IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 11, 2019, accepted April 12, 2019, date of publication April 23, 2019, date of current version April 30, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2911601

Harvesting Indoor Positioning Accuracy by Exploring Multiple Features From Received Signal Strength Vector

MUHAMMAD USMAN ALI[1], SOOJUNG HUR[1], SANGJOON PARK[2], AND YONGWAN PARK[1]
[1]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, South Korea
[2]Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea

Corresponding author: Yongwan Park (ywpark@yu.ac.kr)

This work was supported in part by the (Ministry of Science, ICT) (MSIT) Korea, under the Information Technology Research Center (ITRC) Support Program (IITP-2018-2016-0-00313) supervised by the Institute for Information and Communication Technology Promotion (IITP), in part by the Basic Science Research Program, through the National Research Foundation of Korea (NRF) founded by the Ministry of Science, ICT and Future Planning under Grant 2017R1E1A1A01074345, and in part by the ICT R&D Program of MSIP/IITP. [2017-0-00543, Development of Precise Positioning Technology for the Enhancement of Pedestrian's Position/Spatial Cognition and Sports Competition Analysis.]

ABSTRACT The development of an indoor location information system using ubiquitous resources available in the environment is a challenging problem in the field of Geo-Location technologies, these days. Therefore, instead of relying on a single resource, the fusion of location information from multiple resources into an indoor positioning system (IPS) becomes important. The IPS in which information from multiple sources such as Wi-Fi, geomagnetism, and motion sensors is fused to harvest the next level of accuracy is commonly known as hybrid IPS. The initial estimate of the position with high accuracy is very critical for the hybrid IPS. Wi-Fi fingerprinting is one of the potential candidates for providing the initial position in such systems, whereas due to the multipath, absorption, and fading characteristics of the indoor environment, the accuracy of the Wi-Fi fingerprinting techniques is limited. Many algorithms and techniques have been proposed to improve the accuracy of Wi-Fi-based IPSs. However, most of the solution requires high computing resources and specialized hardware. This article proposes an empirical approach in which the important features present in the received signal strength vector (RSSV) of the Wi-Fi device are selected to exploit the similarity measure and index order of the Access Points (APs). The experimental results show that these features make it possible to avoid long distances outliers and to improve the positioning accuracy of the Wi-Fi fingerprinting technique without the use of specialized hardware.

INDEX TERMS Indoor positioning, IPS, particle swarm optimization, PSO, received signal strength, RSSI, Wi-Fi fingerprinting.

I. INTRODUCTION

The indoor location information of a mobile node using only ubiquitous resources available in an environment is one of the popular topics nowadays in the field of location-based technologies. Global Navigation Satellite Systems (GNSS), such as the Global Positioning System (GPS), are limited only for outdoor location-based services (OLBS) because satellite signals are heavily attenuated in indoor environments because of multipath propagation, absorption, fading effects, etc. become unreliable for the purpose of positioning. As a result, many indoor positioning systems (IPS) solutions have been proposed to enable ILBS (Indoor Location Based Services) services in public and corporate environments (e.g. hospitals, airports, shopping centers). ILBSs help in resource management, including optimal deployment, tracking, and monitoring of resources. IPS solutions are classified as infrastructure-based and infrastructure-free technologies. Infrastructure-based solutions mainly include radio frequency identifier (RFID), RF sensor, Bluetooth, and UltraWideBand (UWB) techniques as resources, while infrastructure-free solutions are typically based on

The associate editor coordinating the review of this manuscript and approving it for publication was Sajjad A. Madani.

Wi-Fi, magnetic field, motion sensors (inertial measurement units (IMU)), and camera vision techniques [1]. Infrastructure-free solutions are preferred because of their low cost and ready-to-use point of view over infrastructure-based solutions that require expensive pre-installation and configuration of specialized hardware in the deployment environment. In recent trends in IPS research, instead of relying only on a single resource, hybrid systems are proposed in which data from multiple sensors are fused (e.g., Wi-Fi, magnetometer, IMU) [2], [3]. In multiple sensor-based solutions, Wi-Fi-based location information is normally used at a global level, geomagnetism-based positioning at a local level and motion sensor-based pedestrian dead reckoning (PDR) helps to further improve accuracy. [4], [5]. As a global position resolver, Wi-Fi positioning also plays an important role as initial position provider [6]. The accurate initial position is very essential for both geomagnetism and motion sensor positioning techniques.

Wi-Fi-based techniques are generally divided into calibration-based and calibration-free techniques [7]. An indoor environment is hybrid, due to the coexistence of the line of sight (LOS) and the non-line-of-sight (NLOS) cases. This complexity in an indoor environment restricts the use of calibration-free techniques such as trilateration or triangulation. Therefore, IPS solutions using time of arrival (TOA) and time difference of arrival (TDOA) [8] techniques exhibit multipath propagation errors. In addition, time synchronization of the receiver/sender side and the measurement accuracy of the short-range flight time make them more challenging, while the angle of arrival (AOA) technique requires complex hardware for the calculation of angle. Position estimation using the value of the received signal strength indicator (RSSI) is a solution that avoids the aforementioned time synchronization problems. The Wi-Fi fingerprinting is a calibration-based technique that uses the RSSI values received at a particular location. Wi-Fi fingerprinting does not require map information and location of APs deployed in the target environment [9]. In contrast, a calibration-free path loss model technique using map and APs location information was proposed in [10] to avoid the laborious task of site survey required for Wi-Fi fingerprinting calibration database. Similarly, several other variants of fingerprinting techniques have been proposed so far, but the major challenge remains the constant accuracy in estimating the position using an unreliable measurement of the RSSI index [11], [12]. PhaseFi is based on deep learning and claims to be able to provide accurate location estimation using phase and amplitude information from sub-carriers' channel state information (CSI) [13]. This improvement in accuracy is due to the use of less attenuated factors in the CSI of the subcarriers compared to RSSI in propagation in indoor environments. In constructing the fingerprint database (DB) of phase and amplitude information CSI, access at the physical layer level to the receiving device is required. Similarly, in [14] author presents a CSI slicing technique that looks for a factor to maximize the match of power delay profiles collected from many CSIs obtained from multiple bands. CSI Splicer evaluate the derived power delay profile by estimating the distance between the sender and the receiver and corrects the ranging error for high resolution positioning purpose. Currently, some sophisticated communication devices can only provide information from multiple subchannels using the MIMO-OFDM transmission technology, which is currently only practical through third-party hardware. ArrayTrack [15] provides an estimate of AoA-based location using an antenna grid with commodity APs. By increasing the number of antennas, multiple Array Track APs overhear the transmission and calculate AoA information from the information transmitted by the client. Moreover, thanks to the AoA technique used in ArrayTrack, it is perfectly suited to low density AP environments. Recently, the addition of Fine Time Measurement (FMT) is a very important development of IPS techniques, with FTM support included in the 801.11AC update. FMT is another direction that claims to provide positioning accuracy inside the cm level range. In [16], the author carried out a complete analysis of the FMT technology and examined its evaluation with an Open Platform which provides the time of flight measurement and the range calibration facility.

At first, the proposed algorithms were tested in an office site that is regarded as an ideal environment, the accuracy was less than 2 meters and the positioning accuracy achieved only with Wi-Fi seems sufficient for a particular requirement. However, while allowing positioning on public sites, the real challenge of indoor positioning is when the average accuracy is very low and positioning with a single resource does not provide sufficient accuracy. Therefore, in public places, several resources are needed to work together to provide a certain level of precision in estimating the position. Another challenge is to improve the accuracy of fingerprinting without the use of special hardware resources or using only mobile devices that use simple transmitters which does not provide easy access to the physical layer. In this study, some of the features of RSSI that exploit the similarity measure and index order of APs in RSSV are examined to improve the accuracy of the fingerprinting technique without using specialized hardware. The proposed solution combines several calculated parameters from the RSSI list received at a particular location by a receiving node and searches for the best match between multiple candidate locations with a minimum distance error. To avoid confusion, the measure of RSSI level similarity between RSSV online APs and RSSV of DB is represented by RMSE (root mean squared error), while the term Euclidean distance is used to represent the geometric distance.

