Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 2976751, 9 pages
https://doi.org/10.1155/2018/2976751

Research Article

A Novel Indoor Positioning System Using Kernel Local Discriminant Analysis in Internet-of-Things

Sajida Imran and Young-Bae Ko

Department of Computer Engineering, Ajou University, Suwon 16499, Republic of Korea

Correspondence should be addressed to Young-Bae Ko; youngko@ajou.ac.kr

Received 8 September 2017; Revised 9 January 2018; Accepted 18 January 2018; Published 19 February 2018

Academic Editor: Waleed Ejaz

WLAN based localization is a key technique of location-based services (LBS) indoors. However, the indoor environment is complex; received signal strength (RSS) is highly uncertain, multimodal, and nonlinear. The traditional location estimation methods fail to provide fair estimation accuracy under the said environment. We proposed a novel indoor positioning system that considers the nonlinear discriminative feature extraction of RSS using kernel local Fisher discriminant analysis (KLFDA). KLFDA extracts location features in a well-preserved kernelized space. In the new kernel featured space, nonlinear RSS features are characterized effectively. Along with handling of nonlinearity, KLFDA also copes well with the multimodality in the RSS data. By performing KLFDA, the discriminating information contained in RSS is reorganized and maximally extracted. Prior to feature extraction, we performed outlier detection on RSS data to remove any anomalies present in the data. Experimental results show that the proposed approach obtains higher positioning accuracy by extracting maximal discriminate location features and discarding outlying information present in the RSS data.

1. Introduction

The outstanding advancement in IoT based applications has provoked the use of location-based systems (LBS) enabling mobile devices to provide a number of personal and commercial services including, but not limited to, object tracking [1], management and security, healthcare monitoring, personal navigation, and context awareness [2]. However, due to complex indoor environment with diverse requirements, despite research efforts for more than a decade, a widely deployed indoor localization system is not yet realized, which makes indoor localization an open research issue. Several candidate technologies are researched to solve indoor positioning problem including radio frequency (RF) [3], ultrawide band (UWB) [4], ultrasonic, and sound [5], visible light [6]. Most of these technologies provide comparatively accurate positioning. However, these technologies need specialized hardware to be deployed. Also, some of the technologies require line of sight (LOS) to work well, which is difficult to realize indoors. Accurate object localization is becoming more important for Wi-Fi based devices due to the increased use of augmented reality, social networking, health care monitoring, personal tracking, and other indoor location-aware applications.

The popularity and low price of Wi-Fi network interface cards are an attractive incentive to use Wi-Fi as the basis for a localization system. WLAN or Wi-Fi based positioning techniques locate the position of virtually every Wi-Fi compatible device without installing extra software or influencing the hardware. Wi-Fi has not been designed for positioning and yet the location can be estimated by leveraging RSS values on any WLAN equipped device without using any specialized hardware. The earnest advantage of Wi-Fi based indoor positioning over other indoor wireless technologies lies in its low cost and widespread deployment. The equipment is relatively cheap, while Wi-Fi infrastructure is the widely deployed communication infrastructure which saves time and money on new installations.

WLAN based positioning techniques are categorized into the following: Time of Arrival (ToA) or Time Difference of Arrival (TDoA) [7], Angle of Arrival (AoA) [8], and fingerprinting based technique [9]. ToA and AoA based

positioning techniques are difficult to implement indoors as these techniques require LOS measurement to work. Fingerprinting-based positioning on the other hand does not require any specialized hardware and makes use of widely deployed WLAN infrastructure. The fingerprinting localization system is based on the behavior of signal propagation and information about the geometry of the building to convert RSS values into distance values. This is a challenging task as, in an indoor environment, RSS values are affected by walls and obstacles which may reflect and propagate the signals, offering a nonlinear transformation between the RSS values and the physical location. Therefore, WLAN fingerprinting-based positioning should be modeled as a nonlinear and non-Gaussian dynamic system. In such case, nonlinear classifier and feature extraction method is preferred.

To improve positioning accuracy by minimizing the above described effects, a large number of access points (APs) are usually deployed. Increased number of APs can help in distinguishing more distinct locations. However, collecting RSS from all seen APs can create too many data dimensions and lead to the curse of dimensionality problem. In the curse of dimensionality, extracting useful information from high-dimensional data becomes restricted in case of limited training data which results in inaccurate position estimations. Remember that, due to changing indoor characteristics and labor intensity, the collected RSS training data are always limited. The curse of dimensionality can be handled by mapping high-dimensional data into low-dimensional data space to preserve inherent information. The solutions for dimensionality can be categorized into two classes: AP selection [10] and feature extraction [11]. In AP selection methods, the most important APs, that is, a subset of all APs, are selected. The AP selection method is less complicated by having RSS values of stable APs only. However, this method limits the classification performance with few distinct RSS values. In feature extraction and dimensionality reduction methods like Linear Discriminant Analysis (LDA) [12] and Principal Component Analysis (PCA) [13], RSS data from all seen APs are used to increase positioning accuracy. In PCA, RSS values are transformed into principal components. It searches for directions in the data that have largest variance and subsequently projects the data onto it. PCA improves performance by reducing the noise level. However, it mainly reduces the dimensionality of data without reducing variability in the data which makes it an inappropriate method for classification. LDA, on the other hand, provide more separate embedding for classification. LDA pick a new dimension that gives maximum separation between means of projected classes and minimum variance within each projected class. Fisher discriminant analysis (FDA), a variant of LDA, discriminates location features by maximizing between-class signal scatter and minimizes within-class signal scatter. However, feature extraction methods like PCA and FDA works only with unimodal and linear data and do not consider the multimodal and nonlinear RSS data. Local FDA (LFDA) [14] can handle multimodality by maximally preserving the local structure of the data. However, if the data is multimodal and nonlinear, direct mapping of signal data into physical location leads to very inaccurate position estimations. Kernel methods [15, 16] are used to extend the linear algorithms into equivalent nonlinear space. Therefore, we applied kernel method on LFDA to convert nonlinear space into kernel space where discriminative location features are extracted from the data.

