

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

5,200

Open access books available

129,000

International authors and editors

155M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Machine Learning Algorithm for Wireless Indoor Localization

Osamah Ali Abdullah and Ikhlas Abdel-Qader

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.74754>

Abstract

Smartphones equipped with Wi-Fi technology are widely used nowadays. Due to the need for inexpensive indoor positioning systems (IPs), many researchers have focused on Wi-Fi-based IPs, which use wireless local area network received signal strength (RSS) data that are collected at distinct locations in indoor environments called reference points. In this study, a new framework based on symmetric Bregman divergence, which incorporates k-nearest neighbor (kNN) classification in signal space, was proposed. The coordinates of the target were determined as a weighted combination of the nearest fingerprints using Jensen-Bregman divergences, which unify the squared Euclidean and Mahalanobis distances with information-theoretic Jensen-Shannon divergence measures. To validate our work, the performance of the proposed algorithm was compared with the probabilistic neural network and multivariate Kullback-Leibler divergence. The distance error for the developed algorithm was less than 1 m.

Keywords: fingerprinting, indoor localization, Wi-Fi, Jensen-Bregman divergence, probabilistic neural network, multivariate Gaussian distribution

1. Introduction

Automatically, identifying the location of a user has recently become a hot topic in research. The study in [1] estimated that the global indoor localization market is expected to grow from its value of \$935.05 million in 2014 to approximately \$4.42 billion in 2019, corresponding to an estimated compound annual growth rate (CAGR) of 36.5%. The estimation of mobile locations has an important role in many computing applications. The global positioning system (GPS) is one of the most common location-based systems, but it cannot be used inside buildings as a direct line of sight (LOS) is required between the GPS receiver and the satellite to identify the

user's location. Therefore, a large number of technologies, such as Bluetooth, radiofrequency identification (RFID), wireless local area network (WLAN or Wi-Fi), magnetic field variations, ultrasound, Zigbee, and light-emitting diode (LED) light bulbs, have been developed to create high-accuracy indoor positioning systems (IPS), with Wi-Fi being the most commonly used technology. Most smartphones can obtain received signal strength (RSS) from the access points (APs) of WLANs because of the low cost and existing WLAN infrastructure [2, 3]

The IPS algorithm that uses RSS-based indoor localization can be classified into two main types: log-distance propagation model (PM) algorithms based on the signal and fingerprinting indoor localization based on the data collected. IPS based on signal propagation is divided into lateration and angulation. The main idea of lateration estimation is to calculate the distance between the smartphone and AP using geometry and signal measurement information, such as the time of arrival (TOA), time difference of arrival (TDOA), and angle of arrival (AOA), of the incoming signals from APs. In general, propagation signals suffer from non-line-of-sight (NLOS) multipath signals due to the presence of walls and furniture and the movement of people. In addition, the signal accuracy decreases if one or more AP coordinates are not accurately calculated. All of these drawbacks have made it difficult to estimate an object's position using signal propagation [4]. Thus, fingerprinting-based localization systems have been proposed as an alternative technology [5] as they do not require infrastructure. Instead, they use the existing WLAN in the building and the smartphone, which relies on the spectrum of the RSS from the APs to the location to estimate the user's location coordinates.

The fingerprint-based technique is divided into offline and online phases. In the offline phase, the entire area of interest is divided into a rectangular set of grid points, and at each point, a site survey is taken by recording the RSS from APs, which is then stored in a database called the radio map [6–10]. In the online phase, the smartphone collects the RSS from the APs and sends it to the server to compare the predefined fingerprint of the offline phase with the RSS in the online phase in order to estimate the location on the grid map, as shown in **Figure 1**.

The k-nearest neighbor (kNN) algorithm is one of the simplest ways to estimate location; it depends upon the Euclidean distance to measure the similarity/dissimilarity between the offline and online phases. Even though this algorithm is easy to implement, it has low accuracy. Other methods such as statistical learning and Bayesian modeling have also been used to estimate the location of an object. Accuracy is one of the most important requirements of IPS.

Mean distance error is typically used as the performance metric and is calculated as the average Euclidean distance between the actual location and the estimated location.

Recently, an important issue was raised about the variation of signal propagation, namely, the question of how signals are able to propagate over time in the same place in the presence of multiple factors, such as physical obstructions, radiofrequency (RF) equipment, and the presence of human bodies. These factors can lead to attenuation and multipath issues, thereby causing gradual changes in the signal that can reduce the accuracy of the localization system [11]. The values stored in fingerprint maps represent the mean value of the received signal strength indicator (RSSI). Some approaches presume that the RSSI distribution is Gaussian