The rest of this article is structured as follows. In Section II, the problem statement is formulated and related techniques are discussed in Section III. Section IV provides details on our proposed system. The operating details of the fingerprint matching algorithm used in our system are discussed in Section V. Section VI covers the experimental setup and the results are presented in Section VII.

II. PROBLEM STATEMENT

An IPS solution based on single resource information can provide location accuracy up to a certain level, while a hybrid solution using information fusion from multiple resources can achieve greater location accuracy. This higher degree of location accuracy is achieved by compensating the limitation of one technique with the strength of another technique. For example, the PDR exhibits low error in position estimation for short distance traveled by the user. However, this error increases over time due to the accumulation of drift in the estimation of the heading angle at each step. This drift is due to the noise present in the technology of the sensors of the microelectromechanical systems (MEMS) [17]. Conversely, in Wi-Fi fingerprinting, the short-range error ratio traveled by the target is high compared to the long distance due to the absence of drift element in this technique. In addition, the uncertainty of resource availability in real-time scenarios is a significant problem that requires the use of Wi-Fi fingerprinting techniques, PDR, and geomagnetism used in parallel or alternatively in a hybrid IPS system to mitigate such unusual situations in the position estimate. [4]. Furthermore, accurate starting position information is very important for PDR and geomagnetism techniques to control drift error in PDRs and to minimize the geomagnetism search space [18]. Therefore, Wi-Fi based positioning plays a key role in most hybrid IPS systems as a technique to provide a starting position for other techniques. However, because of the noise inherent in radio cards, Wi-Fi positioning can lead to long-distance errors, which can lead an IPS hybrid system to an unstable convergence state. In addition, the accuracy of the Wi-Fi positioning as the initial position provider will guarantee a bounded maximum error and the accuracy probability of the hybrid IPS. As discussed in previous section Wi-Fi-based solutions that provide a high level of accuracy by exploiting the CSI from the physical layer. However, access to CSI information is currently only available on Wi-Fi devices that support the MIMO-OFDM Wi-Fi technology. Therefore, the problem is how to improve the accuracy and reduce the maximum error of Wi-Fi fingerprinting techniques using conventional mobile devices instead of using expensive specialized hardware.

III. RELATED WORK

Wi-Fi fingerprinting normally operate in two phases: an offline phase for the generation of a fingerprint database in which RSSI vectors (RSSV) received at particular reference points are stored in the database with accurate location information through an environmental survey. While in the online phase, the location estimate is made when the target node sends a real-time list of the MAC addresses and the corresponding RSSI values in pairs of visible APs received at an unknown position in the Wi-Fi network environment. The system searches for the fingerprint of the best matching candidate in the database and declares the corresponding reference point of the best-identified candidate as the target location. In classical fingerprinting techniques, the matching is performed by calculating the root mean squared error (RMSE) (or root mean square deviation (RMSD)) between RSSI level of matched APs in current scan and all candidate reference points in the fingerprinting DB and then the best (1-NN) or the k best matches (k-NN) are used to calculate the position of target node [19]. The enhanced weighted K-nearest neighbor (EWKNN) algorithm improves accuracy over the simple k-NN approach by selecting the number of neighbors taken into account at run time and giving high weight to candidate positions with fewer RSSI correspondence errors. [20]. Whereas in rank-based approach, the order of APs in the RSSV is considered invariant with respect to bias and scaling; the algorithm finds the candidate reference point by comparing the ranks assigned to the APs in a received RSSV with the ranks of APs in a scan vector stored in the DB with minimum difference [21]. A very important study presented in [22] discusses more the 50 different similarity function and their performance with respect to accuracy. In addition, the need to define a threshold for the elimination of weak RSSI APs is also studied. The author has considered a single site database with multiple floors and also provides the best similarity configuration for k-NN based fingerprinting.

IV. PROPOSED SOLUTION

The proposed algorithm is essentially based on the following principle: instead of relying solely on the RSS similarity measure, the system takes into account the other features available in the RSSV format and their weight in efficiency to identify the candidate positions with a high accuracy and robustness. The measure of similarity of the RSSI level with respect to the number of matched APs and their order with respect to the RSSI level in the RSSV are important measures for identifying candidate positions with respect to spatial confinement. Similarly, there are several features available in the RSSV that can be exploited to bring high accuracy in finding the best candidate from the fingerprint database. The parameter selection in the proposed solution has many dimensions, the Selection of K value and calculation of kRMSE parameters are selected to control the effects of the Wi-Fi environment [19], another parameter named as centroid distance is used to exploit the spatial distribution of candidate reference point around the ground truth [20], whereas to control the diversity effect of different mobile devices, the entropy parameter is adopted from ranked based approach [21]. Furthermore, these features are combined to compensate for the weakness of one with the strength of others. In addition, a heuristic optimization approach is used to calculate the efficiency of each parameter for a particular measurement of accuracy. Therefore, the proposed fingerprint matching algorithm is more efficient and robust than the algorithms described in the previous section. The algorithm proposed in its fingerprint matching process with the conventional RSSI similarity measure also calculates the order difference of the APs in RSSV with respect to the signal strength at a particular reference point, in order to find the best

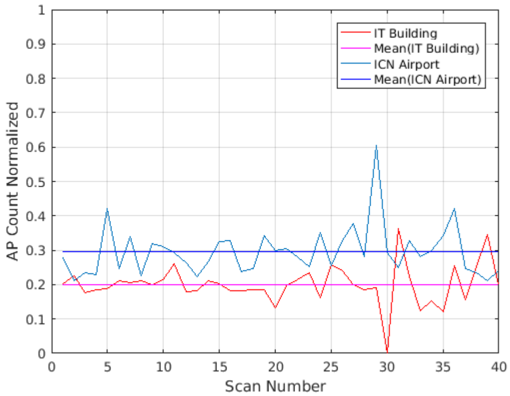

FIGURE 1. Ratio and mean value of good APs vs. Total APs at reference points of both sites.

match among multiple candidate fingerprint locations with minimize error distance.

A. K SELECTION

The first parameter in harvesting the accuracy is K, it is an offline parameter and calculated for each site only once. K in the proposed algorithm is the number of APs compared at each candidate position in the fingerprinting database to retrieve position information. In EWKNN the author has proposed the variable value of k, which is selected by setting a threshold to the RSSI comparison difference, where the threshold selection is manual and it is possible that all RSSI comparison are greater than the threshold while leaving K equal zero. In the proposed algorithm, the selection of K is with respect to Wi-Fi infrastructure present in the environment and is directly proportional to Good APs in the environment. According to Good AP's definition in [23], the APs with consistence presence at particular reference points are given high importance as compare to other which disappear in consecutive scans. For example, if the fingerprint database is constructed using N scans per RP, the APs present in N/2 scans on N are considered as Good AP. Referring to Eq. (1) and Eq. (2) helps us in K selection.

$$K_{RP_i} = (G_{RP_i}/T_{RP_i}) \tag{1}$$

$$K = K_{RSSV} * (Mean(K_{RP_1}, K_{RP_2}, \ldots K_{RP_N}) + C) \tag{2}$$

where G_{RP_i} and T_{RP_i} are Good AP count and Total AP count to each cell of the fingerprinting DB, respectively. K_{RSSV} is the total AP count in a particular RSSV. The Fig. 1 shows the selection of K for selected test sites; for each site 40 random reference points are examined to calculate the G_{RP_i} mean and to estimate the C margin for both sites. The variable C is the standard deviation of the Good APs mean values to give an upper limit to the selection of K. This selection process of K ensures that priority must be given to APs whose frequency of occurrence and order is high in the RSSV signal received at a particular location.