In this work, we used KLFDA to cope with the multimodality and nonlinearity in RSS fingerprint data. KLFDA preserve the multimodal structure of the nonlinear data and provide more separate embedding than LFDA. The KLFDA's advantage over other kernel-based methods lies in its computational simplicity.

In addition to nonlinearity and multimodality, we also dealt with inordinate errors called outliers in RSS data. Due to interfering indoor environment, RSS data are prone to frequent outliers. Position estimation using these outlying RSS data leads to huge positional errors. Despite its huge impact on positioning accuracy, a little attention has been given to this issue. We used Maximum Likelihood Outlier Detection (MLOD) as an outlier detection algorithm, which is an inlier-based outlier detection algorithm.

In the proposed system, we first perform outlier detection to remove any anomalies present in RSS data. Then, for feature extraction, we mapped the raw RSS vectors into kernel feature space applying KLFDA (KLFDA transformation). Location is estimated by measuring Euclidean distance between kernelized offline and online RSS data features. Through experimental results, we witness the higher position accuracy of the proposed system.

The rest of the paper is organized as follows: Section 2 describes the background study on indoor positioning. The proposed kernel localized FDA is explained in Section 3. In Section 4 evaluation and results are provided. Finally, the conclusion is presented in Section 5.

2. Background Study

WLAN fingerprinting positioning consists of two phases: offline and online phase. In the offline phase, RSS values from multiple APs are collected at different reference points (RPs), that is, fingerprints to create the radio map. Furthermore, based on the collected fingerprints, a positioning model is trained to construct the relationship between RSS signals and physical locations. In the online phase, the user's location is estimated by applying the learned model to real-time RSS samples. Use of large RSS samples can help in creating more distinct location features. However, as a result the dimensionality of data is also increased, which leads to the misled feature classifications. Dimensionality reduction techniques like FDA work well by restricting the RSS data to certain low dimensions. However, the FDA only works with unimodal and linear data. The indoor environment is complex and, due to effects like multipath propagation, the RSS data become nonlinear and multimodal. When the FDA is applied on a multimodal and nonlinear data, it will form several different clusters from a single multimodal sample. It is difficult for FDA to keep within-class scatter to a certain level. For dimensionality reduction of multimodal data, the local structure of data needs to be preserved. LFDA does not require multimodal samples to fall into a single cluster.

As a result, more degree of freedom is left for increasing separability. LFDA can preserve multimodal structure of the data better than FDA. However, LFDA also works on linear data structures only.

To deal with multimodality in the data, Sugiyama [14] proposed a dimensionality reduction method called enhanced local Fisher discriminant analysis (ELFDA) which is a localized version of FDA. It takes local structure of the data into account so that the multimodal data can be embedded appropriately. ELFDA can preserve multimodal structure of the data better than FDA and provides more separate embedding than FDA. However, ELFDA only works in linear data structures. Deng and Meng in [16] used kernel Fisher discriminant analysis (KFDA) for indoor localization. KFDA is meant to be use for classifying offline RSS measurements with online RSS measurements. However, the work does not consider any localized variant of FDA to consider multimodal data. In [17], LFDA based indoor positioning is achieved. The work considers the multimodality in data and used LFDA to cope with the problem. A cluster-based approach is used to minimize the search space to a specific cluster. However, this approach did not consider any nonlinearity present in the RSS data. The study [17] used LFDA to cope with multimodality in the data. LFDA is used to extract discriminative features from the RSS data. These discriminative features are extracted in such a way to increase between-class separability, while preserving within-class local structure of the RSS space. The generalization ability of LFDA is further enhanced using signal perturbation, which generates a higher number of representative training samples.

In Hayashi et al. [18], a fingerprinting-based Wi-Fi indoor positioning method robust against temporal fluctuations and spatial instability in Wi-Fi signals is presented. Spatial changes are coped with by splitting the environment of interest into several areas and tailoring several weak estimators to each area so that they can accurately estimate the user's position in the given area. They cope with temporal changes in the Wi-Fi signals by using random subsets of APs in weak estimators. Zheng et al. in [19] consider two problems commonly found in RSS data: unstable positioning fingerprint features and curse of dimensionality. They designed a positioning fingerprint feature using the segment similarity of Wi-Fi access points by considering both the received signal strength value and the Wi-Fi access point. Based on this designed fingerprint feature, a two-stage positioning algorithm for indoor fingerprint-based positioning is proposed. Pan et al. in [20] considered uncertainty and nonlinearity in the signal to map signal and physical space using multidimensional vector regression and kernel method. For feature extraction, kernel canonical analysis is performed to maximally correlate pairwise similarity in both signal and physical space.

In Khalamehrabadi et al. [21], a WLAN based localization scheme is proposed that consists of outlier detection and radio-map interpolation schemes. Outlier detection is mapped as augmented optimization problem. A GS-based positioning system is reformulated to cope with outliers during online phase only. The work [22] proposed a joint WLAN based localization and outlier detection scheme. The scheme consists of three phases of course localization, AP selection, and fine localization using sparse recovery algorithms. The working area is clustered into ROIs where user location is searched within a ROI. Outlier detection is performed during online phase using modified sparse recovery algorithm.