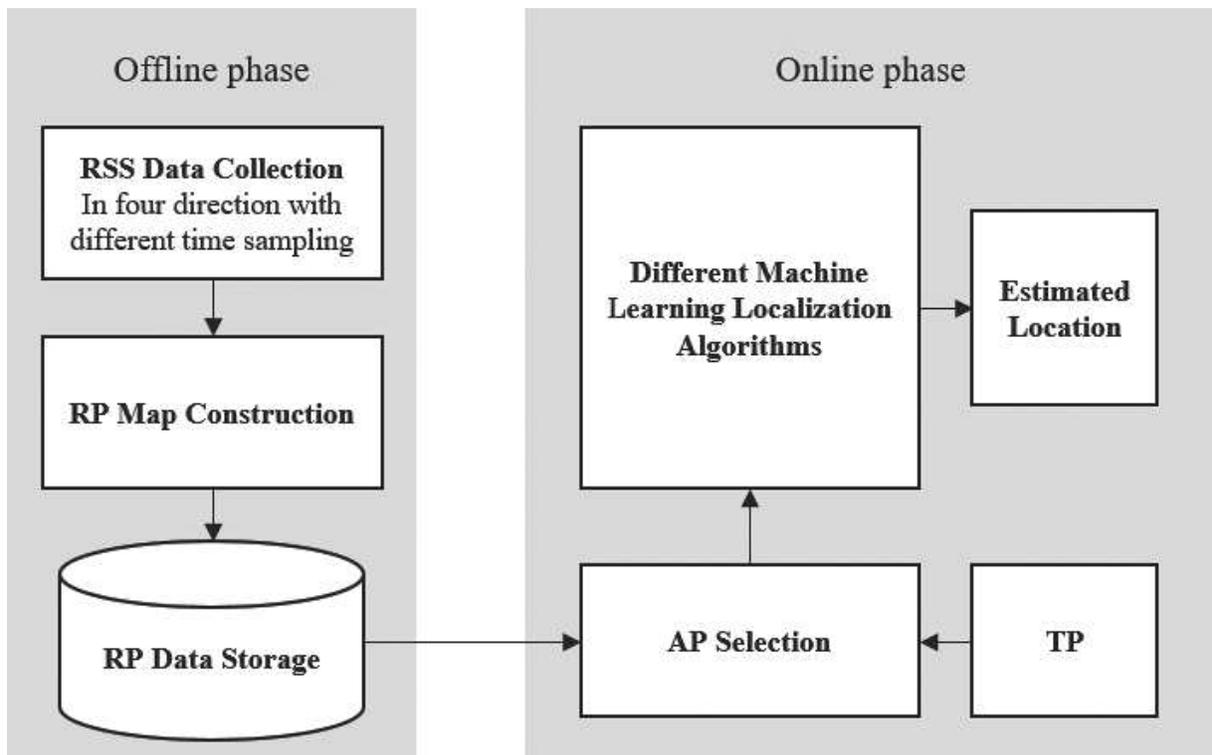


Figure 1. The offline and online stages of location Wi-Fi-based fingerprinting architecture.

[12], whereas others assume a non-Gaussian distribution, such as those described in [13]. Using Wi-Fi localization systems to estimate the location of an object has many advantages compared with other technologies, such as availability and low cost. However, because the RSSI signal uses both offline and online phases, hardware variance can significantly degrade the positional accuracy of these systems. Some studies have investigated this variance; for example, it was reported in [11] that when using different smartphones to collect RSSI data at the same time and same location, some phones consistently had higher RSSI values than others. The orientation of the user can also contribute to the variance of the RSSI signal because the human body can be a significant attenuator.

This hardware variance problem in Wi-Fi localization has also been noticed in Cisco location systems [11]; some signals were found to be omitted when a different device was used in the online phase compared with the offline phase.

This chapter presents the following:

- We propose the use of Jensen-Bregman divergence (JBD) as a WLAN-based method and a Kullback-Leibler multivariate Gaussian (KL_{MVG}) model. The matching stage was performed using probability kernels as a regression scheme.
- We propose a procedure with high characterization distribution. The RSS values were taken from four different orientations (45, 135, 225, and 315°) to prevent body-blocking effects, with a scan performed for 100 s in each direction to reduce the effects of signal variation.

- JBD and KL_{MVG} outperformed the probabilistic neural network (PNN) and kNN with respect to the accuracy and the average error distance, indicating that the proposed combination scheme is more effective in the sensitive environments of WLAN-based positioning systems.