B. MAX MATCH COUNT (MMC)

Empirically, it is observed that RP candidates close to the ground truth and their neighboring RPs have a high number of AP matches, which is obvious. To verify this observation, the index values of the candidate RPs in a list sorted according to RMSE for 100 position requests on each site are observed. In the Fig. 2, the X-axis shows the number of normalized AP matches, while the Y-axis indicates the frequency of the RP closest to the ground truth for both sites. Here, the number of normalized AP matches means that, at each request, the minimum and the maximum number of matches are taken and then these values are normalized between 0 and 1. It is clear from the histogram that the existence of ground truth RP is high in candidate RPs having high match count. Furthermore, it is quite possible that the occurrence of an outlier with a strong correlation with a weak RSSI in paired APs is quite possible. Whereas the headcount of RPs closer to the ground truth having similar AP match count is large as compared to the outlier (mostly single in the count). Therefore, the maximum AP match count (MMC) with respect to neighbor count is an important measure that is used to avoid the outlier that have high AP match count but lower in the neighbor count. To sort candidates for the position, the RMSE of each RP is divided by its MMC value. In addition, the candidate at the top of

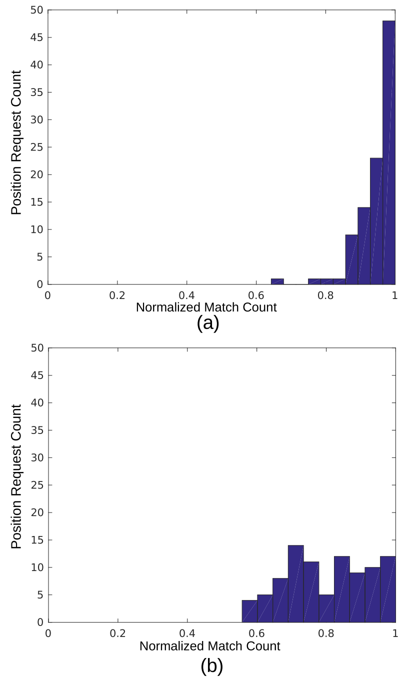

FIGURE 2. Likelihood of RP closest to ground truth with respect to normalized AP match count (a) IT Building (b). Incheon Airport.

TABLE 1. Statistics of K calculation using a standard deviation margin value of C = 0 and C = 0.1 at both sites(IT building and incheon airport).

Building Type	Avg. AP Count/RP	K (C=0.0)	K (C=0.1)
IT Building	73	22	29
ICN Airport	126	25	38

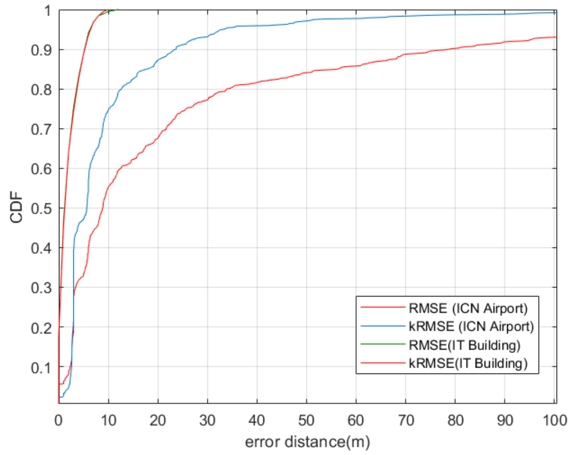

FIGURE 3. **Location accuracy performance graph of RMSE and KRMSE at IT building and ICN airport sites.**

the sorted list is deleted if the number of neighbors is one. This helps us to control the large distance error induced by the outliers.

C. KRMSE RATIO

The RMSE is one of the most powerful parameters and plays a key role in fingerprinting similarity check process. The RMSE is calculated by taking the square root of the sum of squared difference of the RSSI levels of the matching APs between the online user scan vector and an RP from the fingerprint DB (i.e., $RMSE = \sqrt{\frac{\sum_{j=1}^{n}(P_j - Q_j)^2}{n}}$). This parameter helps to sort potential candidates closer to the ground-truth location based on the minimum RMSE value. Experimentation shows that in indoor radio environments, some areas result in the selection of outliers with low RMSE values due to a high correlation between some APs with low RSSI values. Therefore, an additional division with the number of APs compared (i.e. MMC) is added in order to prioritize the candidate locations with a higher number of AP matches. The Fig. 3 shows the performance graph of the simple RMSE value and the kRMSE value with respect to their error distance measurement. It is clear from the graph that the kRMSE ratio provides better control of long-distance errors than simple RMSE. The Fig. 4 shows the spatial visualization of the sorted candidates with respect to the kRMSE value. It is clear from the visualization that candidates with the smallest kRMSE value are closer to the ground-truth.

$$kRMSE_i = \frac{1}{MMC}\sqrt{\frac{\sum_{j=1}^{n}\left(P_j - Q_j\right)^2}{n}} \tag{3}$$

where P and Q refer to two vectors, one of the online scan and the other of the database, respectively.

D. CENTROID DISTANCE

The distance from the center of gravity is also an important parameter for filtering the estimated positions closest to the

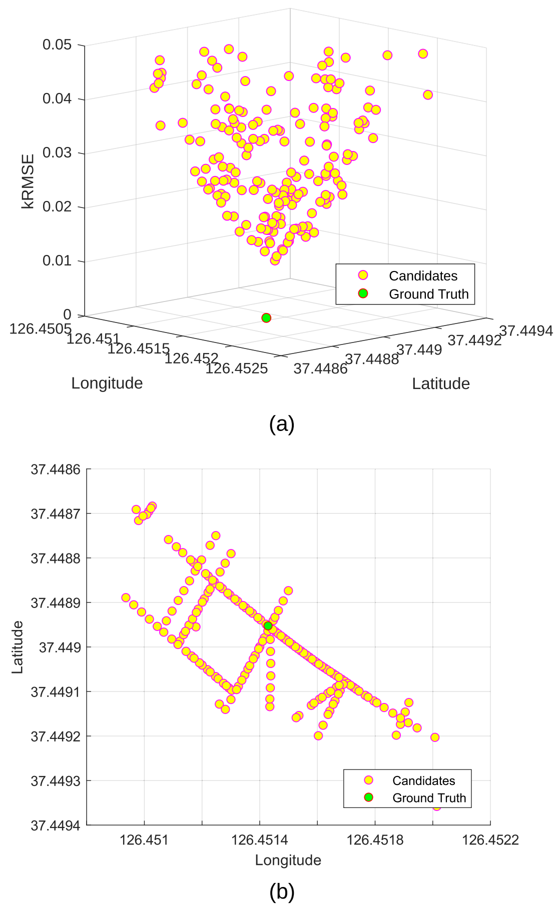

FIGURE 4. **Spatial location of reference Points sort by kMRSE value (a) 3D view, (b) Top view.**

ground [24]. The location of the centroid is the average location of the most ordered N RPs relative to the minimum kRMSE value. The position of the centroid in the proposed algorithm is calculated by taking an average of the coordinate values (X, Y) of each selected candidate position, i.e. the best candidates $N(i.e.8 \leq N \leq K)$ are selected to compute the centroid as follows.

$$P_{cen} = \frac{P(RP_1) + P(RP_2) + \ldots + P(RP_N)}{N} \tag{4}$$

where $P(RP_j) = (X_j, Y_j)$ denotes the position of jth RP and j = 1,2…N. $P_{cen} = (X_{cen}, Y_{cen})$ denotes the calculated position of the centroid from selected N RPs. In a grid an RP is surrounded by 8 neighboring RPs at first hop, therefore, at least top 8 candidates are selected to find the centroid position by taking the mean of their spatial location during a particular position request. The Euclidean distance of a candidate RP from calculated centroid is called its Centroid Distance. Centroid is expected to be the closest point to the ground truth. This distance from centroid is used to priorities RPs that are closer to the centroid. The Fig. 5 (a,b) shows the distribution of indices of candidates in a list sorted by kRMSE value from least to maximum. For example, more than 80% of candidates with a distance less than 1 meter from ground

