A Novel Weighted Fusion based Efficient Clustering for Improved Wi-Fi Fingerprint Indoor Positioning

Pampa Sadhukhan, *Member, IEEE,* Keshav Dahal, *Senior Member, IEEE,*
Pradip K. Das, *Life Senior Member, IEEE.*

***Abstract*—The *received signal strength* (RSS) based *Wi-Fi* fingerprint technique is not only a cost-effective means for indoor positioning but also provides reliable positioning accuracy in the indoor settings. Thus, such positioning technique has drawn many researchers′ attention to address its several limitations like degraded *positioning accuracy* due to continuous changes in surrounding environment, high *positioning overhead*, *storage overhead* etc. To address these issues, we propose a *novel weighted fusion based efficient clustering strategy* (*WF-ECS*) for fingerprint positioning system in this paper. Our proposed technique *WF-ECS* computes a weighted average of the group of *reference points* (RPs) having similar RSS patterns and thus, creates a more perfect match between fused positional co-ordinates and RSS patterns considered for merging to a single entry. Extensive experimentation have been carried out to evaluate and compare the performances of our proposed system *WF-ECS* with the contemporary fingerprint positioning systems including our prior work using the simulation test bed, the dataset collected from our departmental building and also the *benchmark dataset*. The experimental results depict that our newly proposed technique *WF-ECS* can outperform the contemporary techniques in terms of *positioning accuracy* and *positioning overhead* while reducing the *storage overhead* in real indoor settings.**

***Index Terms*—fingerprint, positioning, RSS, clustering, fusion, accuracy, storage overhead.**

I. INTRODUCTION

Wireless positioning is crucial for providing emergency services anywhere as well as for various location-sensitive services required for smart living in urban areas [1]. Although, global navigation satellite system (GNSS) can accurately estimate the location in open-sky areas, its very poor signal strength in indoor areas fails to provide reliable positioning solution in such areas [2]. To resolve this issue, many researches have strived to develop indoor positioning systems (IPSs) using radio frequency (RF) technologies over the past decades. A state-of-the-art review of such positioning systems is provided in [3]. Majority of them use Wi-Fi technology due to its deployment in almost every building in urban areas and also its availability in most of the handheld devices. Among the existing RF-based positioning schemes, range-based positioning that uses either time delay or angle measurements cannot estimate the target's position without knowing the positional coordinates of access points (APs) [4]. Moreover, performances of such positioning schemes are badly affected by the obvious presence of non-light-of-sight propagation and multipath effect in complex indoor settings. Received signal strength (RSS) based fingerprint positioning being a kind of range-free positioning scheme can provide reasonably good accuracy in indoor areas without having the prior knowledge about positions of the APs [5]. Thus, it has drawn many research interests for designing IPSs. However, such positioning methods employ two stages (training and positioning) for location estimation. In the training stage, a radio map which is the database of RSS patterns obtained at the selected locations or reference points (RPs) within the positioning area is constructed. The target's location is estimated in the positioning stage by comparing the currently observed RSS pattern with those stored in the radio map. Moreover, high variation in the RSS values obtained from a certain AP at a particular location over time, i.e., the stochastic nature of RSS dataset, severely degrades the accuracy of fingerprint positioning. As evidenced in [6,7], the positioning accuracy can be significantly enhanced by using more closely spaced RPs and also by deploying more Wi-Fi APs in the positioning area. This, however, significantly increases the radio map's storage requirement and also the comparison overhead in the positioning stage, i.e., positioning overhead. Hence, storage overhead and positioning overhead are undoubtedly two major significant issues with such positioning systems.

We, thus, aim in this paper to lessen the positioning and storage overhead of the fingerprint positioning method as well as the negative impact of the time varying nature of RSS samples on the positioning accuracy. Our previously proposed efficient clustering strategy [8] employs average fusion policy to reduce the storage and positioning overhead of fingerprint positioning. However, such fusion policy fails to diminish the negative impact of RSS dataset's stochastic nature on the positioning accuracy. The average of the positional co-ordinate values of a group of RPs having similar RSS pattern cannot be a perfect match with the average RSS pattern of the same due to the unequal Euclidean distances between each individual RSS pattern included in a group and its average RSS pattern. Smaller the Euclidean distance between a RP's RSS pattern and the average RSS pattern of its subset, larger the contribution it will have to create the fused RP entry. The above reasoning motivates us to propose a weighted fusion based efficient clustering strategy for fingerprint positioning here. Moreover, the previously proposed clustering method selects only the RP obtaining maximum RSS from the AP associated with its cluster as the initial point of data fusion

P. Sadhukhan is with School of Mobile Computing & Comm., Jadavpur University, India. E-mail: pampa.sadhukhan@ieee.org

K. Dahal is with AVCN research centre, School of Computing, Engineering. & Physical Sciences, University of the West of Scotland, UK. E-mail: keshav.dahal@uws.ac.uk

P. K. Das was with Dept. of Computer Science & Engineering, Jadavpur University, India. E-mail:pkdas@ieee.org

within that cluster. So, the effect of considering some other RP as the initial point of data fusion on the performances of the proposed positioning system has not been well investigated in our prior work [8]. We, therefore, propose in this paper a novel weighted fusion based efficient clustering strategy that considers both RPs obtaining either maximum or minimum RSS from the corresponding AP as the potential initiator of data fusion within a cluster to improve the positioning performances of fingerprint method. The contributions of our proposed work in this article are listed below.

- We propose a novel weighted fusion based efficient clustering strategy for RSS fingerprint based indoor positioning to improve the positioning accuracy while reducing the storage and positioning overhead. The proposed weighted fusion policy determines the weighting coefficient for each individual RP included in a group of RPs with similar RSS pattern based on the Euclidean distance of its associated RSS pattern from its group's average RSS pattern. Then, it computes weighted average of the positional co-ordinates and RSS patterns included in that group to create the fused entry.
- The proposed clustering strategy employs two different data fusion methods by selecting two potential RPs obtaining either maximum or minimum RSS from the corresponding AP as the initial point of data fusion within each cluster, to improve the positioning performances of fingerprint system.
- We evaluate and compare the performances of the proposed positioning method with six contemporary methods in a simulation test bed having randomly deployed Wi-Fi APs as well as using two different real RSS datasets including the benchmark dataset of *UJIndoorLoc* [9]. The six other existing methods considered in this paper for performance comparison are *k-means* clustering [10], *affinity propagation clustering* (*AFPC*) [11], *self organizing map* (*SOM*) based clustering [12] and the recently proposed principal component analysis based feature reduction technique coupled with kernel ridge regression (PCA-KRR) [13] along with our previously proposed average fusion based efficient clustering strategy (*AF-ECS*) [8] and *two-way hierarchical clustering strategy* ($2-way\ HCS$) [14].

The remaining part of this paper is organized as follows. Various existing Wi-Fi based IPSs are reviewed in Section II. Section III presents our proposed clustering based fingerprint system. In Section IV, we evaluate our proposed system and compare its performances with the contemporary systems considered in this paper. Lastly, we conclude and present our future research goals in Section V.

II. Related Work

A lot of researches have been carried out in the area of fingerprint based indoor positioning to address several issues like high calibration effort, positioning overhead etc., and also to improve the positioning accuracy. Thus, the existing fingerprint positioning techniques can be divided into three different categories each of which has been reviewed in a separate subsection. Subsection II.A reviews the fingerprint methods that aim to reduce the high calibration overhead incurred by the construction of radio map or its periodic update. In subsection II.B, the fingerprint methods designed to improve the positioning accuracy have been briefly reviewed. On the other hand, a deeper literature review on various fingerprint methods aiming to reduce the positioning overhead is presented in subsection II.C.

A. Fingerprint techniques reducing calibration effort: To reduce the calibration effort, crowd sourcing approach coupled with some machine learning/interpolation method are applied in [15-19]. In the crowd-sourcing approach, *RSS* values and the inertial sensor readings collected automatically via smart phone are processed to construct the radio map. However, such systems cannot provide reasonable positioning accuracy due to the differences in devices used to collect data, data collection point and also the collection time. To construct the radio map from several labeled fingerprints and user traces collected in the target area, the cubic spline interpolation technique is applied by the researchers in [15], whereas an incremental and adaptive process based on some unsupervised learning is used in [16] for the same purpose. The generic interpolation method is used to process the crowd-sourced RSS and magnetometer readings for the construction of radio map in [17]. For periodic update of the radio map, a mobile crowd-sensing based self-updating algorithm and a deep-learning based self-calibration method are proposed in [18] and [19] respectively. Several researchers, on the other hand, have applied some path loss model to build up the radio map. The researchers in [20], employs ray-tracing method to generate the path loss model, which is used along with the AP′s positional data and indoor map for the same purpose. Considering the wall attenuation effect on the path loss model, multi-wall model is employed to construct the radio map in [21]. To ensure faster and accurate construction of the radio map, a hybrid method integrating the crowd-sourcing approach with some path loss model and interpolation technique is proposed in [22]. In [23], the low-rank matrix completion algorithm is used on the RSS data collected from fewer evenly arranged RPs in the positioning area to construct the radio map. A thorough investigation on how the radio map degrades due to several factors like environmental dynamics, changes in wireless infrastructure and/or indoor layout etc., is carried out by the researchers in [24]. The experimental results given there show that radio map's degradation due to changes in Wi-Fi infrastructure is quantitatively larger than the same for environmental effects over time.

B. Fingerprint techniques improving positioning accuracy: To improve the positioning accuracy of fingerprint method in the complex indoor settings where the same differences at the different labels of RSS values do not show equal geometrical distance differences, a feature-scaling-based k-nearest neighbor (FS-kNN) algorithm is proposed in [25], whereas a continuous feature-scaling model which further improves the accuracy is proposed in [26]. The *robust principal component analysis* is applied in both phases of the fingerprint positioning method proposed in [27] in order to remove the sparse noises and the unstable APs from the measured RSS datasets. To address the problem of device heterogeneity that is also responsible for positioning accuracy degradation, the researchers in [28] has proposed a weighted ensemble classifier for the fingerprint positioning system. In [29], the particle swarm optimization method is, at first applied to select the optimum set of APs needed for the positioning service and then a feature-based ensemble model for the selected APs is proposed to keep possession of the generality of positioning performance. A path-loss based positioning method combining Rayleigh fading and log-normal shadowing model to improve the accuracy in indoor IoT environment is proposed in [30]. However, such method cannot determine the target's location without knowing the APs' position. Deep learning (DL) method is also used in fingerprint positioning to improve its accuracy. A multi-layer perceptron-based DL method for fingerprint positioning that uses both channel state information (CSI) and RSS values, is proposed in [31]. Like the RSS, CSI however, cannot be obtained via all types of wireless network interface cards. A recurrent neural network-based fingerprint method that exploit the sequential correlation of RSS values is proposed in [32]. However, such method requires a rigorous training by large number of user traces generated randomly to achieve the desirable accuracy.