Chen and Juang in [23] proposed an outlier detection framework to cope with the outliers in data for localization. The scheme does nothing to remove the outliers and rather repairs the outliers to preserve useful information contained with outliers. Each piece of outlier data is given a confidence value based on MAD-scale score which is introduced to show the usefulness of the data. The proposed outlier detection based scheme is applied in RSS data collection and location tracking phases. Meng et al. in [24] developed a secure and robust indoor localization scheme to address the outliers in RSS data due to accidental environmental changes and access point attacks. A probabilistic region-based fingerprinting method is proposed to reduce the outlier effect and improve the localization accuracy. A three-step location sensing algorithm is proposed.

Proposed positioning system works on both multimodality and nonlinearity in RSS data. Multimodality in data is handled with localized FDA, while kernel method is used to deal with nonlinearity in RSS fingerprinting data. Furthermore, MLOD based outlier detection is performed to remove any anomalies (also called noise) in RSS data. We experienced that the use of outlier detection algorithm prior to feature extraction facilitates more accurate classification of distinguished location features. Experimental results show improved position estimation accuracy as compared to other feature extraction methods of LDA [12], PCA [13], and ELFDA [14].

3. Proposed KLFDA Positioning Algorithm

This section describes the proposed KLFDA based positioning algorithm. We first explain the RSS data collection and radio-map generation step. Then, overview of KLFDA is presented along with feature extraction in nonlinear data. Afterwards, distance calculation of RSS vector in new feature space is described to estimate the user's final position. Choice of Kernel selection is explained afterwards. Finally, outlier detection using MLOD is presented that removes any anomalies present in RSS data to make the KLFDA based positioning more efficient.

3.1. Overview of KLFDA. In offline phase of the proposed KLFDA, the following sequence of steps is followed: at first, a road map is built by collecting RSS samples at different reference points. Then, MLOD is performed to remove any anomalies present in the data. After that, for feature extraction, we first mapped the raw RSS vectors into kernel feature space by applying KLFDA. This process is referred to as KLFDA transformation.

In online phase of the proposed KLFDA, the following sequence of steps is followed: after collection of real-time RSS samples, the outlier detection and KLFDA transformation are applied in the same way as applied in offline data. Then, location is estimated by measuring Euclidean distance

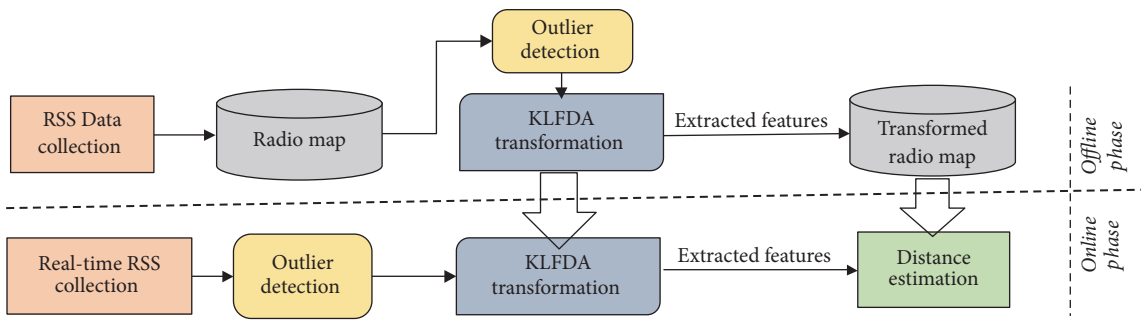

Figure 1: Proposed framework.

between kernelized offline and online RSS data features. Figure 1 shows the proposed indoor positioning framework that tackles multimodal and nonlinear RSS data.

RSS data is multimodal, non-Gaussian, and nonlinear in nature. Kernel methods implicitly map linear data into high-dimensional nonlinear feature space. KLFDA can well preserve the multimodal structure of the nonlinear data and provide more separate embedding than its nonkernel variants.

3.2. Feature Generation. FDA finds a linear combination of features that characterizes or separates two or more classes of objects or events. We used localized FDA (LFDA) to look after the multimodality and then applied kernel method on LFDA to cope with nonlinearity in RSS data (KLFDA). LFDA maximally preserve local structure of the RSS class data than simple FDA which result in better separation on multimodal data. The resulting combination is used as a linear classifier or more commonly for dimensionality reduction. To extract location feature to reduce RSS data dimensionality, the data is modeled using FDA. Let $a_i \in \mathbb{R}^m$ $(i = 1, 2, 3, \ldots, n)$ be m-dimensional RSS samples and $b_i \in \{1, 2, 3, \ldots, c\}$ are related class labels, where n is number of samples and c is the number of classes. Let n_s be the number of samples in class s:

$$\sum_{s=1}^{c} n_s = n. \tag{1}$$

Let X be the matrix of all samples:

$$X \equiv (x_1 \mid x_2 \mid \cdots \mid x_n). \tag{2}$$

Let $z_i \in \mathbb{R}^r$ $(1 \leq r \leq m)$ be the low-dimensional representations of a_i, where r is the reduced dimension. Persuasively, we consider m to be large and r to be small, but not limited to such cases. For linear dimensionality reduction, the embedded samples z_i are given as

$$z_i = T^\top x_i, \tag{3}$$

where $\top$ denotes the transpose of the RSS matrix. The goal of LFDA is to find a linear transformation that maximizes class separability in the reduced dimensional space. Let us have $S_{(b)}$ as between-class scatter and $S_{(w)}$ as within-class scatter. To formulate RSS database, the RSS vector is represented as

$$A = [a_1, a_2, \ldots, a_n], \tag{4}$$

where a_1, a_2 represent the received signal strength at point P from AP_i. The RSS vector $A = [a_1, a_2, \ldots, a_n]$ is formulated with FDA in such a way to maximize