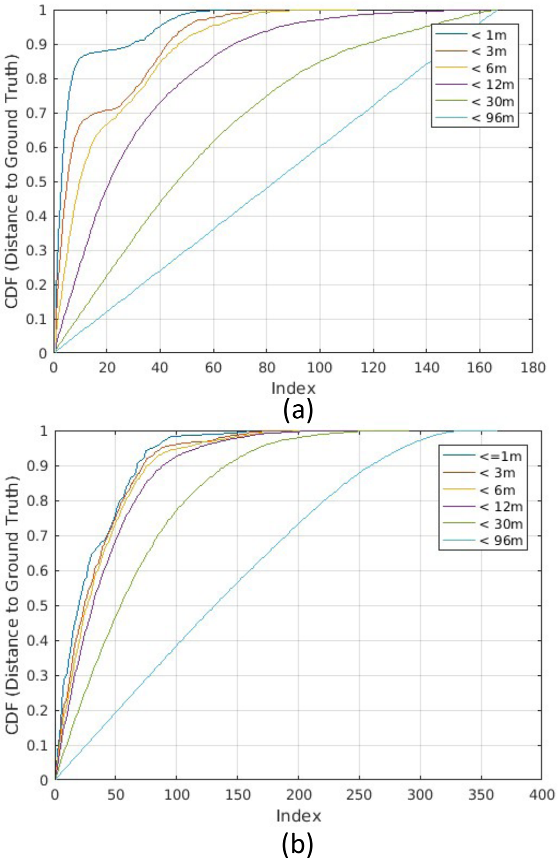

FIGURE 5. Distribution of candidates according to their distance from the centroid location in a list sorted at least to the maximum kMRSE. (a) IT Building (b) Incheon Airport.

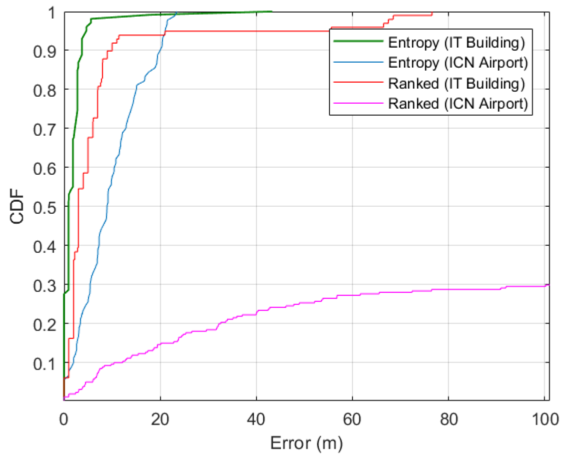

FIGURE 6. Empirical CDF of distance error calculate by using entropy parameter and ranked based algorithm.

truth are present in the first 20 indices of the IT building site (Fig. 5a), while nearly 80% of candidates with a distance less than 3 meters are present in top 50 indices of the ICN Airport site (Fig. 5b).

E. ENTROPY ERROR

Now, in this step, another important parameter called entropy error is calculated. The entropy error is a measure to estimate the similarities between the RSSV received online and each candidate location RSSV in the database. Referring to the Eq. (5), the proposed algorithm calculate Δ_{Vji} for each corresponding AP in both RSSVs, first by sorting the two RSSVs relative to the RSSI level, then measuring the difference of AP index in both lists. In addition, the order difference of an AP with a stronger RSSI is minimized by multiplying it by the RSSI level ($RSSI_{Vji}$) divided by -100. A similar procedure has been followed as a rank calculation in [21]. Our entropy parameter differs from the rank approach in two ways: first, the proposed algorithm use a simple difference formula and, second, it assign a higher weight to APs with strong RSSI values, as shown in Eq. (5) and Eq. (6). Simple difference formula makes it possible to use fewer computation resources and the assignment of higher weight to APs with strong RSSI bring uniqueness to RPs with respect to APs closer to the location of RP. Fig. 6 shows the position accuracy performance of the proposed entropy parameter in comparison with the ranked based approach at both sites. Therefore entropy helps in limiting the strong candidates around the ground truth by calculating the order difference of each AP with respect to its signal strength in online RSSV and a fingerprint from DB.

$$\Delta_{Vji} = \left|(I_{Vji} - I_{Cji}) * \frac{RSSI_{Vji}}{-100}\right| \tag{5}$$

$$Entropy = \frac{\sum^{k} \Delta_{Vji}}{k^2} \; where\ k \leq K \tag{6}$$

where I_{Vji} and I_{Cji}) are indices of matched APs in online RSSV and candidate RSSV from DB, respectively.

F. FACT CALCULATION

The FACT calculation is a measure used in the candidate selection process for the final position. The candidate with the minimum FACT value in the candidate list is selected as the final position of the mobile node or user. Two different approaches for FACT calculation are proposed as follows.

1) WEIGHTED APPROACH

The FACT calculation in weighted approach begins with the normalization of the parameter values of the selected candidate, as indicated in the Eqs. (7), (8), (9). Then the FACT is calculated as a product of *kRMSE* with the sum of *Entropy* and D_{cntr}, as shown in Eq. (10). The sum of the centroid distance and the entropy error is used to weight the priorities of candidates with minimal entropy error and the distance from the center of gravity.

$$\widehat{kRMSE} = \frac{kRMSE - min(kRMSE)}{min(kRMSE) + max(kRMSE)} \tag{7}$$

$$\widehat{Entropy} = \frac{Entropy - min(Entropy)}{min(Entropy) + max(Entropy)} \tag{8}$$

$$\widehat{D_{cntr}} = \frac{D_{cntr} - min(D_{cntr})}{min(D_{cntr}) + max(D_{cntr})} \quad (9)$$

$$FACT = \widehat{kRMSE} * (\widehat{Entropy} + \widehat{D_{cntr}}) \quad (10)$$

2) PSO APPROACH

In PSO approach, the Particle Swarm Optimization (PSO) technique is employed to measure the efficiency of each parameter (i.e., $\widehat{kRMSE}$, $\widehat{Entropy}$, and $\widehat{D_{cntr}}$) in FACT calculation [25]. PSO helps in finding the value of each constant (i.e., α, β, and γ) shown in Eq. (11). The value of each constant is optimized with the minimum distance error cost function in the PSO process.

$$FACT = \alpha \times \widehat{kRMSE} + \beta \times \widehat{Entropy} + \gamma \times \widehat{D_{cntr}} \quad (11)$$

The PSO algorithm runs for each distance error cost function i.e. 3rd quartile(75%), Average, Std. Dev., and Max error given in Eqs. (15), (16), (17),and (18), respectively, with the following configuration.

- 100 particles per Iteration.
- 50 Iterations per test case.

Velocity Update: Each particle's velocity is updated using Eq. (12):

$$v_i(t+1) = wv_i(t) + c_1r_1\,[x_i(t)x_i(t)] + c_2r_2\,[g(t)x_i(t)] \quad (12)$$

where i is the particle index, w is the inertial coefficient $c1$, $c2$ are acceleration coefficients ($0 \leq c_1, c_2 \leq 2$), $r1$, $r2$ are random values ($0 \leq r_1, r_2 \leq 1$) regenerated on every velocity update,

$v_i(t)$ is the particle's velocity at time t. $x_i(t)$ is the particle's position at time t. $x_i(t)$ is the particle's individual best solution as of time t. $g(t)$ is the swarm's best solution as of time t.