C. Fingerprint techniques reducing positioning overhead: To reduce the positioning overhead, several existing fingerprint methods exploit either a single or multiple strongest APs. A clustering-based fingerprint method that groups the set of RPs obtaining highest RSS from a predefined set of APs rather than a single one is proposed in [33]. The number of clusters formed by such clustering method is hard to determine. Another similar work proposed in [34], defines the strongest AP for a fingerprint as its important AP (IAP) and selects the offline fingerprints having same IAP as the observed fingerprints in the positioning stage to estimate the target's location. In [35], the researchers consider interference between the APs to determine the positioning abilities of each group of APs and also calculate the group discriminant value for each such group. However, the interference between a set of APs gets prominent if they are closely spaced. Several researches on fingerprint positioning focus on reducing the fingerprint dimension to lessen the positioning overhead. The researchers in [36,37] aim to reduce the number of available APs that can be used for the positioning purpose. In [38], some clustering strategy and decision tree method are applied to choose an optimum set of APs to reduce the computational power and positioning overhead. Some heuristic rules are applied in [39] to remove the useless APs and RSS samples having low contribution in positioning from the radio map to lessen its size and also the positioning overhead. For the same purpose, the researchers in [13] propose an RSS feature reduction method using the principal component analysis (PCA) and also employ kernel ridge regression (KRR) based positioning model for fingerprint positioning. However, the process of filtering out the useless APs [39] and PCA based feature reduction method [13] will be effective only for the positioning areas having unevenly distributed large set of APs.

Various clustering based fingerprint systems are proposed in literature to reduce the positioning overhead. The *k-means* clustering [10] adopting a recursive process to create k disjoint subsets, are used in [38,40,54]. To limit the positioning error provided by such clustering due to random selection of its initial cluster members, two overlapping based clustering methods are proposed in [40]. However, such overlapping methods incur higher positioning overhead [6]. The self-organizing map (SOM), which is a back propagation neural network, is also used in both phases of fingerprint positioning proposed in [12]. However, such positioning method can provide only room-level accuracy. The *affinity propagation* (*AFP*) clustering [11], which can form an optimal set of clusters without selecting the exemplars randomly and having any prior knowledge of the number of clusters to be formed, are applied in [41-43]. However, such clustering cannot be used to a dataset having complex structure. A semi-supervised affinity propagation clustering, which calculates the similarity measure between two RPs by combining the similarity of AP sets with the RSS distance, is proposed in [44] to improve the positioning accuracy. However, such clustering still depends on the set of input preference values. The clustering method proposed in [45] uses the object′s inertial data in the positioning phase to group the RPs in real time and then set up a target area of positioning on the fly. However, such inertial based clustering accumulates significant cumulative error. The researchers in [46] forms the clusters based on the same first k APs with higher weight values which are computed for all RPs using RSS-probability radio map. However, the number of clusters formed by such method cannot be determined statically. In [47], smallest enclosing circle algorithm is employed to group the RPs based on their positional co-ordinates. Such clustering can only provide improved positioning accuracy in case of continuous tracking of the user. The domain based clustering proposed for fingerprint positioning [48] considers each individual AP from the set of all visible APs rather than relying on some AP selection method. A multi-dimensional RSS feature based fuzzy clustering method that selects only the non-redundant APs is proposed in [49]. However, the calculation of AP correlation and AP fuzzy membership involved in such method induces high computational overhead. In [50], AFP clustering is used at first and then the target′s location is estimated based on a hybrid distance which is a combination of the physical distance and the corresponding signal distance between two RPs. The fingerprint system proposed in [51] uses K-medoids clustering to create a set of overlapped clusters, combines the hamming distance between AP coverage vector with the Euclidean distance between the RSS vectors and then apply KRR method in the positioning phase. Such positioning system incurs high computational overhead. The virtual positions of APs are considered in [52] for cluster formation, which may not work in indoor areas having linear constraints. An adaptive hierarchical clustering-based fingerprint system that uses PCA aided matching process to improve the positioning performance, is proposed in [53]. However, such system may not be practically feasible in large indoor settings due to its relatively high computational overhead. In [54], *k-means* clustering is combined with *Bayesian inference* to address the time varying nature of RSS samples. Such system, however, cannot work without knowing the APs' position. Our previ-

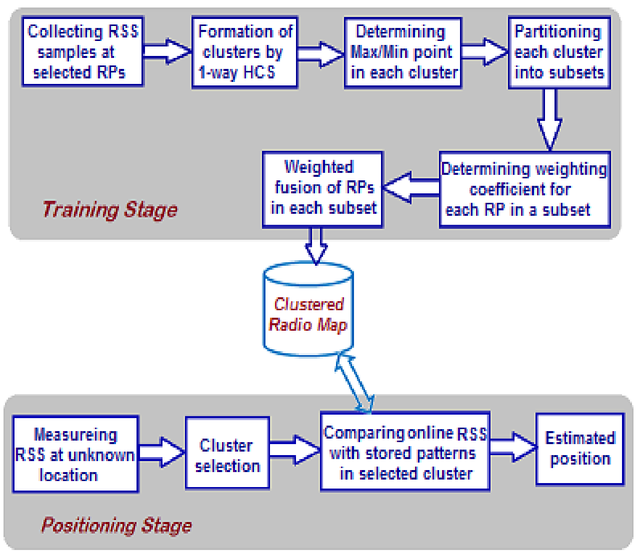


Fig. 1. Block diagram of proposed fingerprint positioning system

ously proposed *hierarchical clustering strategy* (*HCS*) divides the RPs in a hierarchical way based on the set of stronger APs visible at each RP, their order and the level of hierarchy [14]. On the other hand, our recent prior work on clustering based fingerprint system [8] can reduce the storage and positioning overhead by applying an average fusion policy and also aims to improve the positioning accuracy by employing a outlier mitigation technique.

Although, a lot of researches on fingerprint positioning have been conducted over the years to reduce the the positioning overhead, very few works focus on reducing the size of radio map by filtering out the useless APs or using some feature reduction method. Moreover, such methods may not provide satisfactory performance for the indoor areas having properly distributed smaller set of APs as evidenced from the experimental results given in [13]. On the other hand, time-varying nature of the RSS dataset has not been properly addressed by our previously proposed average fusion based efficient clustering. Thus, weighted fusion based efficient clustering is proposed in this article not only to improve the positioning accuracy, but also to reduce positioning and storage overhead.

III. Proposed System

Our proposed system employs a weighted fusion based efficient clustering method and works in two stages (training and positioning) to determine the position. To describe the working methodology of our proposed system, we use several symbols and notations which are summarized in table I. The purpose of applying weighted fusion strategy on the RPs rather than average fusion is to reduce the effect of highly fluctuating RSS values on the positioning accuracy of such system. The block diagram of our proposed positioning system is provided in fig. 1.

A. Training Stage

In this stage, at first, RSS values at several predetermined locations (called reference points) within the positioning area from a set of APs deployed there are measured. A set of RSS samples are collected from each AP for certain duration and then their average is taken as the measured RSS value from that AP due to time-varying nature of the RSS samples. The set of average RSS values measured at a RP from the fixed set of APs deployed in the positioning area constitute the fingerprint for that RP. *Only one fingerprint per reference point is created to maintain one-to-one mapping between a point in the physical space (i.e., RP) and the same in signal space (i.e., RSS vector).* We consider a set of m APs ($AP = \{ap_1, ap_2, \cdots, ap_m\}$) are deployed in the positioning area and n RPs ($L = \{l_1, l_2, \cdots, l_n\}$) are selected within it. The RSS vector observed at location l_i in this stage is denoted by $V_i = [v_{1i},\ v_{2i},\ \cdots, v_{mi}]$, where v_{ji} indicates the average RSS value received from j^{th} AP (ap_j) at l_i. The *radio map* for this positioning area have n records, each of which contains two-dimensional (2D) positional co-ordinates of each RP and RSS vector observed at that RP.

TABLE I
List of Notations & Symbols used in Proposed Clustering Technique

Notation/ Symbol	Meaning
APs	Access Points
m	Total number of APs deployed in the positioning area
n	Total number of reference points selected in positioning area
l_i	i^{th} reference point (RP)
ap_j	j^{th} access point
RM	Radio map
V_i	RSS vector observed at i^{th} RP
v_{ji}	Average RSS value measured at l_i from ap_j
R	RSS vector obtained at unknown location
r_j	Average RSS value observed at unknown location from ap_j
C_j	j^{th} cluster
$n_j = \lvert C_j \rvert$	Number of RPs within the cluster C_j
δ	Value of threshold
$\lvert RM \rvert$	Total number of records within RM
w_p	p^{th} subset within a cluster
SV_p^0	Representative RSS vector of the subset w_p
sv_{jp}^0	j^{th} element of the RSS vector SV_p^0
k_j	Number of subsets within the cluster C_j
w_{kj}	k^{th} subset within the cluster C_j
W_{kj}^a	Weighting coefficient for a RP l_a which lies in subset w_{kj}
SV_{kj}	Conglomerated RSS vector of the subset w_{kj}
sx_{kj}	Conglomerated X co-ordinate of the subset w_{kj}
sy_{kj}	Conglomerated Y co-ordinate of the subset w_{kj}