$$\begin{aligned} j(u) &= \frac{u^\top S_{(b)} u}{u^\top S_{(w)} u}, \\ S_{(w)} &= \frac{1}{A} \sum_{l=1}^{c} \sum_{i:y_{i=l}} (x_i - u_l)(x_i - u_l)^\top, \\ S_{(b)} &= \frac{1}{A} \sum_{l=1}^{c} n_l (u_l - u)(u_l - u)^\top, \end{aligned} \tag{5}$$

where $x_i = (1/n_i) \sum_{j \in N_i} a_j$ is class centroid while $u = (1/n) \sum_{j=1}^{n} a_j$ is global centroid. $\sum_{i:y_{i=l}} \cdot$, which represents the summation over i, that is, $y_i = l, u_l$ is the mean of the samples in class l, and u is the mean of all samples. FDA's transformation matrix T maximizes the between-class scatter, while it minimizes the within-class scatter, where FDA transformation T is defined as follows:

$$T \equiv \underset{T \in R^{d \times r}}{\arg} \left[\left(T^\top S_{(b)} T \left(T^\top S_{(w)} \right)^{-1} \right)^\top \right]. \tag{6}$$

While performing classification on multimodal data, FDA restricts within-class scatter to be small which may result in combining multimodal data into a single cluster. Smaller within-class distance can restrict the increase in feature separability which could result in degraded classification ability of the FDA. Local FDA, on the other hand, works locally and does not necessarily impose the restriction of RSS samples to be close, which gives an increase in between-class distance, resulting in the extraction of more discriminating location features. List of symbols used in this article is listed in the Symbols.

3.2.1. Kernel Local Fisher Discriminant Analysis. LFDA works well to cope with multimodal linear RSS data. However, RSS data usually show nonlinear behavior, especially at points of sudden turns. Kernel methods are proven to deal well with the nonlinearity in the data. To extend LFDA to kernelized LFDA for nonlinear mapping, RSS data can be mapped to

new kernel feature space $\dot{F}$ using function ϕ. We want to maximize the function

$$j(u) = \frac{u^\top S^\phi{}_{(b)} u}{u^\top S^\phi{}_{(w)} u}, \tag{7}$$

in which now $u \in F$ and $S^\phi{}_{(b)}$ and $S^\phi{}_{(w)}$ are the respected matrices in $\dot{F}$ which is represented as

$$\begin{aligned} S^\phi{}_{(w)} &= \frac{1}{A}\sum_{l=1}^{c}\sum_{i:y_{i=l}} (x_i - u_l)(x_i - u_l)^\top, \\ S^\phi{}_{(b)} &= \frac{1}{A}\sum_{l=1}^{c} n_l (u_l - u)(u_l - u)^\top. \end{aligned} \tag{8}$$

The optimal discriminant basis vectors are obtained by solving eigenvalue problem $u^* = \arg\max(j(u))$ in (7). The kernel feature representation of corresponding RSS samples can be collected by the projection of kernel representation onto the optimal discriminant basis vectors u. The optimal value of u that maximizes $j(u)$ can be obtained by the intersection space of the null space of $S^\phi{}_{(w)}$ and the nonzero space of $S^\phi{}_{(b)}$, where nonzero space of $S^\phi{}_{(b)}$ can be procured by keeping the eigenvectors $M = [m_1 \cdots m_E]$ with the E biggest eigenvalues $[\lambda_1 \cdots \lambda_E]$. Smaller eigenvalue vectors are considered as noise as these vectors contain little discriminative information.

3.3. Choice of Kernel. In indoor positioning, kernel methods are popular to implicitly map linear data into high-dimensional nonlinear feature space. For a given kernel function, the corresponding underlying nonlinear transformation function is determined which defines the mapping from original data space to kernel induced feature space. Therefore, it can be said that the form of kernel function plays a very important role in kernel methods. In this work, the choice of kernel mainly relies on the nonlinearity and uncertainty properties of RSS. Kernel parameters are adjusted through a process called generalized-cross-validation. We are interested in kernel K for RSS examples a and b in an input feature space $X = \mathbb{R}^d$ as

$$K(a, b) = \langle \emptyset(a), \emptyset(b) \rangle, \tag{9}$$

where $\emptyset$ nonlinearly maps linear input space X into linear feature space F. Generally, kernels are categorized into stationary, locally stationary, and nonstationary kernels. Stationary kernels, sometimes called anisotropic stationary kernel, depend on lag vector that separates RSS examples a and b. At a fixed location, RSS from one AP may vary as high as 10 dB and always shows Gaussian or semi-Gaussian distribution. Among many, Gaussian kernels are popular in RSS based localization because of their good localization and smoothing ability. Moreover, Gaussian kernels are good at characterizing the uncertainty of RSS and ease in capturing the nonlinear RSS patterns in the Gaussian kernel induced space. Therefore, we chose Gaussian kernel for nonlinear mapping, and a 2D Gaussian kernel is defined as

$$G(a, b, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{a^2 + b^2}{2\pi\sigma^2}\right), \tag{10}$$

where a and b are two RSS vectors and σ determines the kernel width. σ is a similarity measure between two RSS vector samples. Also, it greatly affects shape of the kernelized feature space by controlling the sensitivity of the kernel to change. In this work we used method from [15] to determine kernel width.