Position Update: Each particle's position is updated using Eq. (13):

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (13)$$

The dynamic model used in PSO algorithm to measure the ratio of each parameter for particular measure of accuracy is composed of following functions.

$$FACT = X_1 \times \widehat{kRMSE} + X_2 \times \widehat{Entropy} + X_3 \times \widehat{D_{cntr}} \quad (14)$$

$$cost_{3rdQ} = Quartile3\{D_1, D_2, D_3, \ldots, D_n\} \quad (15)$$

$$cost_{average} = \frac{\sum_{i=1}^{n} D_i}{n} \quad (16)$$

$$cost_{std} = std\{D_1, D_2, D_3, \ldots, D_n\} \quad (17)$$

$$cost_{maxerr} = max\{D_1, D_2, D_3, \ldots, D_n\} \quad (18)$$

where D_i is the Euclidian distance between estimated position and ground truth. The search space of X1 (α), X2 (β), and X1 (γ) is 0 to 1. Each cost function calculates corresponding error value for a set of 120 and 270 position requests for IT Building and Incheon Airport per Iteration, respectively.

Table 2 shows the estimated effectiveness of each parameter for both IT Building and Incheon Airport for selected cost functions of accuracy. Entropy parameter shows no importance in case of IT Building case whereas it shows significant importance in case of Incheon Airport. For Max error and 3rd quartile importance is shifted to parameter D_{cntr} as γ receives the high value which is obvious.

TABLE 2. Estimated α, β, γ values using PSO with different cost function of accuracy.

Cost Function	Test Site	α	β	γ
Average	IT Building	0.3493	0.0578	0.6497
	ICN Airport	0.5899	0.6762	0.4572
Std. Dev.	IT Building	0.0926	0.04573	0.7744
	ICN Airport	0.8704	0.7587	0.7078
3rd quartile	IT Building	0.4257	0.0008	0.3392
	ICN Airport	0.7703	0.2666	0.7342
Max Error	IT Building	0.0	0.0336	0.9023
	ICN Airport	0.3373	0.6885	0.8967

V. ALGORITHM AND WORKING DETAILS

The proposed algorithm works in two phases, similar to the classical fingerprinting algorithm: an offline phase of the survey of the environment to build a fingerprinting DB of the Wi-Fi network, and an online phase in which the position estimation is performed in real time using a mobile device. Before starting the online phase, the estimation of K is made at the end of the offline phase using information stored for each RP in the fingerprinting DB of each site. Algorithm 1 shows the details of the 2nd phase of the fingerprinting algorithm. The algorithm starts with an RSSI list received by a mobile node at an unknown position, i.e., $RSSV = (mac_1, rssi_1, mac_2, rssi_2, \ldots .mac_k, rssi_k)$, using the value K calculated in the first phase and N, the number of candidates needed to find the centroid as input parameters. The output of the algorithm is the position of the selected candidate RP. After sorting the RSSV and number of APs less than expected value K at Line 5, a function retrieves all expected candidate RPs from the database with respect to the list of APs in the RSSV.

Then the algorithm traverses all the cells recursively to find the best candidate with a precise position. By traversing each candidate cell, the algorithm first sorts the list of APs of each cell relative to the RSSI of the corresponding RP point. Then, it goes further through the list of APs in the selected candidate cells and collects the sum of the RSSI differences, the order difference of the APs relative to the order of the sorted list, and counts the number of APs compared. The loops terminate when comparing the number K of APs with the strongest RSSIs in the received vector. All the parameters described above are then calculated one

Algorithm 1 Proposed Algorithm

```
Input: RSSV, kValue, N
Output: Position
    Initialisation :
1:  SortByRSSI(RSSV)
2:  if RSSV.length < kValue then
3:     kValue = RSSV.length
4:  end if
5:  fpCellList = GetCandidateCellsFromFingerprintDB(RSSV);
    LOOP Process:{Loop through all candidate reference positions/cells in fpCellList}
6:  for each fpCell in fpCellList do
7:     SortByRSSI(fpCell.APInfoList)
8:     j = 0
9:     sum = 0
10:    for each APInfo in fpCell.APInfoList do
11:       if RSSV.contains(APInfo) then
12:          index = get index of APInfo in RSSV
13:          rssvalue = get RSSI of APInfo in RSSV
14:          AddIndexandRSSIInfomationToList(IndexRSSIList, j, rssvalue)
15:          sum = sum + (abs(APInfo.RSSI - rssvalue))^2
16:          j = j + 1
17:       end if
18:       if j > kValue then
19:          break Loop
20:       end if
21:    end for
22:    candidate.position = fpCell.position
23:    candidate.krmse = sqrt(sum/j)/j;
24:    candidate.entropy = CalculateEntropyofCurrentCandidate(IndexRSSIList)
25:    AddToList(candidateList, candidate)
26: end for
27: SortBykRMSE(candidateList)
28: SelectFirstHalfCandidates(candidateList)
29: centroid = GetCentroid(candidateList, N)
30: SortByDistanceToCentroid(candidateList)
31: SelectTopKCandidates(candidateList)
32: candidate = SelectCandiateWithMinFACT(CandidateList)
33: Position = candidate.position
34: return Position
```

by one and the candidates are added to the list of candidates expected at the end of the iterations. Then, the half of the candidates with high errors after sorting the list using kRMSE are removed on lines 26 to 27 and the centroid from the N (10 to 15) best candidates is calculated. Finally, a candidate with the least FACT value will be selected as a valid candidate.

VI. EXPERIMENTAL SETUP

Two different sites as real-time testbeds for experimentation have been selected: an office environment in an IT Building (Fig. 7a) of area 92 × 32 m^2 with a ceiling height of 3 meters and a small area at the Incheon Airport Passenger Lounge with an area of 300 × 100 m^2 (Fig. 7b) and a ceiling height of approximately 10 meters. For experimentation, two Samsung Galaxy S8 and LG G6 mobile phones are used for the survey of environments in order to build the fingerprinting DB before an online phase of the positioning; also the same devices as the target devices are utilized. The fingerprinting server consisted of an Intel machine equipped with Core i7 processor and 16 GB RAM running Windows Server Edition with Postgres SQL DB server. The collection of fingerprints is carried out in a grid of 1 meter and 3 meters respectively in IT-building and Incheon airport. Whereas, 4 scan per RP are collected to build the fingerprinting database.

VII. RESULTS AND DISCUSSION

In order to examine the differences in performance between the output positioning results of the underlying test

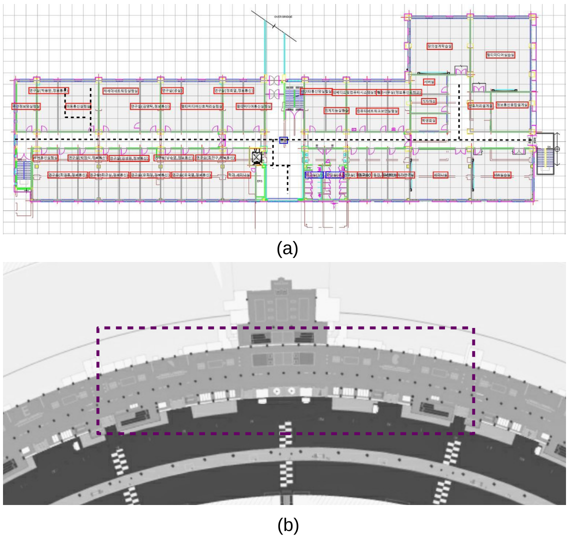

FIGURE 7. Floor plan of real test sites selected for evaluation of proposed algorithm based on Wi-Fi fingerprinting (a) IT Building (b). Incheon Airport.