2. Related work

Global navigation satellite systems (GNSS) such as GLONASS (Russia's version of GPS), GALILEO, and GPS work well in outdoor environments, but their accuracy can significantly decrease in indoor environments due to many factors, such as penetration loss, refraction, multipath propagation, and absorption. Therefore, it is important to develop a system that can work in indoor environments with high accuracy. To this end, many techniques have been proposed for IPS in the last decade. In model-based techniques, the location is estimated based on a geometrical model, such as the log-distance path loss (LDPL) model, in which a semi-statistical function is built on the relationship between the RF propagation function and the RSS value. Several approaches have been proposed that are trade-offs between accuracy and cost, such as TOA, TDOA, AOA, and multidimensional scaling (MDS). MDS is a set of statistical techniques that are used to visualize the information in order to find similarities/dissimilarities in the data. The matrix in MDS begins with item-item dissimilarities, and AP-AP distances are determined by a radio attenuation model [9]. The fingerprinting-based technique depends on matching algorithms (e.g., kNN) that have been used in RADAR [14], which is one of the first Wi-Fi signal strength-based IPS and is considered the basis of WLAN fingerprinting IPS. Many developed kNN algorithms have been proposed for determining the similarity/dissimilarity in metrics, which is usually done using the Manhattan or Euclidean distance, such as in [11–18]. Ref. [19] proposed a new version of kNN that is more efficient than the probabilistic methods, neural networks, and traditional kNN, as it relies upon the decision tree of the training phases and takes into account the average of reference point (RP) measurements instead of needing the entire dataset to estimate the object's location. Ref. [20] performed a modified deterministic kNN technique with Mahalanobis, Manhattan, and Euclidean distances and found the Manhattan distance to be the most accurate. Recently, the use of probabilistic distribution measurements in many IPS applications has increased. The authors in [21] pioneered the use of the probabilistic distribution measurement in IPS and proposed a probabilistic framework by using the Bayesian network to estimate the location. In [22] the authors used a modified probability neural network (MPNN) to estimate the coordinates of the object and found that it outperformed the triangulation method. In [23], a kernel method was proposed to estimate the object's location using a histogram of the RSSI at the unknown location. In [24], the probability density function (PDF) was estimated using the Kullback-Leibler divergence (KLD) framework for composite hypothesis testing between the fingerprinting database and the test point, whereas in [25], the authors assumed that the RSSI distribution was multivariate Gaussian and used the KLD to estimate the impacts of the RPs on the test point in order to estimate the probability of the closest one and to identify the coordinates of the test point.

In [26], the RSS-based Bluetooth low-energy localization technique was used to establish the fingerprint, after which the KLD was used in probabilistic kernel regression to estimate the object's location. The results showed this method to be accurate to approximately 1 m in an office environment. In general, the KLD kernel regression performs better in a multimodal distribution. In [27], the KLD was used to estimate the probabilistic kernel of both Gaussian and non-Gaussian distributions in order to compare them and to determine their limitations.

3. Overall structure of indoor positioning system

We begin with a typical WLAN scenario in which a person carries a smartphone device with WLAN access and takes RSS measurements from different APs within the College of Engineering and Applied Sciences (CEAS) at Western Michigan University (WMU). It is commonly assumed that the RSSI from multiple APs is distributed as a multimodal signal, as noted in [16]. However, in our study, the recorded signal-to-noise ratio for a single device varied significantly at any one location, with the values differing by as much as 10 dBm. Specifically, the signal-to-noise values were recorded for 35 min during rush hour for a single AP and in the same location.

There are many parameters that can affect the shape of the signal, such as reflection, diffraction, and pedestrian traffic. In this study, we sought to find a scenario that would lead to a better distribution of the Wi-Fi signal. During the offline phase, a realistic scenario was created that took into account the variation of the signal. However, because the effects of the body of the person holding the phone as well as pedestrian traffic can change the variation of the signal, a recording of the RSS was taken in four directions (45, 135, 225, and 315°) to reduce these variations. At each RP, a raw set of RSS data were collected as a time sample from the APs in the area of interest, denoted as $\{q_{i,j}^{(\circ)}(\tau), \tau = 1, \dots, t, t = 100\}$, where t represents the number of time samples and (\circ) is the orientation direction. Next, the average and covariance matrix of the RSS were obtained from the four different directions and ten scans used to create the fingerprinting database, known as the Radio Map, represented by $Q^{(\circ)}$ [28]:

$$Q^{(\circ)} = \begin{pmatrix} q_{1,1}^{(\circ)} & q_{1,2}^{(\circ)} & \cdots & q_{1,N}^{(\circ)} \\ q_{2,1}^{(\circ)} & q_{2,2}^{(\circ)} & \cdots & q_{2,N}^{(\circ)} \\ \vdots & \vdots & \ddots & \vdots \\ q_{L,1}^{(\circ)} & q_{L,2}^{(\circ)} & \cdots & q_{L,N}^{(\circ)} \end{pmatrix} \quad (1)$$

where $q_{i,j}^{(\circ)} = \frac{1}{t} \sum_{\tau=1}^t q_{i,j}^{(\circ)}(\tau)$ and $t = 10$, which were randomly chosen from the 100 time samples. This allowed us to obtain the average of the RSS samples over time for different APs, $i = 1, 2, \dots, L, j = 1, 2, \dots, N$, where N represents the number of RPs and L is the number of APs. The variance vector of each RP can be defined as

$$\Delta_j^{(\circ)} = [\Delta_{1,j}^{(\circ)}, \Delta_{2,j}^{(\circ)}, \Delta_{3,j}^{(\circ)}, \dots, \Delta_{L,j}^{(\circ)}] \tag{2}$$

where

$$\Delta_{i,j}^{(\circ)} = \frac{1}{t-1} \sum_{\tau=1}^t (q_{i,j}^{(\circ)}(\tau) - q_{i,j}^{(\circ)})^2 \tag{3}$$

where $\Delta_{i,j}^{(\circ)}$ is the variance for AP i at RP j with orientation (\circ) ; thus, the database table of the Radio Map is $(x_j, y_j, q_j^{(\circ)}, \Delta_j^{(\circ)})$ with $q_j^{(\circ)}$ defined as

$$q_j^{(\circ)} = [q_{1,j}^{(\circ)}, q_{2,j}^{(\circ)}, q_{3,j}^{(\circ)}, \dots, q_{L,j}^{(\circ)}] \tag{4}$$