3.4. Outlier Detection. Due to variability in the measurements of a dataset, there could be observations that are different from other observations present in the data. These different observations are called outliers. Outlier in a dataset are present due to a number of reasons like shadowing and multipath and variability in transmitting power of APs to manage network traffic; if traffic loads during offline and online data collection are different, then the RSS readings will certainly be different. Also, due to impermanent effects in any AP, the RSS readings may not be available in any of the offline or online phases. The described phenomena generate outliers in the RSS data which results in difference in offline and online RSS measurements. This difference can significantly influence position estimation accuracy. In this paper we adopt Maximum Likelihood Outlier Detection, which is an inlier-based outlier detection algorithm that works on detecting outliers in sample data based on some model dataset. The sample dataset (evaluation set) is organized based on degree of outlining. The degree of outlyingness is measured as

$$f = \frac{E_s}{M_s}. \tag{11}$$

The ratio is estimated by the density-ratio estimation method KL importance estimation procedure (KLIEP). KLIEP's performance is estimated through basic function $\Psi(x)$ [19]. In such case, use of Gaussian kernel is preferred which is expressed as follows:

$$r(x) = \sum_{l=1}^{n_{nu}} \emptyset_l K(x, x_l{}^{nu}), \tag{12}$$

where $K(x, x_l)$ represents Gaussian kernel. Choosing large n_{nu} needs Gaussian centers to be chosen by using all $\{x_i^{nu}\}_{i=1}^{n_{nu}}$ samples, which is computationally expensive. This problem can be solved by using subset of $\{x_i^{nu}\}_{i=1}^{n_{nu}}$ as Gaussian centers, which is expressed as follows:

$$r(x) = \sum_{l=1}^{z} \emptyset_l K(x, c_l), \tag{13}$$

where c_l is a randomly chosen template point from $\{x_i^{nu}\}_{i=1}^{n_{nu}}$ and z ($\in \{1, \ldots, n_{nu}\}$) represents a prefix number. KLFDA transforms raw RSS data into high-dimensional kernel space, which contains enhanced nonlinear discriminative location information. Moreover, it also adheres to Fisher criterion, which maximizes the between-class scatter while it minimizes within-class scatter at the same time. Distance calculation between offline and online RSS data is then applied for location estimation. It is achieved by matching the k-nearest RSS vectors through KLFDA distance measure.

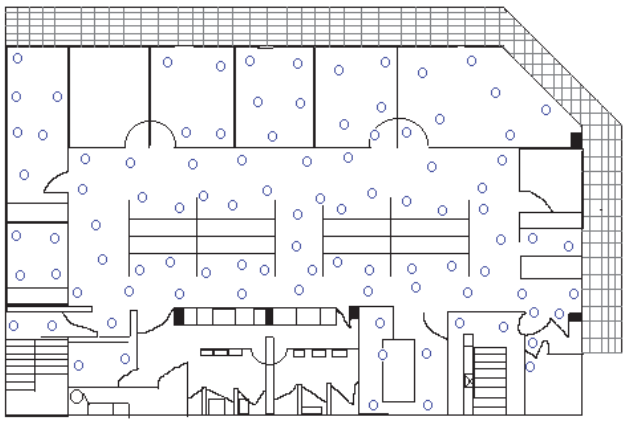

Figure 2: Experimental setup. Purple circles show the reference locations used during offline phase.

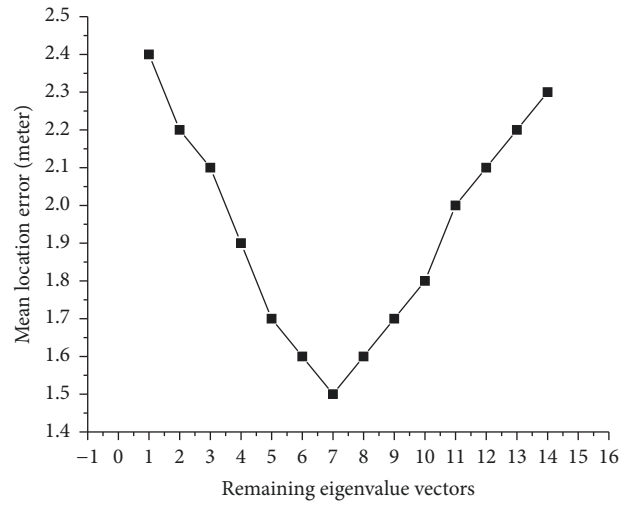

Figure 3: Effect of different number of remaining eigenvectors E on the mean location error.

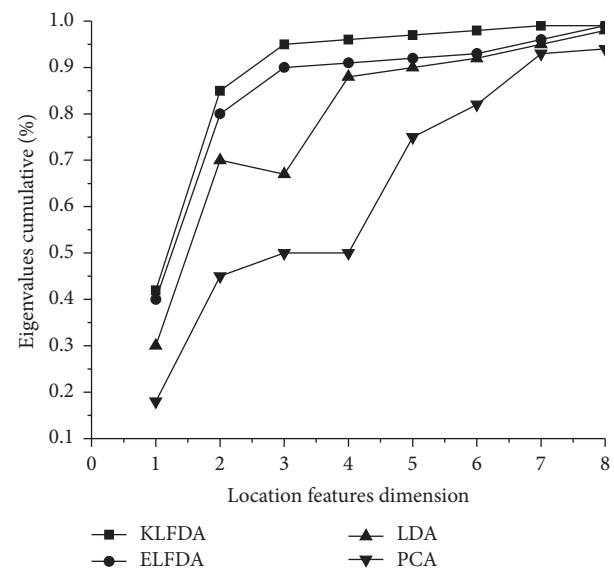

Figure 4: Cumulative eigenvalues percentage against feature dimensions.