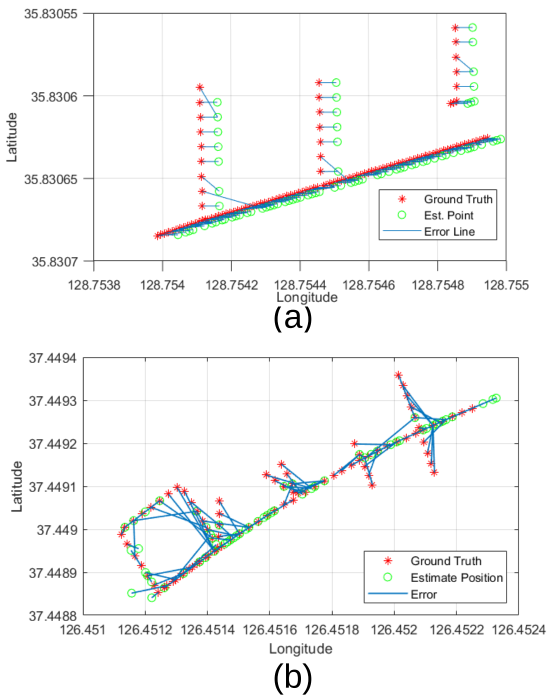

FIGURE 8. The "*"s are representing Ground truth points, the "o"es are estimated locations, and blue lines represents the distance error in estimation. (a) IT Building (b). Incheon Airport.

environments, the dynamics of the two environments is examined and it has been determined that many factors are responsible for the degradation of the accuracy of the estimation of location in overcrowded environments (public sites). The Fig. 9 shows the cross-section of the RSSI propagation change in real time on a single path with respect to the distance to the APs with the highest level RSSI, one to IT

TABLE 3. Statistics of location accuracy of k-nearest neighbor (KNN), ranked based (ranked), enhanced weighted KNN (EWKNN), and proposed algorothm with weighted approach (mClips) and particle swam optimization (mClips + POS) algorithm in meters with respect to quantities: average, Std. Dev., 3rd quartile, and maximum positioning error at all test sites.

Site	Technique	Average	Std. Dev.	3rd Quart	Max Error
IT Building	kNN	1.00	3.93	3.00	52.00
	Ranked	7.28	14.27	7.00	68.47
	EWKNN	2.25	3.04	3.51	11.723
	mClips	1.67	2.59	1.86	14.93
	mClips+PSO	0.93	2.23	1.86	11.20
Terminal 1	kNN	51.91	38.25	76.17	168.00
	Ranked	126.53	72.09	187.90	225.4
	EWKNN	20.46	18.83	23.09	129.00
	mClips	9.26	7.99	12.57	30.95
	mClips+PSO	8.18	8.51	10.06	30.35
CS Dept.	kNN	11.50	8.13	14.48	33.20
	Ranked	7.13	10.58	7.00	41.01
	EWKNN	2.51	4.06	2.12	24.71
	mClips	0.90	0.97	1.00	4.33
	mClips+PSO	0.98	0.97	1.00	4.00
Terminal 2	kNN	38.35	25.90	56.28	93.08
	Ranked	74.22	49.57	105.02	178.2
	EWKNN	40.30	25.01	57.99	108.8
	mClips	5.56	6.26	6.14	24.03
	mClips+PSO	6.03	6.44	7.99	25.32

Building and the other at Incheon Airport (Terminal 1). The Fig. 9 a and 9 b show the view from the side while 9 c and 9 d represent the top view of the modification of the RSSI level along a path for APs to the IT and ICN Airport buildings, respectively. The differences in RSSI level changes between the two sites are due to the following reasons: firstly, the APs were mounted at high altitude at Incheon airport, while those of the IT building were raised to around 10 feet above the ground. Secondly, Incheon Airport is a high roof atrium building, while the IT building is an office space with corridors. In addition, Incheon Airport has more pedestrian traffic than the IT building. Figs. 9a and 9b show that in the IT Building, the change is the RSSI with respect to distance is quite clear and monotonic as compared to that in Incheon airport. This RSSI level change behavior results in high accuracy in the IT building (i.e. the office environment) compared to Incheon Airport (i.e. a public place) for the estimation of the position. The Fig. 8 shows the result of the positioning test performed on both sites, parts (a) and (b) show the ground truth in the form of red marks "*", the corresponding estimated position being indicated under form of green marks "o", and the blue line represents the distance error in the estimate of the position at the IT Building and Incheon Airport sites, respectively. For the IT building, the estimated positions are shifted one meter horizontally in the figure to avoid overlap.

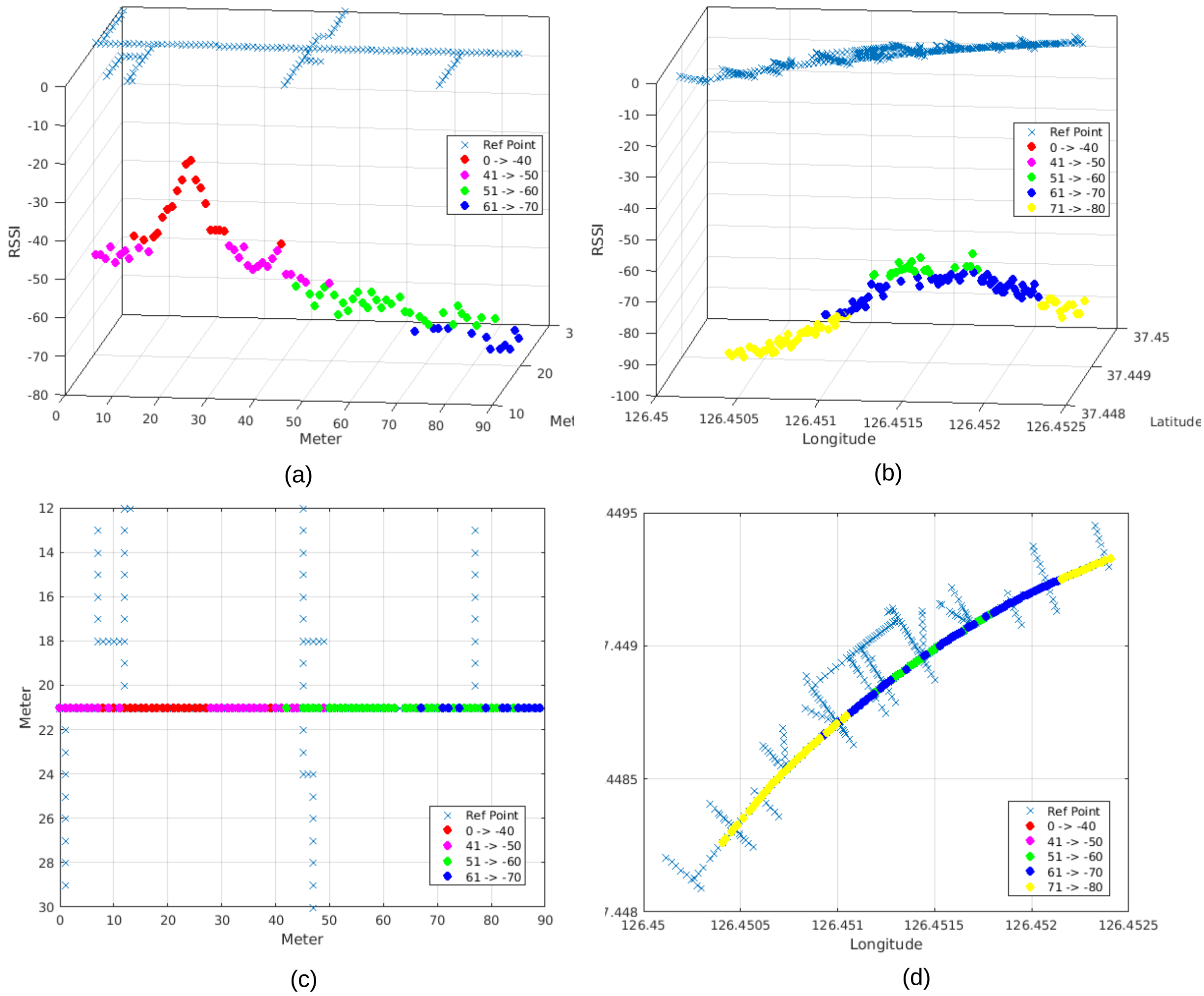

FIGURE 9. **Cross section of radio map of APs along a straight path to visualize change in received signal strength against distance (a),(b) Side view and (c)(d) top view of IT building and incheon airport(Terminal 1), respectively.**