During the online phase, the RSS measurement is denoted as

$$p_r = [p_{1,r}, p_{2,r}, \dots, p_{L,r}] \tag{5}$$

4. The Kullback-Leibler multivariate Gaussian (KLMvG) model

Another approach, specifically the KLMvG model, has recently been used in fingerprinting-based methods to estimate the position of the objects. This model exploits the interdependencies between the RPs, such as the signal model and the geometry that can be quantified to find the correlations among the RPs. Milioris [29] proposed a KLMvG model to measure the similarity between the RSS measurements of test points and the RPs, defined as

$$KL_{MVG}(p||q_j^{(\circ)}) = \frac{1}{2} \left((\mu_{q,j}^S - \mu_p^S)^T (\Sigma_{j,q}^S)^{-1} (\mu_{q,j}^S - \mu_p^S) + tr \left(\Sigma_R^S (\Sigma_{j,q}^S)^{-1} - I \right) - \ln \left| \Sigma_p^S (\Sigma_{j,q}^S)^{-1} \right| \right) \tag{6}$$

where S represents the matrix of RSS values from the different APs at specific locations and j represents the cell of the fingerprint location where

$$S_j^{(\circ)} = \{ \mu_j^{(\circ)}, \Sigma_j^{(\circ)} \} \tag{7}$$

$\mu_j^{(\circ)}$ is the mean of J th column of the RSS measurement and $\Sigma_j^{(\circ)}$ represents the covariance matrix, where $|\Sigma|$ is the determinant of Σ . Now, using a KLMvG model, we can formulate a probability kernel-based approach. The kernel regression scheme allows us to estimate the PDF of the training datasets and the true positives (TPs) from the online phase that are used to estimate the location of the object. The KL_{MVG} model is used to measure the distance between the likelihood of the input sample and the RPs in order to determine which class it belongs to. The RSS distribution can be defined as

$$D(p, q_\ell) = \exp \left(- \frac{KL_{MVG}(p||q_j^{(\circ)})}{2\sigma^2} \right) \tag{8}$$

where σ is the kernel smoothing factor. The probability will be equal to 1 if $p = q$, and the output will decrease when the difference between p and q becomes larger.

Algorithm 1. The Kullback-Leibler multivariate Gaussian positioning method.

1. During the offline phase, RSS measurements are taken at different known locations, and 10 scans with 10 second time delays are used to generate the Radio Map.
 2. During the online phase, RSS measurements are taken at unknown locations of the smartphone.
 3. During the online phase, the following steps are performed:
 - A database for each RP is set using RSS measurements from different locations.
 - The RSS measurements from the APs of smartphones from unknown locations are set in the same way as the database of the offline phase with respect to the similar media access control (MAC) address.
 - The minimum KLMvG is estimated using Eq. 8.
 - The previous step is repeated for different APs until the minimum distance is obtained.
 4. The maximum outputs to the output layer are transferred.
-

5. Bregman divergence algorithm formulation

In recent times, approaches that measure the distortion in classes have become more common, instead of depending on a single distance. Indeed, the analysis of distortion is being used in many applications of machine learning, computational geometry, and IPS. Using Bregman divergence to measure the similarity/dissimilarity has recently become an attractive method because it encapsulates both information-theoretic relative entropy and the geometric Euclidean distance, which is a meta-algorithm [30]. The Bregman distance D_φ between two sets of convex space data, $p = (p_1, \dots, p_d)$ and $q = (q_1, \dots, q_d)$, that is associated with φ (defined as a strictly convex and differentiable function) can be defined as

$$D_\varphi(p, q) = \varphi(p) - \varphi(q) - \langle \nabla\varphi(p), p - q \rangle \quad (9)$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product:

$$\langle p, q \rangle = \sum_{i=1}^d p^{(i)} q^{(i)} = p^T q \quad (10)$$

and $\nabla\varphi(p)$ denotes the gradient decent operator:

$$\nabla\varphi(p) = \left[\frac{\partial\varphi}{\partial p_1} \dots \frac{\partial\varphi}{\partial p_d} \right]^T \quad (11)$$

The Bregman divergence unifies the statistical KLD with the squared Euclidean distance by defining the distortion measurement in classes:

- The Euclidean distance is obtained from the Bregman divergence by considering the convex function as $\varphi(p) = \sum_{i=1}^d p_i^2 = \langle p, p \rangle$, which is the parabolic potential function in **Figure 2**.