*1) **Proposed clustering strategy**:* To Construct the compressed clustered radio map, our previously proposed *one-way hierarchical clustering strategy* ($1-way\ HCS$) [14] is used at first, to divide all RPs considered in the positioning field into several disjoint clusters. The proposed $1-way\ HCS$ creates a set of clusters equal to the number of APs deployed in the positioning field and denoted by $\{C_1, C_2, \cdots, C_m\}$ such that $\bigcup_{j=1}^{m} C_j = L$. Each such cluster ($C_j$) created by $1-way\ HCS$ include only those RPs that obtain strongest RSS from only one particular AP (ap_j). Although, the RSS values obtained at some RP from a certain AP varies over time, such measurements collected at the nearby RPs from a set of APs within the positioning area holds *spatial correlation property* as proved theoretically in our earlier work [8]. Based on the above principle, the RPs included in each cluster are further divided into several smaller subsets and then the RPs included within each such subset are conglomerated by employing our proposed weighted fusion strategy. The process of partitioning and data fusion should begin with some suitable RP within the cluster rather than selecting it randomly. Unlike our earlier work, the proposed clustering presented in this paper initiates the process of partitioning and data fusion within a cluster either with the RP receiving maximum signal strength (Max point) or with the RP receiving minimum signal strength (Min point) from the AP associated with its cluster. So, the first subset (w_1) of a cluster (say C_j, where $1 \leq j \leq m$) is formed by either the Max or Min point (say l_a, where $1 \leq a \leq n_j$ and n_j is the number of RPs within C_j) of that cluster. The RSS vector associated with the initial point of each subset is defined as its representative RSS vector and the latter does not change even if more RPs are added to that subset. Next, each of the remaining RPs (say l_b, where $1 \leq b \leq n_j,\ a \neq b$) belonging to the cluster (C_j) will be added to the existing subset (w_1) if the Euclidean distance between the RSS vector (V_b) associated with that RP (l_b) and the subset′s representative RSS vector (SV_1^0) is upper bounded by a certain numerical value called *threshold* (δ). Otherwise, that RP (l_b) will form a new subset (w_2). We apply the following rule iteratively to create the various subsets within a cluster.

$$l_b \in w_p \leftrightarrow \left\{l_b, w_p\right\} \in C_j \wedge \left\|V_b, SV_p^0\right\| \leq \delta,$$
$$where \left\|V_b, SV_p^0\right\| = \sqrt{\sum_{j=1}^{m}\left(v_{ja} - sv_{jp}^0\right)^2}\ and\ p \geq 1$$

After partitioning the cluster into several subsets, the positional coordinates and RSS vectors of the RPs included within each subset are merged according to our proposed weighted fusion method. The proposed fusion method determines a weighting coefficient for each individual RP based on the Euclidean distances between its associated RSS vector and the mean of all RSS vectors belonging to its subset. More close a RP's RSS vector to its subset's average RSS vector, that means, a RP whose RSS vector has smaller Euclidean distance from its subset's average RSS vector, will be assigned larger weight to create the fused RP entry. So, the weighting coefficient for an RP should be proportional to the inverse of the Euclidean distance between its associated RSS vector and its subset's average RSS vector. We, at first, compute the inverse of Euclidean distance (denoted by u_{kj}^c for the RP l_c, which lies in the k^{th} subset of j^{th} cluster denoted as w_{kj}) between each individual RP′s RSS vector and its subset's average RSS vector ($\overline{SV}_{kj}$) by using the following equation.

$$u_{kj}^c = \frac{1}{\left\|V_c, \overline{SV}_{kj}\right\|} \tag{1}$$

To make the summation of all RPs' weighting coefficients included in a subset equal to 1, the weighting coefficient denoted by W_{kj}^c for each individual RP (l_c which lies in the subset w_{kj}) is computed as follows.

$$W_{kj}^c = \frac{u_{kj}^c}{\sum_{c=1}^{h_{kj}} u_{kj}^c}, \tag{2}$$

where h_{kj} is number of RPs in the subset w_{kj}.

According to our proposed weighted fusion method, the conglomerated RSS vector (denoted as SV_{kj}) and positional co-ordinates (denoted as $\left\{sx_{kj},\ sy_{kj}\right\}$) of some subset w_{kj} are determined by the following equations.

$$SV_{kj} = \sum_{a=1}^{h} W_{kj}^a \times V_a, \tag{3}$$

where subset w_{kj} contains the set of RSS vectors $\{V_1, V_2, \cdots, V_h\}$ and W_{kj}^a denotes the weighting co-efficient for RP l_a included in the subset w_{kj}.

$$sx_{kj} = \sum_{a=1}^{h} W_{kj}^a \times x_a, sy_{kj} = \sum_{a=1}^{h} W_{kj}^a \times y_a \tag{4}$$

where $\{[x_1\ y_1], [x_2\ y_2], \cdots, [x_h\ y_h]\}$ represent the positional co-ordinates of h RPs included in the subset w_{kj}. For each subset, only one radio map entry consisting of the conglomerated RSS vector and positional co-ordinates of that subset is created to reduce the amount of storage. Two different weighted fusion based efficient clustering strategies (*WF-ECS*), based on whether the Max point (RP obtaining maximum RSS from the associated AP) or the Min point (RP obtaining minimum RSS from the associated AP) is selected as the initial point of partitioning and data fusion process, are presented in this section. These are named as *WFmax-ECS* and *WFmin-ECS* respectively. Our proposed fusion based clustering strategy reduces the number of fingerprint records in the created radio map since it merges the fingerprint records associated with a subset of RPs included in the same cluster and having similar RSS vector, into a single record. *The resultant radio map created by our proposed clustering-based fingerprint system get reduced in size compared to the original radio map and thus is defined as compressed radio map*. The algorithm for constructing the clustered and compressed radio map using either *WFmax-ECS* or *WFmin-ECS* is presented below.

Algorithm 1: Construction of clustered and compressed radio map using WF-ECS

Input:

i. Set of clusters $[C_1, C_2, \cdots, C_m]$.
ii. Set of positional co-ordinates of all training locations $[(x_1,y_1), (x_2,y_2), \cdots, (x_n,y_n)]$
iii. Set of average *RSS* vectors measured at n RPs $[V_1, V_2, \cdots, V_n]$
iv. Value of threshold (δ)

Output: *A clustered and compressed radio map*

Procedure:

1) For each cluster C_j (j = 1 to m), following iterative process is executed.
   a) $n_j = \left|C_j\right|$ and $L_j = \left\{l^{(1)}, l^{(2)}, \cdots, l^{(n_j)}\right\}$, where L_j is the set of RPs belonging to C_j
   b) $V^{(a)} \leftarrow$ *RSS* vector measured at $l^{(a)}$.
   c) Determine either the Max point (for *WFmax-ECS*) or Min point (for *WFmin-ECS*) within C_j and set it to $l^{(m_j)}$, where $1 \leq m_j \leq n_j$.

d) $k \leftarrow 1$ and $l^{(m_j)} \in w_1$, where w_1 is the first subset within cluster C_j.

e) $SV_1^0 \leftarrow V^{(m_j)}$, where $V^{(m_j)}$ is RSS vector obtained at $l^{(m_j)}$ and SV_1^0 is the representative *RSS* vector of subset w_1.

f) Repeat the following steps for $1 \leq i \leq n_j$ and $i \neq m_j$.

   i) If $\left\|V^{(i)}, SV_p^0\right\| \leq \delta$ for some $p, 1 \leq p \leq k$, then $l^{(i)} \in w_p$

   ii) If $\left\|r^{(i)}, SV_p^0\right\| > \delta$ for all $p, 1 \leq p \leq k$, then a new subset is created as follows.

      A) $k \leftarrow k+1$

      B) $l^{(i)} \in w_k$

      C) $SV_k^0 \leftarrow V^{(i)}$

g) Conglomerated RSS vector (SV_{pj}) of subset w_{pj} is estimated by equations 1 and 2.

h) Conglomerated positional co-ordinates ([sx_{pj} sy_{pj}]) of subset w_{pj} is estimated by equations 1 and 3.

i) A record containing the conglomerated RSS vector and positional co-ordinates of each subset ([sx_{pj} sy_{pj} SV_{pj}]) is included in the **radio map**.

2) The final clustered and compressed radio map is represented as

$$RM = \begin{bmatrix} \left\{ \begin{matrix} (sx_{11}, sy_{11}), & SV_{11} \\ \vdots & \vdots \\ (sx_{k_1 1}, sy_{k_1 1}), & SV_{k_1 1} \end{matrix} \right\} C_1 \\ \left\{ \begin{matrix} (sx_{12}, sy_{12}), & SV_{12} \\ \vdots & \vdots \\ (sx_{k_2 2}, sy_{k_2 2}), & SV_{k_2 2} \end{matrix} \right\} C_2 \\ \vdots \\ \left\{ \begin{matrix} (sx_{1m}, sy_{1m}), & SV_{1m} \\ \vdots & \vdots \\ (sx_{k_m m}, sy_{k_m m}), & SV_{k_m m} \end{matrix} \right\} C_m \end{bmatrix}, \quad (5)$$

where m is the number of available APs in the positioning area and k_j is number of generated subsets in j^{th} cluster (C_j, $1 \leq j \leq m$). Total number of records within the **radio map** is determined as follows.

$$|RM| = \sum_{j=1}^{m} k_j \quad (6)$$

B. Positioning Stage

The positioning stage of both fingerprint systems based on our two newly proposed clustering techniques, i.e., *WFmax-ECS* and *WFmin-ECS*, works in two steps as shown in fig. 1. In the first step, most appropriate cluster for the RSS vector observed at the unknown location is selected by determining the index of its element having the maximum value. In the second step, the RSS vector of unknown location, at first, is compared with the radio map entries belonging to the selected cluster only and then the positional coordinates associated with the best matched entry within the selected cluster is provided as estimated position of the unknown location. The positioning algorithm is given below.

Algorithm 2: Positioning Algorithm

Input:

i. Compressed clustered *radio map*

ii. Average *RSS* vector observed at unknown location ($R = [r_1, r_2, \cdots, r_m]$, where m is number of *APs*)

Output: Estimated position ($[x' \ y']$)

Procedure:

1) Selected cluster $\leftarrow C_i$, if following condition satisfies $\forall j \exists i((1 \leq j, i \leq m) \wedge (j \neq i)) \rightarrow (r_i \geq r_j)$.

2) $RM_i \leftarrow$ records of radio map belonging to cluster C_i.