4. Evaluation

4.1. Experimental Setup. We used KIOS RSS dataset to train and validate our positioning system. Wi-Fi RSS data is collected through extensive measurements at KIOS Research Center, Cyprus [25]. The area consists of a 560 m^2 typical office environment that includes a conference room, several open cubicles and private offices, laboratories, and corridors. Total 9 stable APs with full coverage were found to collect RSS data at the floor. Moreover, varying number of unstable APs were found at different points in corridors during RSS data collection. Five different devices were used to collect RSS data at different points. The mobile devices used for RSS data collection are an Asus PC T101MT laptop running Windows 9, a HP iPAQ hw6915 PDA with Windows Mobile, an HTC Flyer Android tablet, HTC Desire, and Samsung Nexus S. The area is divided into 201 reference points. Out of 201 reference points, 2100 RSS readings at 105 reference points are used to train the model with 20 RSS readings per reference point, while 960 RSS readings at 96 with 10 readings per reference points are used to test the model. Figure 2 shows the schematics of the experimental setup. The proposed approach is compared with PCA [12], LDA [11], and ELFDA [13]. The whole training set for radio map is divided into five parts: four parts are for tentatively building the model, while the remaining one is a validation set for evaluating the positioning performance. Each part is chosen as a validation set for one time and the validation performance is the average result of five times. The real model is set by parameter values with the found best validation performance. The parameters of the other compared algorithms are all set to the found optimal values by this method.

4.2. Analysis of Simulation Results. Maximally extracted location features can better classify individual locations. For accurate position estimation, it is important to fully extract discriminative features in RSS data. This discriminating quantity is obtained by the cumulative percentage of extracting eigenvalues compared with the remaining eigenvectors E. Figure 3 compares accuracy of different feature generation methods. It is shown in the figure that the mean location error is the lowest when E is equal to 7, which shows that 97.0% nonlinear information is retained. The remaining 3.0% is considered as noise which is discarded. We see a decrease in mean location error until within-class scatter reaches 7, and after that the mean error starts increasing. The eigenvectors with larger eigenvalues are discarded since these larger values increase the within-class scatter which can result in degraded positioning accuracy. The proposed approach can maximally extract nonlinear distinct location feature. The information that increases within-class scatter or decreases between-class scatter is considered noise. Although extremely larger between-class scatter can lead to increased uncertainty, however, marginally higher between-class scatter preserves more location information. Figure 4 shows the eigenvalues cumulative percentage against feature

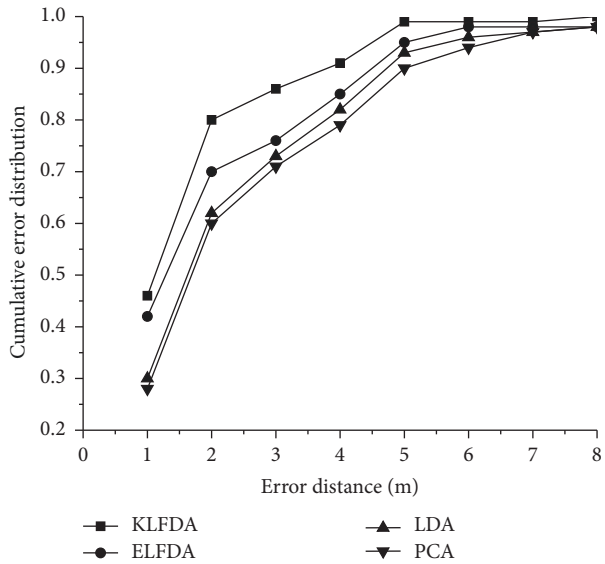

Figure 5: Accuracy comparison between different feature generation methods.

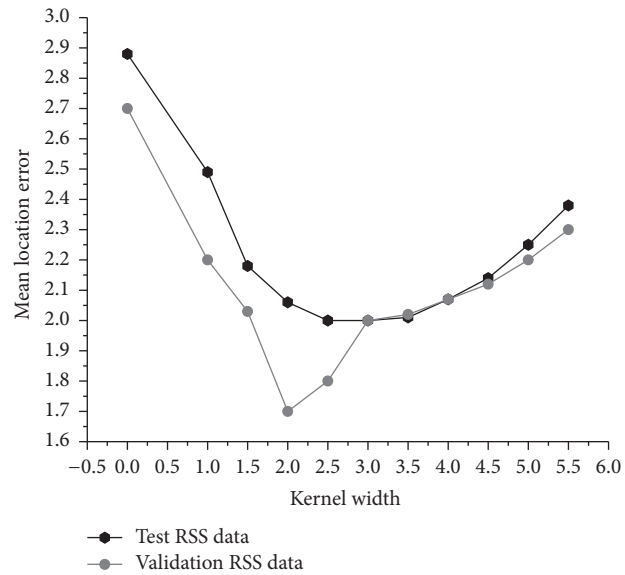

Figure 6: Kernel width versus mean location error.

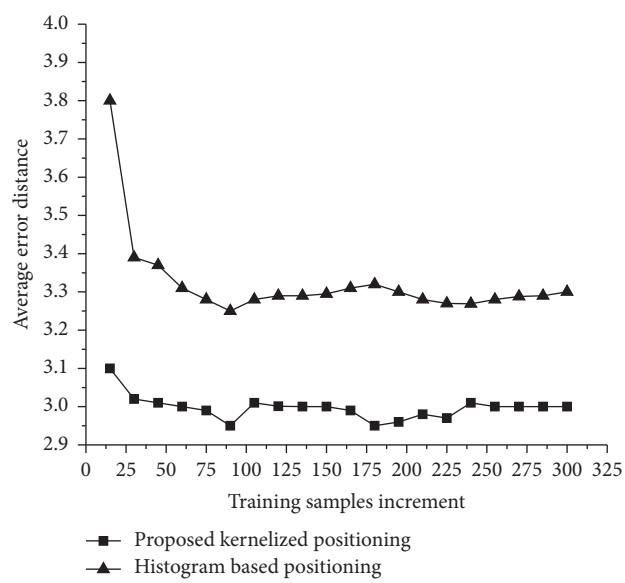

Figure 7: Number of training samples versus average error distance.