- The KLD is also a Bregman divergence if the convex function used is $\varphi(p) = \sum_{i=1}^d p_i \log p_i$, which is defined as negative Shannon entropy. The KLD is defined for two discrete distributions as

$$KL(p||q) = \sum_s p(S = s) \log \left(\frac{p(S = s)}{q(S = s)} \right) \tag{12}$$

In information theory, the Shannon differential entropy measures the amount of uncertainty of a random variable:

$$H(p) = - \sum_s p(S = s) \log p(S = s) \tag{13}$$

The KLD is equal to the cross entropy of two discrete distributions minus the Shannon differential entropy [31]:

$$KL(p||q) = \sum_s H^x(p(S = s)||q(S = s)) - H(p(S = s)) \tag{14}$$

where H^x is the cross-entropy:

$$H^x(p(S = s)||q(S = s)) = - \sum_s p(S = s) \log q(S = s) \tag{15}$$

and S is the set of vectors of the RSS. In general, the Bregman divergence is not symmetrical, but it can symmetrize as follows:

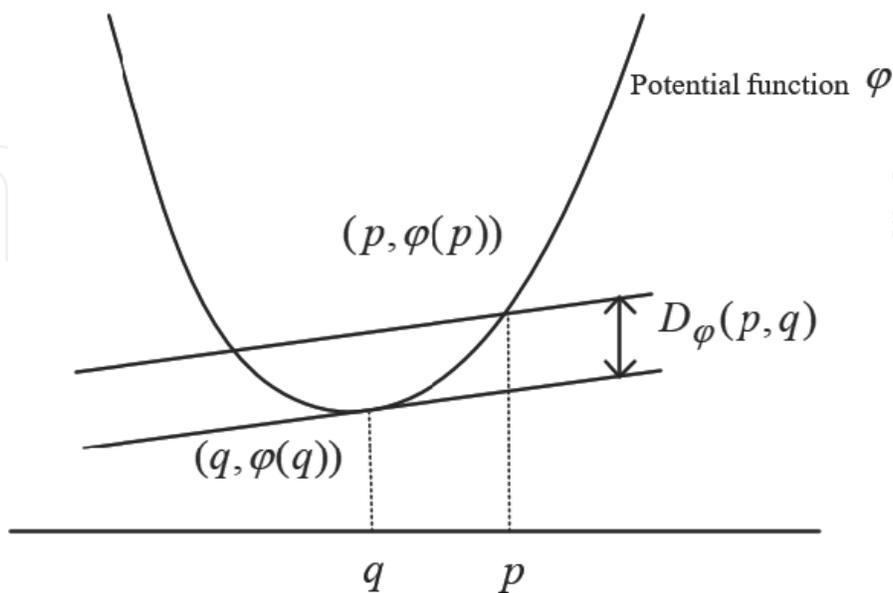


Figure 2. The Bregman divergence represents the vertical distance between the potential function and the hyperplane at q.

$$SD_{\varphi}(p, q) = \frac{D_{\varphi}(p, q) + D_{\varphi}(q, p)}{2} \quad (16)$$

$$= \frac{1}{2} \langle p - q, \nabla\varphi(p) - \nabla\varphi(q) \rangle \quad (17)$$

In the same manner, Jeffreys' divergence symmetrizes the oriented KLD as follows:

$$J(p, q) = KL(p||q) + KL(q||p) \quad (18)$$

$$= H(p||q) + H(q||p) - (H(p) + H(q)) \quad (19)$$

$$= \sum_s \left((p(S=s) - q(S=s)) \log \left(\frac{p(S=s)}{q(S=s)} \right) \right) \quad (20)$$

Such information-theoretic divergence has two major drawbacks: first, the output can be undefined if $q = 0$ and $p \neq 0$, and second, the J-divergence is not bound by terms of metric distance. To avoid these drawbacks and avoid the $\log(0)$ or to divide by 0, the authors in [32] proposed a new divergence called K-divergence:

$$K(p||q) = KL\left(p, \frac{p+q}{2}\right) \quad (21)$$

By introducing the K-divergence, [30] produced the Jensen-Shannon divergence (JSD) as follows:

$$JSD(p||q) = \frac{1}{2} \left(KL\left(p, \frac{p+q}{2}\right) + KL\left(q, \frac{p+q}{2}\right) \right) \quad (22)$$

$$= \frac{1}{2} \left(H\left(p||\frac{p+q}{2}\right) - H(p) + H\left(q||\frac{p+q}{2}\right) - H(q) \right) \quad (23)$$

$$= \frac{1}{2} \left(\sum_{i=1}^L \left(p_i \log \frac{p_i}{\frac{1}{2}q_i + \frac{1}{2}p_i} + q_i \log \frac{q_i}{\frac{1}{2}q_i + \frac{1}{2}p_i} \right) \right) \quad (24)$$

The JSD can be defined, bound by an L1-metric, and finite. In the same vein, the Bregman divergence can be symmetrized as

$$SD_{\varphi}(p, q) = \frac{1}{2} \left(D_{\varphi}\left(p, \frac{q+p}{2}\right) + D_{\varphi}\left(q, \frac{q+p}{2}\right) \right) \quad (25)$$

$$= \frac{\varphi(p) + \varphi(q)}{2} - \varphi\left(\frac{p+q}{2}\right) \quad (26)$$

for d-dimensional multivariate data:

$$SD(p, q) = \sum_{i=1}^L \frac{\varphi(p_i) + \varphi(q_i)}{2} - \varphi\left(\frac{p_i + q_i}{2}\right) \quad (27)$$