3) $[x' \ y'] = \left[sx_{qi} \ sy_{qi}\right] \leftrightarrow SV_{qi} = \arg\min_{SV_{pi}} \left\|R, \ SV_{pi}\right\|$, where $q \neq p$, $1 \leq q, p \leq k$ and k is number of subsets within C_i.

The searching overhead of our newly proposed clustering strategy *WF-ECS* in the positioning stage remain same as that of our previously proposed average fusion based clustering method given in [8] and it is $O(t_{cs} + \bar{k_j})$, where t_{cs} denotes some fixed duration needed to select an appropriate cluster and

IV. Performance Analysis

In this section, we evaluate the performances of our two proposed clustering methods, *WFmax-ECS* and *WFmin-ECS*, in terms of three performance metrics which are *average positioning error*, *average positioning time* and *percentage of reduction in storage*. We also compare their performances with the existing *k-means* clustering [10], *AFPC* [11], *SOM* based clustering [12], recently proposed PCA-KRR [13] and also our previously proposed *AF-ECS* [8] and $2-way$ *HCS* [14] in the simulation test bed as well as using the dataset collected from two different real indoor environments.

A. Performance Metrics

Various performance metrics used to evaluate and compare the performances of our proposed and other existing systems have been defined below.

i. **Average Positioning Error:** The distance between the actual position of an object and its estimated position is stated as the *positioning error* obtained by the given positioning system. So, the *average positioning error* (**APE**) of a positioning system is determined by the following equation.

$$APE = \frac{1}{q} \sum_{l=1}^{l=q} \sqrt{(x_l' - x_l)^2 + (y_l' - y_l)^2}, \quad (7)$$

where q is number of test points (TPs) whose positions are needed to be estimated, $[x_l' \, y_l']$ and $[x_l \, y_l]$ denote the estimated position and actual position of l^{th} ($1 \leq l \leq p$) TP respectively.

ii. **Average Positioning Time:** The duration of time gap between the instance estimated position obtained and the instance positioning request made is defined as the positioning time. So, the *average positioning time* (**APT**)

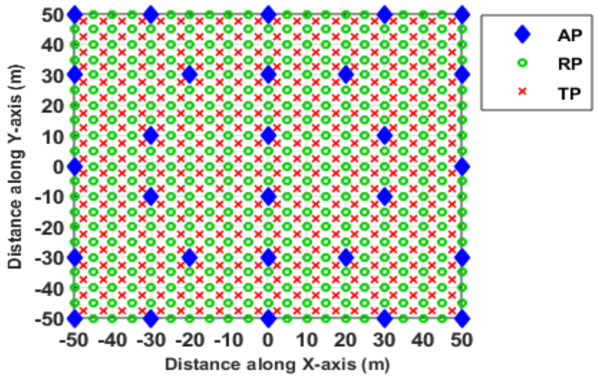


Fig. 2. Simulation test bed

experienced by a positioning system is determined as follows.

$$APT = \frac{1}{q}\sum_{l=1}^{l=q} pt_l, \tag{8}$$

where pt_l is estimated positioning time for l^{th} TP and q is number of test points whose positions have been estimated by the given positioning system.

iii. **Percentage of Reduction in Storage:** With the increasing value of percentage of reduction in storage (**PRS**) the amount of storage required by the radio map of a fingerprint positioning system decreases. It is calculated by the following equation.

$$PRS = \left(1 - \frac{|RM|}{n}\right) * 100, \tag{9}$$

where $|RM|$ refers the number of fingerprint records stored within the radio map of some fingerprint system and n is number of RPs considered within the positioning area.

B. *Simulation based Performance Analysis*

We consider an area of size 100×100 $meter^2$ as the positioning area for the simulation test bed. We assume existence of several horizontal and vertical walls along the line segments $[-50, i\times10]$, $[50, i\times10]$ and $[i\times10, -50]$, $[i\times10, 50]$ ($-5 \leq i \leq 5$) respectively in the simulated positioning area as shown by thin grey color lines in fig. 2. Fig. 2 also explicitly shows the positions of the available APs as well as the selected RPs and test points within the simulation test bed. The *training grain size* (TGS), i.e., the distance between two consecutive RPs is set to 5 meter and the same between two TPs is also kept the same. Unlike our earlier work [8], a set of non-equidistant APs are deployed in the simulation test bed as usually found in any real indoor settings. To simulate various characteristics of the signal propagation usually observed in real indoor areas like non-isotropic propagation, dynamic signal fading, signal attenuation due to presence of walls/obstacles etc. and also the effect of hardware variance, a modified version of radio irregularity model (RIM) [55] is integrated into log-distance path loss model based simulation test bed presented here. So, the RSS values at different RPs and test points within the simulation area are obtained by applying the following equation.

$$\begin{aligned} P_r(l', ap_j) = P_t^{vsp}(ap_j) - PL^{doi}(l', ap_j) \\ - PL^{waf}(l', ap_j) + n(0, \sigma), \end{aligned} \tag{10}$$

where $P_t^{vsp}(ap_j)$ denotes the transmitted power of j^{th} AP (ap_j), $PL^{doi}(l', ap_j)$ denotes the path loss due to the presence of radio irregularities in the wireless channel at location l' from ap_j, $PL^{waf}(l', ap_j)$ denotes the same at location l' due to the attenuation effects of walls and other obstacles between location l' and ap_j and $n(0, \sigma)$, a zero-mean normally distributed random variable with standard deviation σ, denotes the noise incorporated into the measured RSS values. Since the transmitted power of different APs varies depending on the hardware, thus it is calculated based on the values of the parameter named as variance of sending power (VSP) by using the following equation.

$$P_t^{vsp}(ap_j) = P_t \times (1 + Q(0, vsp)), \tag{11}$$

where P_t denotes a constant transmitted power and $Q(0, vsp)$ is a normally distributed random variable with mean value 0 and a standard deviation vsp. The amount of path loss in different directions in RIM is determined using values of the parameter called degree of irregularity (DOI) and it is obtained as follows.

$$PL^{doi}(l',\ ap_j) = PL(\|l',\ ap_j\|) \times k_i,\ 0 < i < 360 \tag{12}$$

where $PL(\|l',\ ap_j\|)$ denotes the obstacle-free path loss from j^{th} AP (ap_j)at location l' and the difference in path loss in different directions is computed using the coefficient k_i. The values of k_i can be determined as follows.

$$k_i = \begin{cases} 1, & \text{if } i = 0, \\ k_{i-1} \pm rand \times doi, & \text{otherwise.} \end{cases} \tag{13}$$

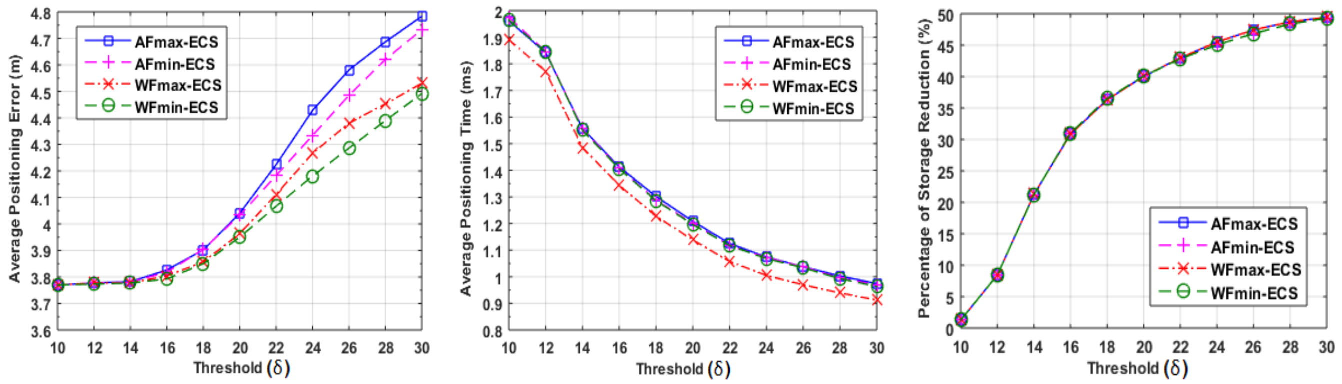
where $0 < i < 360$, $|k_0 - k_{359}| \leq doi$ and $rand$ takes values from range $[0, 1]$. At location l', the obstacle free path loss from j^{th} AP is computed as follows.

$$\begin{aligned} PL\left(\|l',\ ap_j\|\right)[dB] = PL\left(d_0\right) + 10\gamma \cdot log\left(\frac{\|l', ap_j\|}{d_0}\right) \\ + n(0,\ \sigma), \end{aligned} \tag{14}$$

where d_0 is the reference distance which is set to 1, $\|l',\ ap_j\|$ denotes the distance between j^{th} AP and location l', γ is the path loss exponent, $n(0,\ \sigma)$ is a normally distributed random variable with zero mean and standard deviation σ to denote the amount of noise. The path loss due to presence of walls or other obstacles in RIM is determined as follows.

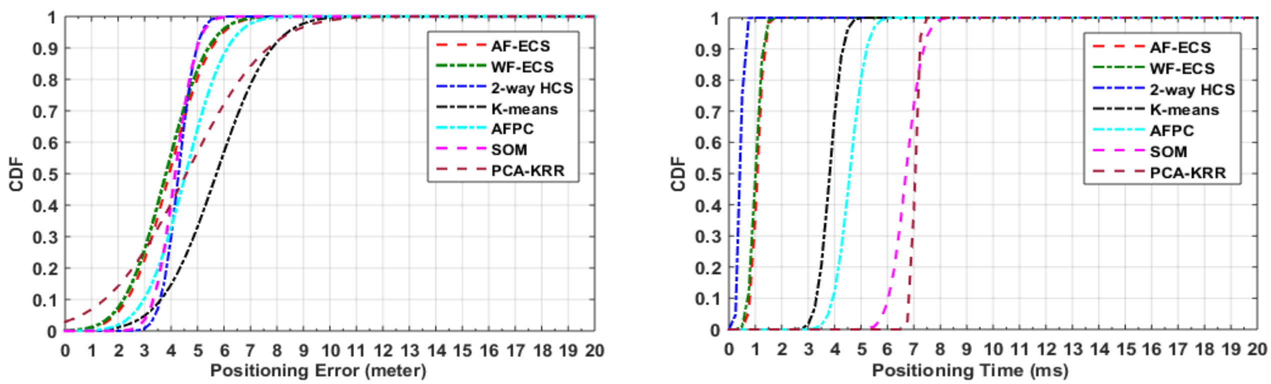
$$PL^{waf}(\|l',\ ap_j\|) = min(n_{obs},\ n_{max}) \times waf, \tag{15}$$

where n_{obs} is the number of walls or obstacles along the line-of-sight path between location l' and j^{th} AP, n_{max} denotes the maximum number of obstacles and waf is the amount of signal attenuation caused by one wall. The default values of various parameters used to model the RSS values using above mentioned equation are mentioned below.
$P_t = 20$ dBm, $D_0 = 1$m, $PL(D_0) = 37.3$ dBm, $\gamma = 4$, $\sigma = 4$, $VSP = 0.2$, $DOI = 0.02$, $WAF = 3$, $N_{max} = 4$.