dimensions. The proposed approach is compared with ELFDA, LDA, and PCA. PCA can maximize the variance of extracted features, whereas ELFDA and LDA work on maximizing the between-class separability. However, we see in the figure that the proposed KLFDA preserves local within-class structure without being effected by nonlinearity in RSS data. The first three dimensions of KLFDA already have a cumulative percentage of more than 90% as compared with ELFDA, LDA, and PCA which indicates its strongest discriminative power. Figure 5 shows the average testing positioning accuracy that indicates the cumulative positioning error distribution. Accuracy is also compared between PCA, LDA, ELFDA, and proposed KLFDA. Error distance is the Euclidean distance between true and estimated location coordinates. The cumulative probability distribution of error distance is used as positioning accuracy. From the figure we see that, within error distance of 2 m, the accuracy of the proposed KLFDA is 80%, while accuracy of ELFDA, LDA, and PCA is 70%, 62%, and 60%, respectively.

The shape of the kernelized feature space is mainly defined by σ which is a parameter to restraint the kernel width. Determining the parameter σ is a complex task. Figure 6 shows the mean location error for different kernel widths because of constraints in computations and facing the variances between offline and online measurements. Near-optimal kernel width is observed in validation data for the test data, which encourages urging for parameter tuning on valid validation data. At σ value 2, the larger gap between validation and test data is because of larger variance between offline and online measurements.

While computing distance, kernel-based system's computation complexity is increased. Smaller test samples result in less calibration time in offline phase. Figure 7 shows the effect of number of training samples on distance error. Comparison is made between well-known histogram positioning method [26] and the proposed scheme. Histogram is a popular technique of estimating PDF in a nonparametric way. The figure shows the number of samples collected at each reference point. We see that histogram method shows larger variance with number of samples. This is because a substantial number of samples are required at each training location to produce satisfactory results in histogram based methods. The proposed approach on the other hand shows almost consistent results with the increased number of samples, which shows the invariability of the proposed scheme against number of samples. We also see that both the systems show highest accuracy with almost 90 training samples. Figure 8 shows the mean location error observed with increasing number of outliers. It is shown in the figure that KLFDA with outlier detection has mean location error less than that of KLFDA where there is no outlier detection. The results

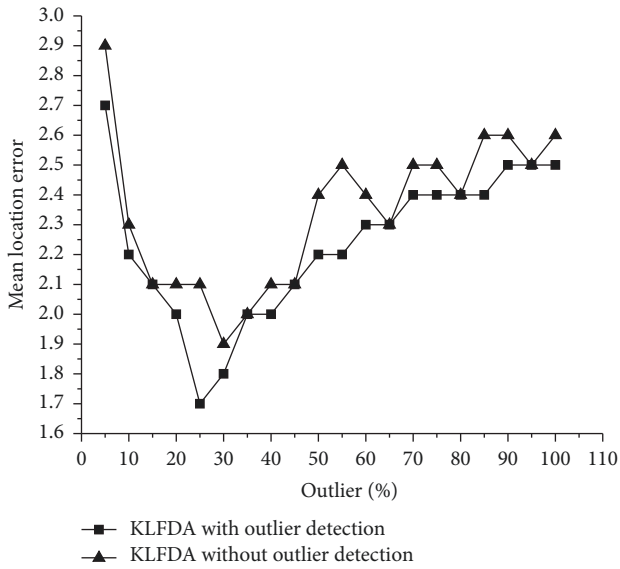

Figure 8: Increase in outliers' versus mean location error.

of Figure 8 acknowledge usefulness of outlier detection in proposed scheme.

5. Conclusion and Future Work

Accuracy of fingerprinting-based localization increases with higher number of collected WLAN signals. Due to complex environment, the collected RSS values are highly uncertain, multimodal, and nonlinear in nature, which makes the positioning system perform inaccurately. We performed nonlinear discriminative feature extraction of RSS using KLFDA which extracts location features in a kernel space where the nonlinear RSS features are well characterized and captured. Along with handling nonlinearity in data, KLFDA copes well with the multimodality in the RSS data. By performing KLFDA, the discriminative information contained in RSS is reorganized and maximally extracted. Outlier detection on RSS data removes anomalies in the data which makes KLFDA based location system efficiently perform position estimation. The proposed approach obtains higher accuracy by maximally extracting discriminative features in nonlinear space. We aim to extend this work by performing extensive experiments to thoroughly analyze further discriminative features and effects of outlier detection on feature extraction efficiency.

Symbols

a_i: m-dimensional RSS samples
b_i: Class labels
n: Number of samples
c: Number of classes
n_s: Number of samples in class s
X: Matrix of all samples
z_i: Low-dimensional representation of a_i
r: Reduced dimension
$S_{(b)}$: Between-class scatter
$S^{\Phi}{}_{(b)}$: Between-class scatter in kernelized feature space
$S_{(w)}$: Within-class scatter
$S^{\Phi}{}_{(w)}$: Within-class scatter in kernelized feature space
T: FDA transformation matrix
$\dot{F}$: Kernel feature space
ϕ: Function to map RSS vectors to new kernel feature space
E_s: Sample/evaluation dataset
M_s: Model dataset.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01059049).