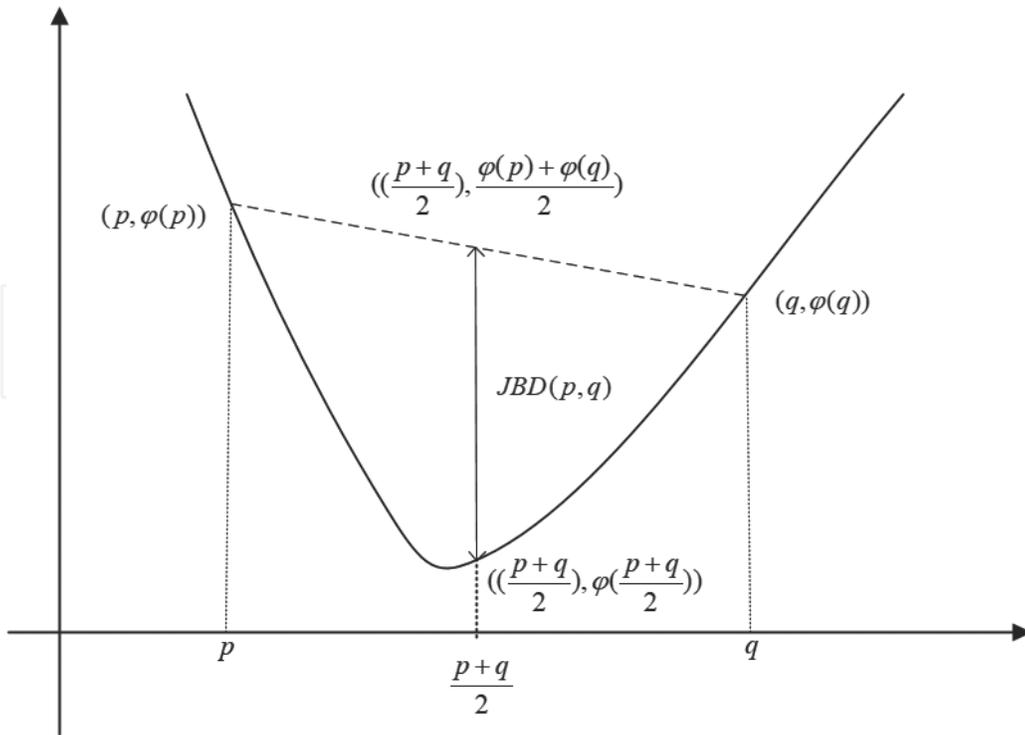


Figure 3. Interpreting the Jensen-Bregman divergence.

where q represents the fingerprint dataset and p , which is the dataset of the test points, represents the APs that the mobile device received. Because φ is a strictly convex function and $SD(p, q)$ equals zero if and only if $p = q$, this family of distortions is termed JSD. The geometric interpretation is represented in **Figure 3**, where divergence represents the vertical distance between $((\frac{p+q}{2}), \varphi(\frac{p+q}{2}))$ and the midpoint of the segment $[(p, \varphi(p)), (q, \varphi(q))]$.

In general, for a positive definite matrix, the Jensen-Bregman divergence contains the generalized quadratic distance, which is known as the Mahalanobis distance:

$$\begin{aligned}
 SD(p, q) &= \frac{\varphi(p) + \varphi(q)}{2} - \varphi\left(\frac{p+q}{2}\right) \\
 &= \frac{2\langle Qp, p \rangle + 2\langle Qq, q \rangle - 2\langle Q(p+q), p+q \rangle}{4} \\
 &= \frac{1}{4} (\langle Qp, p \rangle + \langle Qq, q \rangle - 2\langle Qp, q \rangle) \\
 &= \frac{1}{4} \langle Q(p-q), p-q \rangle \\
 &= \frac{1}{4} \|p - q\|_Q^2
 \end{aligned}
 \tag{28}$$

To improve accuracy, we present Algorithm 2:

Algorithm 2. The Kullback-Leibler multivariate Gaussian positioning method

1. During the offline phase, RSS measurements are taken at different known locations, and 10 scans with 10 second time delays are used to generate the Radio Map.
 2. During the online phase, RSS measurements are taken from unknown locations of the smartphone.
 3. During the online phase, the following steps are performed:
 - A database for each RP is set using RSS measurements from different locations.
 - The RSS measurements from the APs of smartphones from unknown locations are set in the same way as the database of the offline phase with respect to the similar media access control (MAC) address.
 - The minimum symmetric Bregman divergence is estimated using Eq. 27.
 - The previous step is repeated for different APs until the minimum distance is obtained.
 4. The maximum outputs are transferred to the output layer.
-

6. Performance analysis

The proposed algorithm evaluations will be presented in the subsequent subsections; the algorithms were implemented on the first floor of the CEAS at WMU. To collect the data sample, a Samsung S5 smartphone with operating system 4.4.2 was used. The proposed algorithms were implemented on an HP Pavilion using Java software with an Eclipse framework. Cisco Linksys E2500 Advanced Simultaneous Dual-Band Wireless-N Routers were used in the area of interest. Most of this work discounted the variation of the RSS from the APs.

To evaluate the performance of the different fingerprinting techniques, the localization error was computed as the Euclidean distance between the actual reported coordinates of the test points and the coordinates of the mobile user during the online phase. The number of RSS of the APs and the number of nearest neighbors were noted, as they can affect the accuracy of the algorithms. The number of APs can play an important role in the accuracy of the distance error, which can distinguish near RPs from those further away.