The experimental results of Fig.10 show that our algorithm outperformed the other techniques in both environments, IT Building part(a) and ICN Airport part(b) and achieved an accuracy of 1.86 meters and 10.06 meters in the 3rd quartile in office and public environment, respectively. Here, the choice of the third quartile measure is due to several indoor location competition experiences organized by the International Conference on International Conference on Indoor Positioning and Indoor Navigation [26] where it is used to rank IPS approaches [27]. Furthermore, statistics in Table 3 summarizes the location accuracy of kNN approach from RADAR system (KNN) [19], Rank based approach (Ranked) [21], Enhanced Weighted KNN approach (EWKNN) [20], and Proposed Algorithm with the weighted approach (mClips) and Particle Swarm Optimization (mClips + POS) algorithm in meters with respect to quantities: Average, Std. Dev., 3rd quartile, and Maximum Positioning Error at all test sites. Moreover, it is clear from the results that the proposed approach minimizes the maximum error in the office and in public places, which is a very important factor for the hybrid IPS to minimize the search space of the geomagnetism and drift errors over long distances in PDR techniques. To verify results the location estimation has performed at two more sites i.e. CS Department of area $60 \times 32\ m^2$ and Terminal 2 site of area $200 \times 60\ m^2$ as an office and a public environment, respectively. The collection of fingerprints is carried out on a grid of 1 meter in the department CS while on a grid of 2 meters in the terminal 2, by carrying out 4 scans per RP. The mobile devices used on the new sites are Samsung Galaxy 9, LG G7. The results show that the proposed algorithm achieves high precision and limits the maximum error at the two new sites, as shown in Fig. 10 part(c) CS Department and part(d) Incheon Airport Terminal 2. To estimate the cost of accuracy, the processing time of all algorithms, for a single position request without DB connection and access delay time is calculated. Table 4 indicates the time required to estimate a signal position request by different algorithms against a database size of fingerprint records at each site.

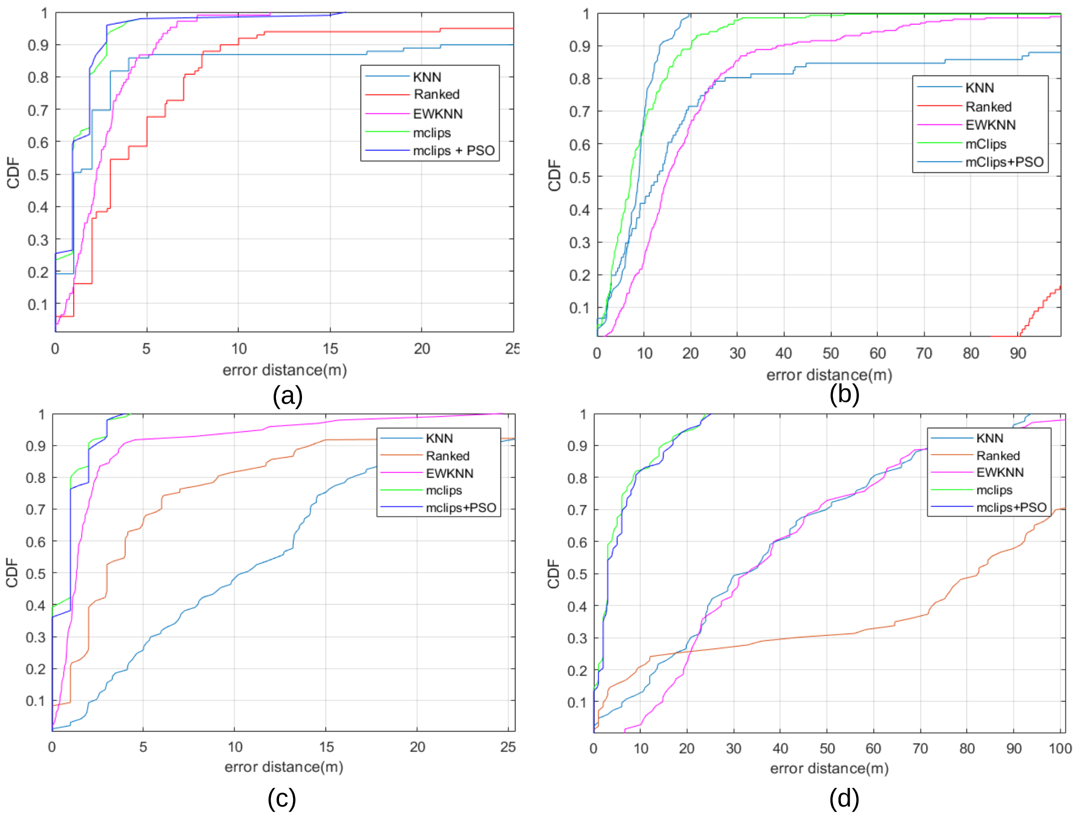

FIGURE 10. Distance error CDF of k-nearest neighbor (KNN), ranked based (Ranked), enhanced weighted KNN (EWKNN), and proposed algorithm with the weighted approach (mClips) and particle swarm optimization (mClips + POS) approaches is calculated for all sites, IT building.(a), Incheon Airport (Terminal 1)(b), CS Department(c) and Incheon Airport (Terminal 2) (d).

TABLE 4. Statistics of computation time in millisecond(ms) required for estimation of single position request by k-nearest neighbor (KNN), ranked based (ranked), enhanced weighted KNN (EWKNN), and proposed (mClips) algorothms with respect to database size of all test sites.

Test Site	DB Records	kNN (ms)	Ranked (ms)	EWKNN (ms)	mClips (ms)
IT Building	12620	0.65	1.03	0.93	0.97
Terminal 1	29374	1.04	1.62	1.19	1.61
CS Dept.	18033	1.29	1.92	1.74	2.11
Terminal 2	32215	1.39	1.74	1.71	1.89

The cost of processing time relative to the precision of our algorithm is almost minimal, since the normal delay in two consecutive scans of a mobile device is usually 1 to 3 seconds. However, the algorithm requires an automated estimate of the K value for each site before starting to estimate the online position.

VIII. CONCLUSION

In this study, an empirical approach to improve the accuracy of Wi-Fi fingerprinting without the use of infrastructure or special devices is investigated, in order to serve as a starting point estimation technique for hybrid IPS systems. An empirical approach to harvest parameters in the RSSI vector was discussed. In addition, an algorithm with several improvements has been proposed; The new algorithm has significantly improved the positioning accuracy of Wi-Fi fingerprinting for office environments and especially for crowded public areas. The results show that the proposed approach has achieved significant accuracy in real-time environments. Possible directions for future work include adding parameters for trajectory tracking and verifying that a hybrid approach to the proposed technique can effectively improve the tracking performance of indoor positioning systems. Integrating deep learning with these selected parameters is another direction to explore.

REFERENCES

[1] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 6, pp. 1067–1080, Nov. 2007.

[2] U. Shala and A. Rodriguez, "Indoor positioning using sensor-fusion in android devices," M.S. thesis, School Health Society, Dept. Comput. Sci., Kristianstad Univ., Kristianstad, Sweden, 2011.

[3] L.-H. Chen, E. H.-K. Wu, M.-H. Jin, and G.-H. Chen, "Intelligent fusion of Wi-Fi and inertial sensor-based positioning systems for indoor pedestrian navigation," *IEEE Sensors J.*, vol. 14, no. 11, pp. 4034–4042, Nov. 2014.