(a) Average positioning error vs threshold
(b) Average positioning time vs threshold
(c) Percentage of reduction in storage vs threshold

Fig. 3. Performance comparisons of proposed techniques under different values of threshold based on simulation results.



(a) Distributions of positioning error
(b) Distributions of positioning time

Fig. 4. Comparisons of PE and PT of the proposed and 6 other techniques in terms of CDF based on simulation results.

TABLE II
COMPARISONS OF PE & PT ACQUIRED BY VARIOUS TECHNIQUES BASED ON SIMULATION RESULTS

Positioning System	PE(meter)		PT(ms)	
	50%	95%	50%	95%
AF-ECS	3.9	6.0	1.1	1.4
WF-ECS	3.8	5.8	1.0	1.4
2-way HCS	4.3	5.2	0.4	0.7
k-means	5.7	8.5	3.8	4.4
AFPC	4.6	6.7	4.6	5.4
SOM	4.2	5.2	6.7	7.6
PCA-KRR	4.6	8.7	7.0	7.2

We compare, at first, the performances of our proposed clustering strategies, *WFmax-ECS* and *WFmin-ECS* with two similar variants of our previously proposed *AF-ECS* named as *AFmax-ECS* and *AFmin-ECS* with respect to the different values of threshold (δ) to demonstrate the effectiveness of the former (*WF-ECS*) over the latter (*AF-ECS*) and also to select the optimum value of threshold for our proposed strategies. From the set of different values of threshold considered here, one particular value at which our proposed strategies provide lower APE and APT while reduces the amount of storage by some extent, is selected as the optimum threshold value. This optimum value is also made as the default value of threshold for our proposed strategies to further compare their performances with the other existing methods. The performance comparison of our proposed strategies in terms of **APE**, **APT** and *percentage of reduction in storage* with respect to different values of threshold are depicted in figs $3(a)$, $3(b)$ and $3(c)$ respectively. To determine the appropriate value for threshold, at first our proposed method is evaluated against a broad range of values on coarse scale for this parameter (say $[10, 20, \cdots, 90, 100]$). Based on the required positioning performances, two intermediate values within the above-said broad range are chosen for it and then our proposed clustering method is further evaluated on a finer scale within the selected intermediate values of threshold to find out a more appropriate value for this parameter. Figs. $3(a)$ and $3(b)$ depict that our proposed strategies *WFmin-ECS* and *WFmax-ECS* achieve comparatively smaller APE and APT respectively with the increasing value of threshold in the range of $[10 - 30]$. Fig. $3(c)$, on the other hand, depicts that the performances of all four strategies remain same in terms of percentage of reduction in storage, which increases with the increasing value of threshold. Since, our newly proposed weighted fusion policy creates a more perfect match between the fused RSS pattern and the fused positional co-ordinates compared to that provided by previously proposed average fusion policy, hence, both schemes of *WF-ECS* reduce APE and APT compared to those of *AF-ECS* with the increasing value of threshold. It is

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2022.3225796

TABLE III
COMPARISONS OF **APE** (METER) W.R.T. DIFFERENT VALUES OF **TGS**.

TGS (m)	*AF-ECS*	*WF-ECS*	2-*way HCS*	*KM*	*AFPC*	*SOM*	*PCA-KRR*
2	2.88	2.37	2.62	2.65	4.5	2.6	4.42
4	3.28	3.12	3.64	4.67	4.69	3.5	4.6
6	4.55	4.44	4.93	6.17	4.91	4.85	9.6
8	5.99	5.89	6.27	8.85	6.46	6.28	11.38

TABLE IV
COMPARISONS OF **APT** (MS) W.R.T. DIFFERENT VALUES OF **TGS**.

TGS (m)	*AF-ECS*	*WF-ECS*	2-*way HCS*	*KM*	*AFPC*	*SOM*	*PCA-KRR*
2	2.0	2.0	2.3	10.0	9.7	9.3	9.7
4	1.3	1.2	0.6	4.5	5.8	7.0	8.1
6	1.0	0.9	0.3	3.5	3.8	6.8	7.8
8	0.7	0.6	0.2	3.1	2.8	6.0	7.7

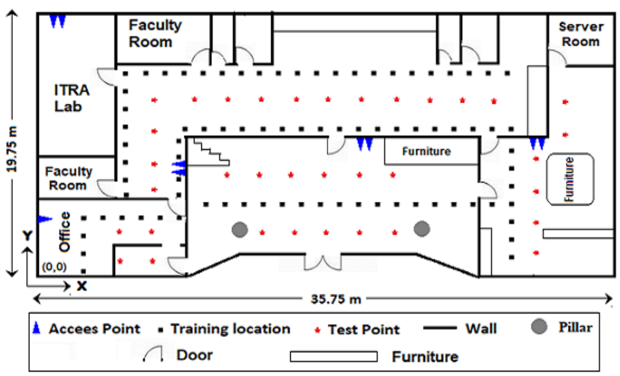


Fig. 5. Experimental area inside departmental building of School of Mobile Computing & Communication.

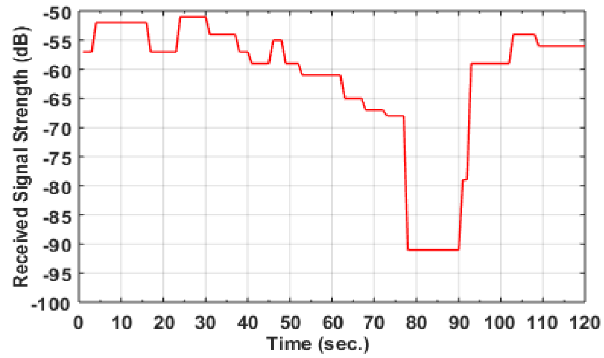


Fig. 6. Distribution of RSS over time from a certain Wi-Fi AP deployed inside the departmental building of SMCC, JU

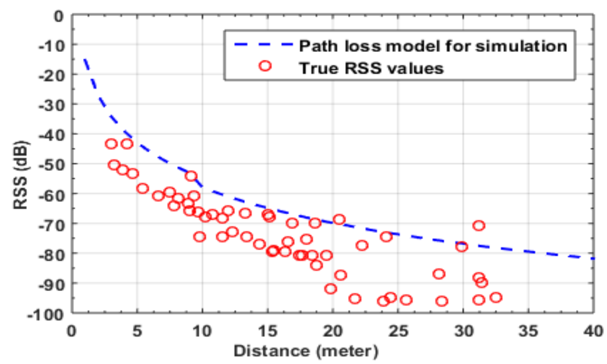
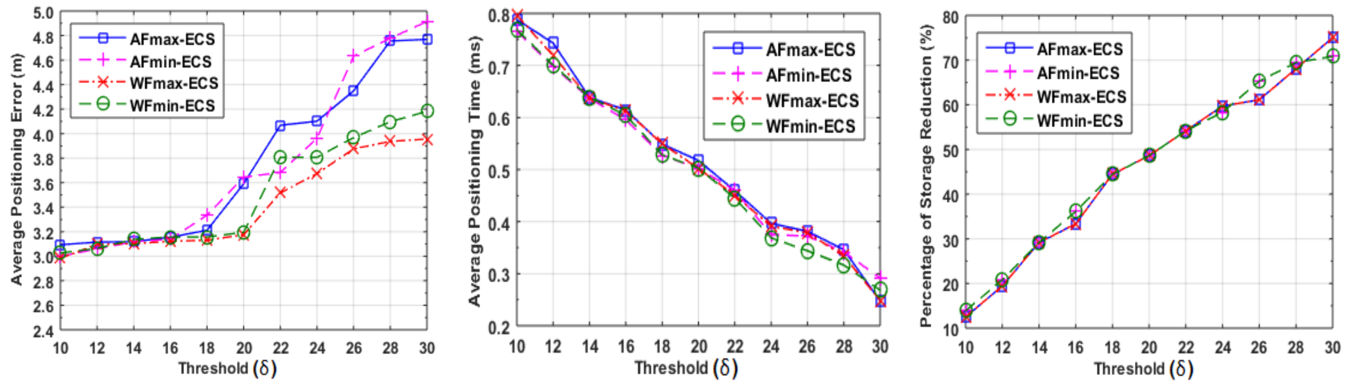