To evaluate the performance of the different fingerprinting techniques, the localization error was computed as the Euclidean distance between the actual reported coordinates of the test points and the coordinates of the mobile user during the online phase. The number of RSS of the APs and the number of nearest neighbors were noted, as they can affect the accuracy of the algorithms. The number of APs can play an important role in the accuracy of the distance error, which can distinguish near RPs from those further away.

In order to measure the impact of the APs on the accuracy, we used a specific number of nearest neighbors with a variety of APs. However, that resulted in a longer RSS scanning interval, which slowed the process down. As a result, the online phase comprised five time samples, which took 1 s for Wi-Fi scanning on the device. To investigate the accuracy of our proposed algorithm, different algorithms were used, such as PNN and KNN, and compared with our proposed algorithm. Different numbers of nearest neighbors were used to estimate the location of the object and to evaluate the performance of our system framework.

Figure 4 shows the impact of different APs when five nearest neighbors were used. The lowest localization error was obtained when 22 APs were used: 0.98 m for kJBD, 1.12 m for kJSD, 1.16 m for KLMvG, 1.34 m for PNN, and 1.38 m for kNN. Greater accuracy was obtained when more nearest neighbors were used, as illustrated in **Figure 5**.

The lowest localization accuracy was also obtained when 22 APs were used: 0.92 m for kJBD, 1.01 m for kJSD, 1.02 m for KLMvG, 1.097 m for PNN, and 1.19 m for kNN. More improvements in system accuracy were noticed when 80 nearest neighbors were used: 0.865 m for kJBD, 0.96 m for kJSD, 0.99 m for KLMvG, 0.995 m for PNN, and 1.12 m for kNN, as shown in **Figure 6**.

Figure 7 illustrates the corresponding cumulative probability distributions of the localization error for the three methods. In particular, the median errors for kJBD were 0.89 m, 0.98 m for kJSD, and 1.02 m for KLMvG. Furthermore, an accuracy of 90% was achieved at 2.13 m for KLMvG and 1.93 m for kJSD, with the best accuracy obtained at 2.13 m for kJSD.

To validate our work, a comparison was made between the proposed algorithms with other algorithms from prior works, such as kNN [14], compressive sensing [28], and the kernel-based method [33], as illustrated in **Table 1**.

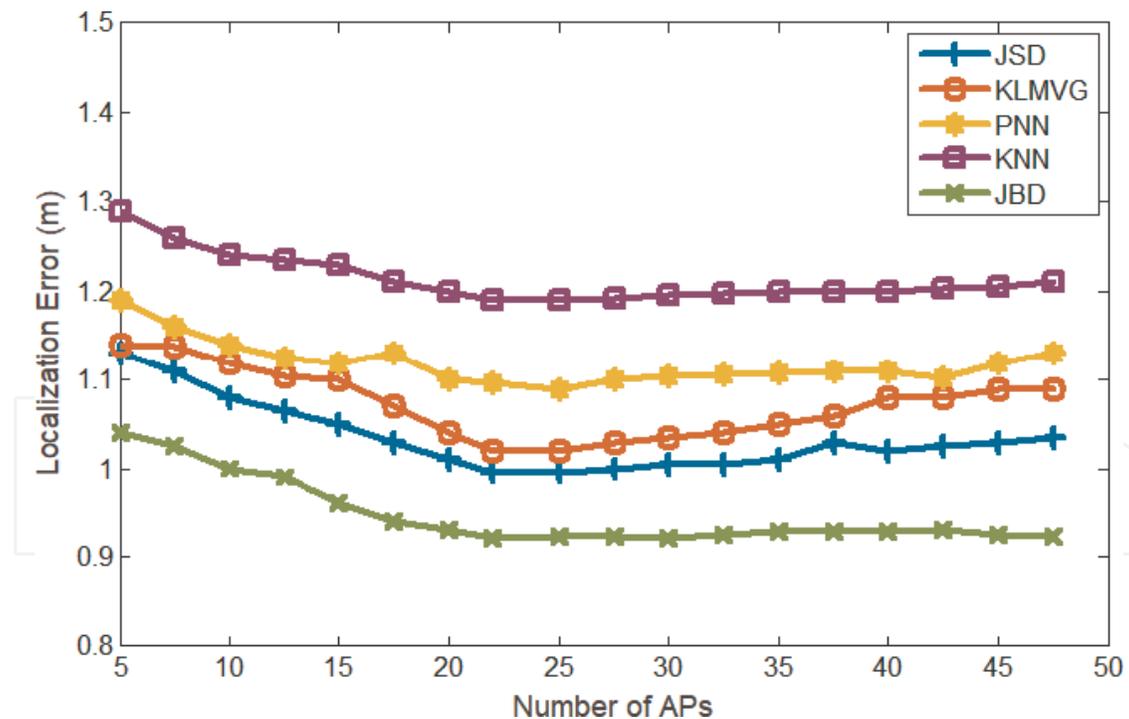


Figure 4. Error distance estimation with respect to APs with five nearest neighbors.