References

[1] S. Imran and Y.-B. Ko, "A continuous object boundary detection and tracking scheme for failure-prone sensor networks," *Sensors*, vol. 17, no. 2, 2017.

[2] J. Chon and H. Cha, "LifeMap: a smartphone-based context provider for location-based services," *IEEE Pervasive Computing*, vol. 10, no. 2, pp. 58–67, 2011.

[3] S. He and S. G. Chan, "Wi-Fi fingerprint-based indoor positioning: recent advances and comparisons," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 466–490, 2016.

[4] S. Gezici, Z. Tian, G. B. Giannakis et al., "Localization via ultra-wideband radios: a look at positioning aspects of future sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 70–84, 2005.

[5] F. Ijaz, H. K. Yang, A. W. Ahmad, and C. Lee, "Indoor positioning: a review of indoor ultrasonic positioning systems," in *Proceedings of the 15th International Conference on Advanced Communication Technology: Smart Services with Internet of Things!, ICACT 2013*, pp. 1146–1150, 2013.

[6] Z. Yang, Z. Wang, J. Zhang, C. Huang, and Q. Zhang, "Wearables can afford: light-weight indoor positioning with visible light," in *Proceedings of the the 13th Annual International Conference*, pp. 317–330, Florence, Italy, 2015.

[7] Y. Qi, H. Kobayashi, and H. Suda, "Analysis of wireless geolocation in a non-line-of-sight environment," *IEEE Transactions on Wireless Communications*, vol. 5, no. 2, pp. 672–681, 2006.

[8] C. Wong, R. Klukas, and G. Messier, "Using WLAN infrastructure for angle-of-arrival indoor user location," in *Proceedings of the 68th Semi-Annual IEEE Vehicular Technology (VTC '08)*, pp. 1–5, IEEE, 2008.

[9] S. Chan and G. Sohn, "Indoor localization using wi-fi based fingerprinting and trilateration techiques for LBS applications," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVIII-4/C26, pp. 1–5, 2012.

[10] H. Miao, Z. Wang, J. Wang, L. Zhang, and L. Zhengfeng, "A novel access point selection strategy for indoor location with Wi-Fi," in *Proceedings of the 26th Chinese Control and Decision Conference, CCDC 2014*, pp. 5260–5265, Changsha, China, 2014.

[11] Z.-A. Deng, Y.-B. Xu, and L. Ma, "Indoor positioning via nonlinear discriminative feature extraction in wireless local area network," *Computer Communications*, vol. 35, no. 6, pp. 738–747, 2012.

[12] C. H. Park and H. Park, "A comparison of generalized linear discriminant analysis algorithms," *Pattern Recognition*, vol. 41, no. 3, pp. 1083–1097, 2008.

[13] S.-H. Fang and T. Lin, "Principal component localization in indoor wlan environments," *IEEE Transactions on Mobile Computing*, vol. 11, no. 1, pp. 100–110, 2012.

[14] M. Sugiyama, "Local fisher discriminant analysis for supervised dimensionality reduction," in *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, pp. 905–912, ACM, 2006.

[15] A. Kushki, K. N. Plataniotis, and A. N. Venetsanopoulos, "Kernel-based positioning in wireless local area networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 689–705, 2007.

[16] Y. Xu, Z. Deng, and W. Meng, "An indoor positioning algorithm with kernel direct discriminant analysis," in *Proceedings of the 53rd IEEE Global Communications Conference (GLOBECOM '10)*, pp. 1–5, Miami, Fla, USA, December 2010.

[17] Z.-A. Deng, Y. Xu, and L. Chen, "Localized local fisher discriminant analysis for indoor positioning in wireless local area network," in *Proceedings of the 2013 IEEE Wireless Communications and Networking Conference, WCNC 2013*, pp. 4795–4799, Shanghai, China, 2013.

[18] T. Hayashi, D. Taniuchi, J. Korpela, and T. Maekawa, "Spatio-temporal adaptive indoor positioning using an ensemble approach," *Pervasive and Mobile Computing*, vol. 41, pp. 319–332, 2017.

[19] Z. Zheng, Y. Chen, T. He, L. Sun, and D. Chen, "Feature learning for fingerprint-based positioning in indoor environment," *International Journal of Distributed Sensor Networks*, vol. 2015, Article ID 452590, 2015.

[20] J. J. Pan, J. T. Kwok, Q. Yang, and Y. Chen, "Multidimensional vector regression for accurate and low-cost location estimation in pervasive computing," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 9, pp. 1181–1193, 2006.

[21] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Structured group sparsity: a novel indoor wlan localization, outlier detection, and radio map interpolation scheme," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 6498–6510, 2017.

[22] A. Khalajmehrabadi, N. Gatsis, D. J. Pack, and D. Akopian, "A Joint Indoor WLAN Localization and Outlier Detection Scheme Using LASSO and Elastic-Net Optimization Techniques," *IEEE Transactions on Mobile Computing*, vol. 16, no. 8, pp. 2079–2092, 2017.

[23] Y.-C. Chen and J.-C. Juang, "Outlier-detection-based indoor localization system for wireless sensor networks," *International Journal of Navigation and Observation*, vol. 2012, Article ID 961785, 2012.

[24] W. Meng, W. Xiao, W. Ni, and L. Xie, "Secure and robust Wi-Fi fingerprinting indoor localization," in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '11)*, pp. 1–7, IEEE, Guimarães, Portugal, 2011.

[25] C. G. P. Christos Laoudias and R. Piché, *KIOS WiFi RSS Dataset*, https://www.researchgate.net/publication/256482916_KIOS_WiFi_RSS_dataset.

[26] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen, "A probabilistic approach to WLAN user location estimation," *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 155–164, 2002.