Fig. 7. Values of RSS vs distance from a certain Wi-Fi AP of experimental area

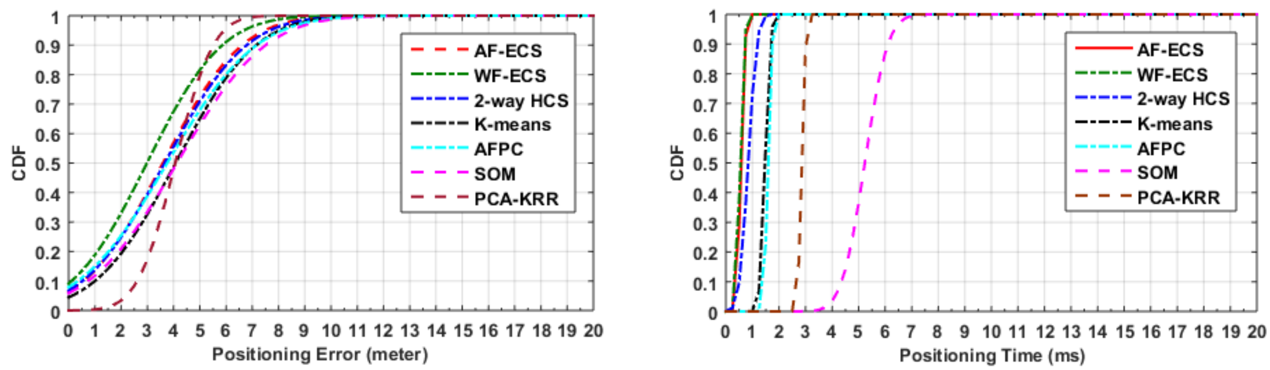
also found from figs. $3(a)$, $3(b)$ and $3(c)$ that newly proposed *WF-ECS* acquires better performances at the threshold value of 18 compared to the previously proposed *AF-ECS*, so the default threshold value for our proposed schemes is set to 18 to compare their performances with the other contemporary methods. Since positioning accuracy is more important performance metric compared to other metrics for any positioning system, hence *WFmin-ECS* is selected as the best performing strategy among the four proposed strategies in the simulation test bed. Fig. $4(a)$ illustrates the cumulative distribution of positioning error (**PE**) acquired by our proposed technique and other existing techniques considered in this paper. The same for the positioning time (**PT**) incurred by various techniques is illustrated in fig. $4(b)$. Moreover, table II provides average (50%) as well as 95% PE and PT incurred by these techniques. It is evident from the comparative results given in figs. $4(a)$, $4(b)$ and table II that the performances of our proposed method *WF-ECS*, in terms of PE, is comparatively better than our previously proposed *AF-ECS* and other existing methods considered in this paper. In terms of positioning time, our previously proposed $2-way$ HCS achieves best performance, whereas the proposed *WF-ECS* and *AF-ECS* have almost similar performances which are better than that of other contemporary methods.

The comparative performance analysis of our proposed methods (*AF-ECS* and *WF-ECS*) and other existing methods in terms of *APE* and *APT* with respect to different values of *training grain size* (*TGS*) are provided in table III and IV respectively. The results given in table III shows that our newly proposed *WF-ECS* achieves better performance compared to the other existing methods and APE for each fingerprint method increases with the increasing value of TGS. This is so because the number of RPs reduces with the increasing value of *TGS*. Table IV, on the other hand, shows that, in terms of APT, our previously proposed $2-way$ *HCS* achieves best performances at the higher values of *TGS*, whereas our proposed *WF-ECS* and *AF-ECS* have almost similar performances which are better than that of other existing methods under different values of *TGS*. With the increasing value of *TGS*, number of RPs in the positioning field reduces. So, the positioning time for each fingerprint technique also reduces with the increasing value of *TGS*. Since all algorithms including our proposed one and other existing techniques considered in this paper have been implemented according to their operating principles and also using the same platform, hence their performance comparison in terms of APT should be regarded as fair comparison of positioning overhead incurred by various fingerprint positioning techniques. From the above comparative performance analysis in terms of positioning error and positioning time, it can be inferred that our newly proposed *WF-ECS* acquires better positioning accuracy compared to the other methods considered in this paper as



(a) Average positioning error vs threshold (b) Average positioning time vs threshold (c) Percentage of reduction in storage vs threshold

Fig. 8. Performance comparisons of proposed techniques under different values of threshold based on real experimental results.



(a) Distributions of positioning error (b) Distributions of positioning time

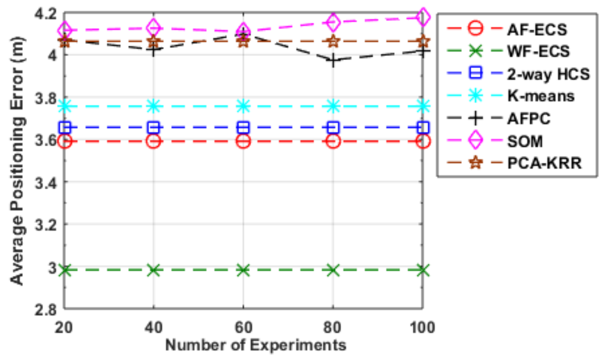
Fig. 9. Comparisons of positioning accuracy and positioning time of 7 fingerprinting techniques including proposed *WF-ECS* in terms of CDF based on experimentation in departmental building.

well as lower positioning overhead at more than 35% reduced storage requirement than those except our previously proposed $2-way\ HCS$.

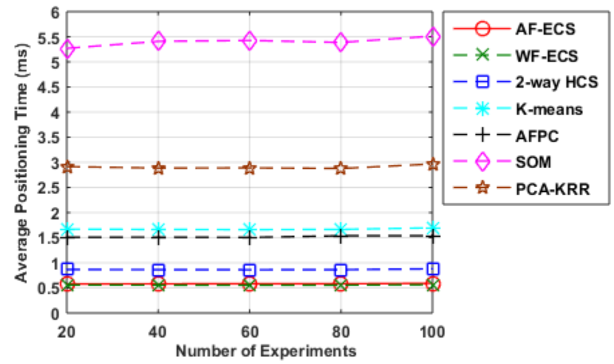
C. Performance Analysis using Dataset Collected in Departmental Building

The performances of our newly proposed efficient clustering *WF-ECS* have been analyzed and compared with the other techniques using the same dataset collected in our departmental building and also used in our previous work [8] to demonstrate its effectiveness in the real indoor setting. The floor map of the experimental site is depicted by fig. 5 which also shows the positions of APs deployed there, selected 84 RPs and 35 test points along with presence of other obstacles in the experimental site. Most pairs of the consecutive RPs are separated by a distance of 1.2 meter, whereas a few pairs of them have a distance of 1.8 meter. The process of RSS data collection have been carried out during the office hours throughout a duration of several months. So, influences of various external factors like variation in weather parameters, presence & movements of human beings etc., on the collected RSS values have been considered. The distribution of RSS values collected over a time span of 2 minutes from a particular Wi-Fi AP deployed within our departmental building is shown in fig. 6, whereas the RSS distribution at various distances from a particular Wi-Fi AP within our departmental building along with *log-distance path loss model* used for simulation are shown in fig. 7. The performances of two variants of our newly proposed and previously proposed efficient clustering strategies, i.e., *WF-ECS* and *AF-ECS* are compared, at first, in terms of various performance metrics with respect to the different values of threshold in the range of $[10-30]$ in figs. $8(a)$, $8(b)$ and $8(c)$. Figs. $8(a)$ and $8(b)$ show that our proposed methods *WFmax-ECS* and *WFmin-ECS* acquire lower APE and APT respectively than the other proposed methods with the increasing value of threshold. Fig. $8(c)$ depicts that all four proposed methods provide almost similar performances in terms of *percentage of storage reduction* which increases with the increasing value of threshold. Considering positioning accuracy as more important performance metric compared to the others, we have selected *WFmax-ECS* as the best performing strategy in this real indoor settings. The comparative experimental results given in figs. $8(a)$, $8(b)$ and $8(c)$ show that our newly proposed *WF-ECS* achieves optimum performances at the threshold value of 20 and so, the default threshold value for our proposed strategies is set to 20 to compare their performances with the other contemporary methods considered in this paper.

Fig. $9(a)$ shows the cumulative distribution of positioning



(a) **APE** vs no. of experiments



(b) **APT** vs no. of experiments

Fig. 10. Comparisons of **APE** and **APT** w.r.t different number of experiments conducted in real environment of departmental building.

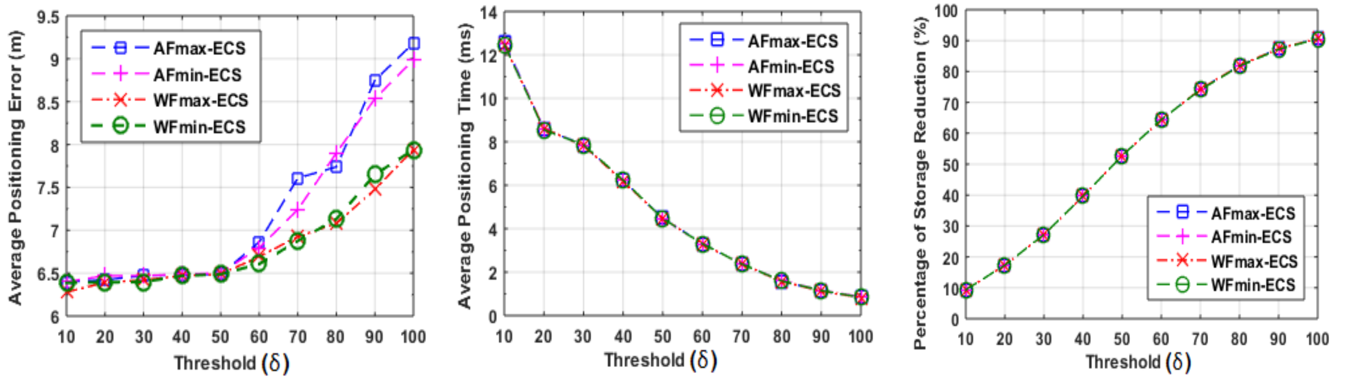
TABLE V
COMPARISONS OF PE & PT ACQUIRED BY VARIOUS METHODS IN REAL INDOOR AREA OF DEPARTMENTAL BUILDING

Positioning system	PE(meter)		PT(ms)	
	50%	95%	50%	95%
AF-ECS	3.6	7.5	0.6	1.0
WF-ECS	3.0	6.7	0.5	0.8
2-way HCS	3.7	7.8	0.8	1.4
k-means	4.1	8.2	1.5	1.7
AFPC	3.8	8.3	1.6	1.9
SOM	4.1	8.5	5.2	6.4
PCA-KRR	4.1	6.0	2.9	3.1

error (**PE**) acquired by 7 clustering methods including our proposed one based on the experimentation conducted using the dataset collected in our departmental building. The same for the positioning time (**PT**) incurred by various techniques is shown in fig. $9(b)$. Moreover, table *V* provides average (50%) as well as 95% PE and PT incurred by those methods including our proposed one. It is evident from the comparative results given in figs. $9(a)$, $9(b)$ and table *V* that our newly proposed clustering method *WF-ECS* outperforms the contemporary methods including our previously proposed two clustering strategies in terms of positioning error and positioning time. Figs. $10(a)$ and $10(b)$, on the other hand, illustrates the performance comparisons of our newly proposed clustering strategy *WF-ECS* and the other existing methods considered here in terms of **APE** and **APT** respectively under the different number of experiments (varying in $[20-100]$) conducted using the dataset collected in our departmental building. The above experimental results also demonstrate that our newly proposed clustering strategy *WF-ECS* acquires lower positioning error and positioning overhead compared to various other techniques including our two previously proposed methods even while reducing the amount of storage required for the radio map by almost 50%. Moreover, both variants of our proposed efficient clustering (*AF-ECS* and *WF-ECS*) provides better positioning performances in terms of positioning accuracy and positioning overhead compared to the contemporary methods. This happens because the fusion process subsides the ambiguity of selecting the best matched RP for the RSS vector observed at the unknown location. Moreover, the weighted fusion of a subset of RPs having similar RSS pattern based on the inverse of Euclidean distance of each individual RSS pattern from the average RSS pattern of its subset creates a more perfect conglomerated RP entry in the real indoor settings compared to the average fusion of the same.