[4] Z. Chen, H. Zou, H. Jiang, Q. Zhu, Y. C. Soh, and L. Xie, "Fusion of WiFi, smartphone sensors and landmarks using the Kalman filter for indoor localization," *Sensors*, vol. 15, no. 1, pp. 715–732, Jan. 2015.

[5] R. Harle, "A survey of indoor inertial positioning systems for pedestrians," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1281–1293, 3rd Quart., 2013.

[6] R. Faragher and R. Harle, "SmartSLAM—An efficient smartphone indoor positioning system exploiting machine learning and opportunistic sensing," in *Proc. ION GNSS*, vol. 13, Sep. 2013, pp. 1–14.

[7] A. K. M. M. Hossain and W.-S. Soh, "A survey of calibration-free indoor positioning systems," *Comput. Commun.*, vol. 66, pp. 1–13, Jul. 2015.

[8] Y. T. Chan and K. C. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 1905–1915, Aug. 1994.

[9] M. N. Husen and S. Lee, "Indoor location sensing with invariant Wi-Fi received signal strength fingerprinting," *Sensors*, vol. 16, no. 11, p. 1898, 2016.

[10] M. U. Ali, S. Hur, and Y. Park, "LOCALI: Calibration-free systematic localization approach for indoor positioning," *Sensors*, vol. 17, no. 6, p. 1213, 2017.

[11] V. Honkavirta, T. Perala, S. Ali-Loytty, and R. Piche, "A comparative survey of WLAN location fingerprinting methods," in *Proc. 6th Workshop Positioning, Navigat. Commun.*, Mar. 2009, pp. 243–251.

[12] G. Minaev, A. Visa, and R. Piché, "Comprehensive survey of similarity measures for ranked based location fingerprinting algorithm," in *Proc. Int. Conf. Indoor Positioning Navigat. (IPIN)*, Sep. 2017, pp. 1–4.

[13] X. Wang, L. Gao, and S. Mao, "CSI phase fingerprinting for indoor localization with a deep learning approach," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1113–1123, Dec. 2016.

[14] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity Wi-Fi," *IEEE Trans. Mobile Comput.*, to be published.

[15] J. Xiong and K. Jamieson, "ArrayTrack: A fine-grained indoor location system," in *Proc. Usenix*, vol. 13, Apr. 2013, pp. 71–84.

[16] M. Ibrahim *et al.*, "Verification: Accuracy evaluation of WiFi fine time measurements on an open platform," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, Nov. 2018, pp. 417–427.

[17] K. Khoshelham and S. Zlatanova, "Sensors for indoor mapping and navigation," vol. 16, no. 5, p. 655, May 2016.

[18] J. Chung, M. Donahoe, C. Schmandt, I.-J. Kim, P. Razavai, and M. Wiseman, "Indoor location sensing using geo-magnetism," in *Proc. 11th Proc. 9th Int. Conf. Mobile Syst., Appl., Services*, Jun. 2011, pp. 141–154.

[19] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proc. IEEE 9th Annu. Conf. Comput. Commun. Societies*, vol. 2, May 2000, pp. 775–784.

[20] B. Shin, J. H. Lee, T. Lee, and H. S. Kim, "Enhanced weighted K-nearest neighbor algorithm for indoor Wi-Fi positioning systems," in *Proc. 8th Int. Conf. Comput. Technol. Inf. Manage. (ICCM)*, vol. 2, Apr. 2012, pp. 574–577.

[21] J. Machaj, P. Brida, and R. Piché, "Rank based fingerprinting algorithm for indoor positioning," in *Proc. Int. Conf. Indoor Positioning Navigat.*, Sep. 2011, pp. 1–6.

[22] J. Torres-Sospedra, R. Montoliu, S. Trilles, O. Belmonte, and J. Huerta, "Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems," *Expert Syst. Appl.*, vol. 42, no. 23, pp. 9263–9278, 2015.

[23] A. Aksu and P. Krishnamurthy, "Sub-area localization: A simple calibration free approach," in *Proc. 13th Int. Conf. Modeling, Anal., Simulation. Wireless Mobile Syst.*, Oct. 2010, pp. 63–72.

[24] Z. Liu and J. Chen, "A new sequence-based iterative localization in wireless sensor networks," in *Proc. Int. Conf. Inf. Eng. Comput. Sci.*, Dec. 2009, pp. 1–4.

[25] M. Clerc and J. Kennedy, "The particle swarm-explosion, stability, and convergence in a multidimensional complex space," *IEEE Trans. Evol. Comput.*, vol. 6, no. 1, pp. 58–73, Feb. 2002.

[26] (2010). *International Conference Indoor Positioning Indoor Navigation (IPIN)*. Accessed: Mar. 10, 2019. [Online]. Available: http://ipin-conference.org/

[27] J. Torres-Sospedra *et al.*, "The smartphone-based offline indoor location competition at IPIN 2016: Analysis and future work," *Sensors*, vol. 17, no. 3, p. 557, 2017.

MUHAMMAD USMAN ALI received the M.S. degree in computer network engineering from the University of Engineering and Technology Taxila, Pakistan, in 2008. He is currently pursuing the Ph.D. degree in information and communication engineering with Yeungnam University, South Korea, where he is also a Research Assistant with the Mobile Communication Laboratory. His current research interests include multisensor fusion-based indoor positioning systems, Wi-Fi fingerprinting, indoor navigation and mapping, and computer vision technologies.

SOOJUNG HUR received the B.S. degree from Daegu University, Gyeongbuk, South Korea, in 2001, the M.S. degree in electrical engineering from the San Diego State University, San Diego, CA, USA, in 2004, and the M.S. and Ph.D. degrees in information and communication engineering from Yeungnam University, South Korea, in 2007 and 2012, respectively, where she is currently a Research Professor with the Mobile Communication Laboratory. Her current research interests include performance of mobile communication, indoor/outdoor location, and unnamed vehicle.

SANGJOON PARK received the B.S. and M.S. degrees in electronics engineering from Kyungpook National University, in 1988 and 1990, respectively, and the Ph.D. degree from the Computer Science Department, North Carolina State University, in 2006. He is currently the Director of the Defense Warrior ICT Platform Research Section in Electronics and Telecommunications Research Institute (ETRI), South Korea. He has also worked as a Senior Researcher with the Agency for Defense Development (ADD), from 1990 to 2001. His current research interests are in positioning, wireless sensor networks, next generation embedded sensor networks, multi sensor data fusion, and target tracking.

YONGWAN PARK received the B.E. and M.E. degrees in electrical engineering from Kyungpook University, Daegu, South Korea, in 1982 and 1984, respectively, and the M.S. and Ph.D. degrees in electrical engineering from The State University of New York at Buffalo, Buffalo, NY, USA, in 1989 and 1992, respectively. He was a Research Fellow of the California Institute of Technology, from 1992 to 1993. From 1994 to 1996, he has served as the Chief Researcher for developing IMT-2000 System at SK Telecom, South Korea. Since 1996, he has been a Professor of information and communication engineering with Yeungnam University, South Korea. In 2000, he was an Invited Professor of the NTT DoCoMo Wireless Lab., Japan. He was also a Visiting Professor with UC Irvine, Irvine, CA, USA, in 2003. From 2008 to 2009, he has served as the Director of the Technology Innovation Center for Wireless Multimedia by the Korean Government. From 2009 to 2017, he has also served as the President of the Gyeongbuk Institute of IT Convergence Industry Technology (GITC), South Korea. He is also serving as the Chairman of the 5G Forum Convergence Service Committee in South Korea. His current research areas of interests include 5G systems in communication, OFDM, PAPR reduction, indoor location-based services in wireless communication, and smart sensors (LIDAR) for the smart car.