "Cricket location-support system," in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (Mobicom '00)*, pp. 32–43, Boston, Mass, USA, August 2000.
- [10] D. Niculescu and B. Nath, "Ad hoc positioning system (APS) using AOA," in *Proceedings of the 22nd Annual Joint Conference on the IEEE Computer and Communications Societies (IEEE INFOCOM '03)*, pp. 1734–1743, April 2003.
- [11] Y.-g. Ha, H. Kim, and Y. Byun, "Energy-efficient fire monitoring over cluster-based wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 460754, 11 pages, 2012.
- [12] R. Duda, P. Hart, and D. Strok, *Pattern Classification*, John Wiley & Sons, New York, NY, USA, 2001.
- [13] A. K. Chandra-Sekaran, P. Dheenathayalan, P. Weisser, C. Kunze, and W. Stork, "Empirical analysis and ranging using environment and mobility adaptive RSSI filter for patient localization during disaster management," in *Proceedings of the 5th International Conference on Networking and Services (ICNS '09)*, pp. 276–281, April 2009.
- [14] U. Maurer, A. Rowe, A. Smailagic, and D. P. Siewiorek, "eWatch: a wearable sensor and notification platform," in *Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks (BSN '06)*, pp. 142–145, April 2006.
- [15] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, 2009.
- [16] Y.-T. Kim, Y.-S. Jeong, and G.-C. Park, "Design of RSSI signal based transmit-receiving device for preventing from wasting electric power of transmit in sensor network," in *Proceedings of the 2nd International Conference on Ubiquitous Computing and Multimedia Applications (UCMA '11)*, vol. 151 of *Communications in Computer and Information Science*, pp. 331–337, 2011.
- [17] S. H. Fang and C. H. Wang, "A dynamic hybrid projection approach for improved Wi-Fi location fingerprinting," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 3, pp. 1037–1044, 2011.