D. *Performance Analysis using Benchmark dataset*

To demonstrate the effectiveness of our proposed methods in other real indoor areas, we have evaluated and also compared their performances with the contemporary methods using the benchmark dataset, i.e., *UJIndoorLoc dataset* provided in [9]. To determine the optimum threshold value and the best performing efficient clustering strategy for the *UJIndoorLoc dataset* among our four proposed strategies, we evaluate and compare their performances in terms of APE, APT and *percentage of storage reduction* under different values of threshold in the range of $10-100$. Since the separating distance between two consecutive RPs located on the same floor of UJIndoor building is much larger than that of the experimental site in our departmental building, hence higher ranges of threshold values have been considered for the *UJIndoorLoc dataset*. Fig. $11(a)$ shows that our proposed strategy *WFmin-ECS* achieves lower positioning error compared to our other proposed strategies with the increasing values of threshold. The performances of all four proposed strategies remain same in terms of APT and percentage of reduction in storage under the different values of threshold as shown in figs $11(b)$ and $11(c)$ respectively. The comparative results given in figs. $11(a)$, $11(b)$ and $11(c)$ depicts that *WFmin-ECS* will be selected as the best performing strategy in terms of positioning accuracy for the *UJIndoorLoc dataset* and it achieves optimum performances at the threshold value of 50. So, the default threshold value for our both proposed strategies (*AF-ECS* and *WF-ECS*) is set to 50 to compare their performances with the other existing methods using the benchmark dataset. Table *VI* provides average (50%) as well as 95% PE and PT incurred by seven fingerprint techniques including our proposed one based on the experimentation conducted using the *UJIndoorLoc dataset*. It is evident from the comparative results given in figs. $11(a)$, $11(b)$, $11(c)$ and table *VI* that both the proposed clustering techniques *WF-ECS* and *AF-*



(a) Average positioning error vs threshold (b) Average positioning time vs threshold (c) Percentage of reduction in storage vs threshold

Fig. 11. Performance comparisons of proposed techniques under different values of threshold using UJIndoorLoc dataset.

TABLE VI
COMPARISONS OF PE & PT ACQUIRED BY VARIOUS METHODS USING UJINDOORLOC DATASET.

Positioning system	PE(meter) 50%	PE(meter) 95%	PT(ms) 50%	PT(ms) 95%
AF-ECS	6.5	17.5	7.6	16
WF-ECS	6.5	17.3	7.7	16.2
2-*way HCS*	6.7	18.5	10.9	23
k-means	7.6	20.2	16.1	21.5
AFPC	9.1	27.5	23.7	27.5
SOM	6.8	19.7	31.4	49
PCA-KRR	7.1	19.8	22.6	35.2

ECS achieve lower positioning error than the other existing techniques considered here while reducing the storage requirement by more than 50% and our newly proposed clustering technique *WF-ECS* provides better positioning accuracy than our previously proposed technique *AF-ECS* with the increasing value of threshold.

V. CONCLUSIONS AND FUTURE WORK

An weighted fusion based *efficient clustering strategy* for RSS based fingerprint positioning system has been proposed in this paper. The proposed clustering method aims to improve the positioning accuracy of RSS fingerprint method compared to contemporary methods by addressing the problem of inherent stochasticity in the RSS measurements collected in any real setting. Apart from improving the positioning accuracy, it also aims to reduce the amount of storage required for the radio map and also the positioning overhead. The proposed *WF-ECS* employs a novel weighted fusion strategy to conglomerate the similar RSS patterns belonging to the same cluster in order to improve the performances of fingerprint positioning system. The proposed weighted fusion policy assigns larger weight to a RP whose RSS pattern has smaller Euclidean distance from the average RSS pattern of the subset. Thus, it creates a more perfect match between the fused positional co-ordinates and the fused RSS pattern compared to our previously proposed average fusion strategy. The comparative performance analysis carried out in the simulation test bed and real indoor settings between our newly proposed strategy *WF-ECS* and six other existing fingerprint methods demonstrates the effectiveness of our newly proposed method to provide improved positioning accuracy over the others. The extensive experimentation carried out using dataset of departmental building depicts that our proposed method outperforms contemporary methods in terms of positioning accuracy and positioning overhead even by reducing the amount of storage needed for the radio map by about 50%. Moreover, the experimental results obtained from both simulation test bed and real indoor areas demonstrate that the newly proposed *WF-ECS* provides better performances in terms of positioning error and positioning time in real indoor areas compared to that of simulation test bed. This happens because the proposed *WF-ECS* can better address the stochasticity of the collected RSS measurements in real indoor areas where it is more prevalent.

As part of future research, we have a plan to redesign our proposed positioning scheme to provide either *2D* with floor *ID* or *3D* location estimate while evaluating it within a multi-floor building. We also strongly intend to carry out further researches in future to enable our proposed indoor positioning method to locate or track a moving user. Moreover, evaluating the performances of our proposed strategies using various public RSS datasets as given in [56] to demonstrate their feasibility and to devise a closed-loop solution for finding out optimum threshold value under different datasets is another goal of our future researches.

REFERENCES

[1] P. Sadhukhan, N. Mukherjee, P. K. Das, ″Location-Based Services for Smart Living in Urban Areas,″ In Precision Positioning with Commercial Smartphones in Urban Env., Springer, 2021, pp. 53-69.
[2] M. Lemmens, ″Global Navigation Satellite Systems and Inertial Navigation,″ In Geo-information. Geotech. and the Env., Vol 5. Springer, 2011.
[3] S Subedi, JY Pyun, ″A Survey of Smartphone-Based Indoor Positioning System Using RF-Based Wireless Technologies,″ Sensors(Basel). 2020 Dec 17;20(24):7230.
[4] P. Sadhukhan, and P. K. Das. ″MGALE: A modified geometry-assisted location estimation algorithm reducing location estimation error in 2D case under NLOS environments,″ In Int'l Workshop on Mobile Entity Loc. and Tracking in GPS-less Env., pp. 1-18. Springer, 2009.
[5] K. Kaemarungsi and P. Krishnamurthy, ″Modeling of indoor positioning systems based on location fingerprinting,″ Proc. of 23th Annu. Joint Conf. IEEE Comput. Commun. Soc., Hong Kong, 2004, pp. 1012–1022.
[6] P. Sadhukhan, ″Performance Analysis of Clustering-based Fingerprinting Localization Systems″, Wireless Netw., Vol. 25, Issue 5, pp. 2497–2510, July 2019, Springer US.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/TWC.2022.3225796

[7] P. Sadhukhan, K. Dahal and Z. Pervez, ″Impact of beacon coverage on clustering strategies for fingerprinting localization system,″ Proc. of Int'l Conf. on Comput., Networking and Comm, CA, 2017, pp. 184-188.

[8] P. Sadhukhan et al. (2021). An efficient clustering with robust outlier mitigation for Wi-Fi fingerprint based indoor positioning. Applied Soft Computing, 107549.

[9] J. Torres-Sospedra et al., ″UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems,″ in Proc. of Int. Conf. Indoor Pos. and Indoor Navig.(IPIN), 2014, pp. 261–270.

[10] E. Gokcay and J. Principe, ″Information Theoretic Clustering,″ IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 2, Feb. 2002, pp. 158-172.

[11] B.J. Frey and D. Dueck, ″Clustering by Passing Messages Between Data Points,″ Science, vol. 315, no. 1, Feb. 2007, pp. 972-976.

[12] L. Mengual et al., ″Clustering-based location in wireless networks,″ Expert Systems with Applications 37, 2010, pp. 6165-6175.

[13] Y. Le et al., ″Fingerprinting Indoor Positioning Method Based on Kernel Ridge Regression with Feature Reduction,″ Wireless Comm. and Mobile Comput., vol. 2021, Article ID 6631585, 12 pages, 2021.

[14] A. Saha and P. Sadhukhan, ″A novel clustering strategy for fingerprinting-based localization system to reduce the searching time,″ Proc. of IEEE 2nd Int'l Conf. on Recent Trends in Information Syst., 2015, Kolkata, pp. 538-543.

[15] M. Zhou et al., ″Semi-Supervised Learning for Indoor Hybrid Fingerprint Database Calibration With Low Effort,″ in IEEE Access, vol. 5, pp. 4388-4400, 2017.

[16] S. Jung and D. Han, ″Automated Construction and Maintenance of Wi-Fi Radio Maps for Crowdsourcing-Based Indoor Positioning Systems,″ in IEEE Access, vol. 6, pp. 1764-1777, 2018.

[17] C. Laoudias et al. Indoor Quality-of-position Visual Assessment Using Crowdsourced Fingerprint Maps. ACM Trans. Spatial Algorithms Syst., 7, 2, Article 10 (February 2021)

[18] L. Xi et al., A radio map self-updating algorithm based on mobile crowd sensing, Jour. of Netw. and Comp. Appl., Vol. 194, 2021, 103225.

[19] L. Zheng et al., ″A Deep-Learning-Based Self-Calibration Time-Reversal Fingerprinting Localization Approach on Wi-Fi Platform,″ IEEE Internet of Things Jour., vol. 7, no. 8, Aug. 2020, pp. 7072-7083.

[20] Y. Ji et al. 2006. ARIADNE: a dynamic indoor signal map construction and localization system. In Proc. of 4th Int'l Conf. on Mobile syst., appl. and services(MobiSys ′06), NY, USA, pp. 151–164, 2006.

[21] J. Talvitie, M. Renfors and E. Lohan, ″Distance-Based Interpolation and Extrapolation Methods for RSS-Based Localization With Indoor Wireless Signals,″ in IEEE Trans. on Vehicular Tech., vol. 64, no. 4, pp. 1340-1353, April 2015.

[22] J. Bi et al. Fast Radio Map Construction by using Adaptive Path Loss Model Interpolation in Large-Scale Building. Sensors (Basel). 2019 Feb 10;19(3):712.

[23] Z. Wang et al., ″Fast Construction of the Radio Map Based on the Improved Low-Rank Matrix Completion and Recovery Method for an Indoor Positioning System,″ Jour. of Sensors, Vol. 2021, Article ID 2017208, 12 pages.

[24] I. Silva et al., ″Quantifying the Degradation of Radio Maps in Wi-Fi Fingerprinting,″ 2021 Int'l Conf. on Indoor Pos. and Indoor Navigation (IPIN), 2021, pp. 1-8.

[25] D. Li et al., ″A feature scaling based k-nearest neighbor algorithm for indoor positioning system,″ 2014 IEEE Global Comm. Conf., 2014, pp. 436-441.

[26] Y. Fu et al., ″An Indoor Localization Algorithm Based on Continuous Feature Scaling and Outlier Deleting,″ in IEEE Internet of Things Jour., vol. 5, no. 2, pp. 1108-1115, April 2018.

[27] H. Li et al, ″TILoc: Improving the Robustness and Accuracy for Fingerprint-Based Indoor Localization,″ IEEE Internet of Things Jour., vol. 7, no. 4, April 2020. pp. 3053-3066.

[28] P. Roy et al., 2021. Novel weighted ensemble classifier for smartphone based indoor localization. Expert Syst. with Appl., 164, p.113758.

[29] A. Panja et al., A novel feature based ensemble learning model for indoor localization of smartphone users, Engg Appl. of Artificial Intelligence, Volume 107, 2022, 104538.

[30] A. Pandey et al., ″Adaptive Mini-Batch Gradient-Ascent-Based Localization for Indoor IoT Networks Under Rayleigh Fading Conditions,″ in IEEE Internet of Things Jour., vol. 8, no. 13, pp. 10665-10677, 1 July, 2021.

[31] C. -H. Hsieh, J. Chen and B. Nien, ″Deep Learning-Based Indoor Localization Using Received Signal Strength and Channel State Information,″ in IEEE Access, vol. 7, pp. 33256-33267, 2019.

[32] M.T. Hoang et al., ″Recurrent Neural Networks for Accurate RSSI Indoor Localization,″ in IEEE Internet of Things Jour., vol. 6, no. 6, pp. 10639-10651, Dec. 2019.

[33] M. Youssef, A. Agrawala and A. Shankar, ″WLAN location determination via clustering and probability distributions,″ Proc. of 1st IEEE Int'l Conf. on Pervasive Computing and Commun., 2003, pp. 143-150.

[34] P. Jiang et al.,″Indoor Mobile Localization Based on Wi-Fi Fingerprint's Important Access Point.″ Int'l Jour. of Dist. Sensor Netw., (April 2015).

[35] T. Lin et al., ″A Group-Discrimination-Based Access Point Selection for WLAN Fingerprinting Localization,″ in IEEE Trans. on Vehicular Tech., vol. 63, no. 8, pp. 3967-3976, Oct. 2014.

[36] A Kushki, K. Plataniotis, and A Venetsanopoulos, ″Kernel-based positioning in wireless local area networks,″ IEEE Trans. on Mobile Comp., vol. 6,no. 6, pp. 689-705,2007.

[37] C. Laoudias, C. Panayiotou, and P. Kemppi, ″On the RBF-based positioning using WLAN signal strength fingerprints,″ in Proc. of 7th Workshop on Pos., Navigation and Comm. (WPNC), 2010, pp. 93-98.

[38] Y. Chen et al., ″Power-efficient access-point selection for indoor location estimation,″ IEEE Trans. on Knowledge and Data Engg, vol. 18, no. 7, pp. 877-888, 2006.

[39] S. Eisa et al., ″Removing useless APs and fingerprints from WiFi indoor positioning radio maps,″ Int'l Conf. on Indoor Pos. and Indoor Navigation, 2013, pp. 1-7.

[40] S.-P. Kuo et al., ″Cluster-Enhanced Techniques for Pattern-Matching Localization Systems,″ Proceedings of IEEE Int'l Conf. Mobile Ad-Hoc and Sensor Syst., 2007.

[41] C. Feng et al., ″Received-Signal-Strength-Based Indoor Positioning Using Compressive Sensing,″ in IEEE Trans. on Mobile Comp, vol. 11, no. 12, Dec. 2012, pp. 1983-1993.

[42] Z. Tian et al., ″Fingerprint indoor positioning algorithm based on affinity propagation clustering,″ EURASIP Jour. on Wireless Comm. and Networking, 2013.

[43] S. Subedi et al., ″Improving Indoor Fingerprinting Positioning With Affinity Propagation Clustering and Weighted Centroid Fingerprint,″ in IEEE Access, vol. 7, pp. 31738-31750, 2019.

[44] X. Hu et al. ″Improving Wi-Fi Indoor Positioning via AP Sets Similarity and Semi-Supervised Affinity Propagation Clustering.″ Int'l Jour. of Dist. Sensor Netw., (January 2015).

[45] F. Yang et al., ″Dynamic active area clustering with inertial information for fingerprinting based indoor localization systems,″ Proc. of 14th IFIP Networking Conference, May 2015, Toulouse, France.

[46] P. Tang, ″Wi-Fi Fingerprint Localization Using RSSI-Probability Radio Map and AP Weight Clustering,″ Jour. of Advances in Computer Netw., 2016, pp. 121-124.

[47] W. Liu, F. Xiao and D. Zhongliang, ″Coordinate-Based Clustering Method for Indoor Fingerprinting Localization in Dense Cluttered Environments″ Sensors 16, no. 12: 2055.

[48] W. Zhang et al., ″Radius based domain clustering for WiFi indoor positioning,″ Sensor Review, 37(1), 54-60, 2017.

[49] X. Yang et al., ″AP Optimization for Wi-Fi Indoor Positioning-Based on RSS Feature Fuzzy Mapping and Clustering,″ in IEEE Access, vol. 8, pp. 153599-153609, 2020.

[50] J. Bi et al., Improved Indoor Fingerprinting Localization Method Using Clustering Algorithm and Dynamic Compensation. ISPRS Int. J. Geo-Inf. 2021, 10, 613.

[51] Y. Le et al, ″ Received signal strength based indoor positioning algorithm using advanced clustering and kernel ridge regression, ″ Front Inform Technol Electron Eng 22, pp. 827–838, 2021.

[52] W. Xue et al., ″APs′ Virtual Positions-Based Reference Point Clustering and Physical Distance-Based Weighting for Indoor Wi-Fi Positioning,″ IEEE Internet of Things Jour., vol. 5, no. 4, Aug. 2018, pp. 3031-3042.

[53] A. Li et al., ″A Cluster-Principal-Component-Analysis-Based Indoor Positioning Algorithm,″ in IEEE Internet of Things Jour., vol. 8, no. 1, pp. 187-196, 1 Jan.1, 2021.

[54] B. Pinto et al., ″Robust RSSI-Based Indoor Positioning System Using K-Means Clustering and Bayesian Estimation,″ in IEEE Sensors Jour., vol. 21, no. 21, pp. 24462-24470, 1 Nov.1, 2021.

[55] G. Zhou et al., ″Impact of Radio Irregularity on Wireless Sensor Networks,″ in Proc. of 2nd Int'l Conf. on Mobile syst, appl. and services, 2004, pp. 125-138.

[56] N. Saccomanno, A. Brunello and A. Montanari, ″What You Sense is Not Where You are: on the Relationships Between Fingerprints and Spatial Knowledge in Indoor Positioning,″ in IEEE Sensors Jour., vol. 22, no. 6, pp. 4951-4961, 15 March, 2022.

Pampa Sadhukhan is presently an Assistant Professor in School of Mobile Computing & Communication, Jadavpur University, India. She was recipient of Erasmus Mundus postdoctoral fellowship and pursued her postdoctoral research in the Artificial Intelligence, Virtual Communication and Network (AVCN) Research Centre at the University of the West of Scotland (UWS), UK from November, 2015 to August, 2016. She received the PhD degree in 2012 and completed ME in Computer Science in 2005 from Jadavpur University, India. Her research interests include wireless localization & tracking, machine learning, IoT and mobile computing. She is also a member of IEEE.

Keshav Dahal is a Professor of Intelligent Systems and the leader of the Artificial Intelligence, Virtual Communication and Network (AVCN) Research Centre at the University of the West of Scotland (UWS), UK. Before joining UWS he was with Bradford and Strathclyde Universities in UK. He obtained his Ph.D. and Master from Strathclyde. His research interests lie in the areas of applied AI to intelligent systems, trust and security modeling in distributed systems, and scheduling/optimization problems. He has published extensively with award winning papers, and has sat on organizing/program committees of over 60 international conferences including as the General Chair and Programme Chair. He is a senior member of the IEEE.

Pradip K. Das served as a Professor & Head in the Department of Computer Science & Engineering, Jadavpur University, Kolkata. He also served as the President of the Institute for Open Technology Applications under the Department of Information Technology & Electronics, Govt. of West Bengal. He was the founder director of the School of Mobile Computing & Communication, a centre of excellence at Jadavpur University created by the University Grants Commission, Govt. of India. Dr. Das received his B.E., M.E. and Ph.D. degrees all from Jadavpur University. He was a senior visiting fellow in the Department of Computer Science, The Queen's University of Belfast, Northern Ireland, UK. He was a recipient of the prestigious ″Technology for Teaching" award in the Asia Pacific & Japan region in the year 2006 instituted by the Hewlett Packard Philanthropy trust. His research interests are in the areas of distributed computing, mobile computing, ad hoc and sensor networks and application of technology for differently abled persons. Dr. Das is a Life Senior Member of IEEE.