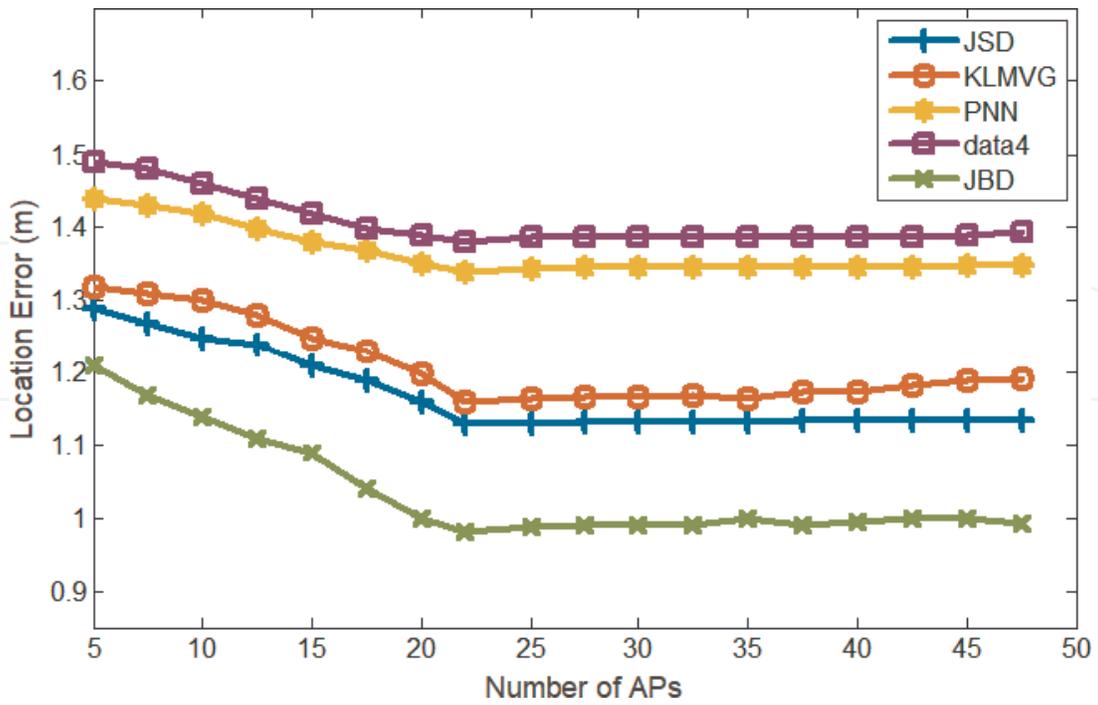


Figure 5. Error distance estimation with respect to APs with 20 nearest neighbors.

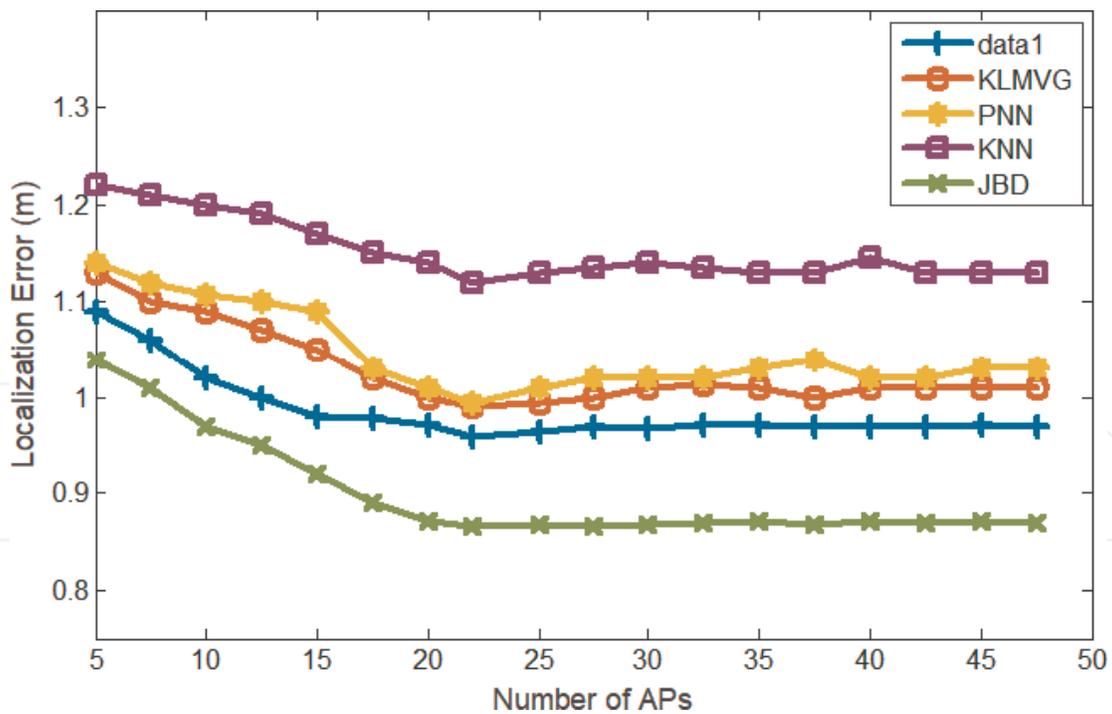


Figure 6. Error distance estimation with respect to APs with 80 nearest neighbors.

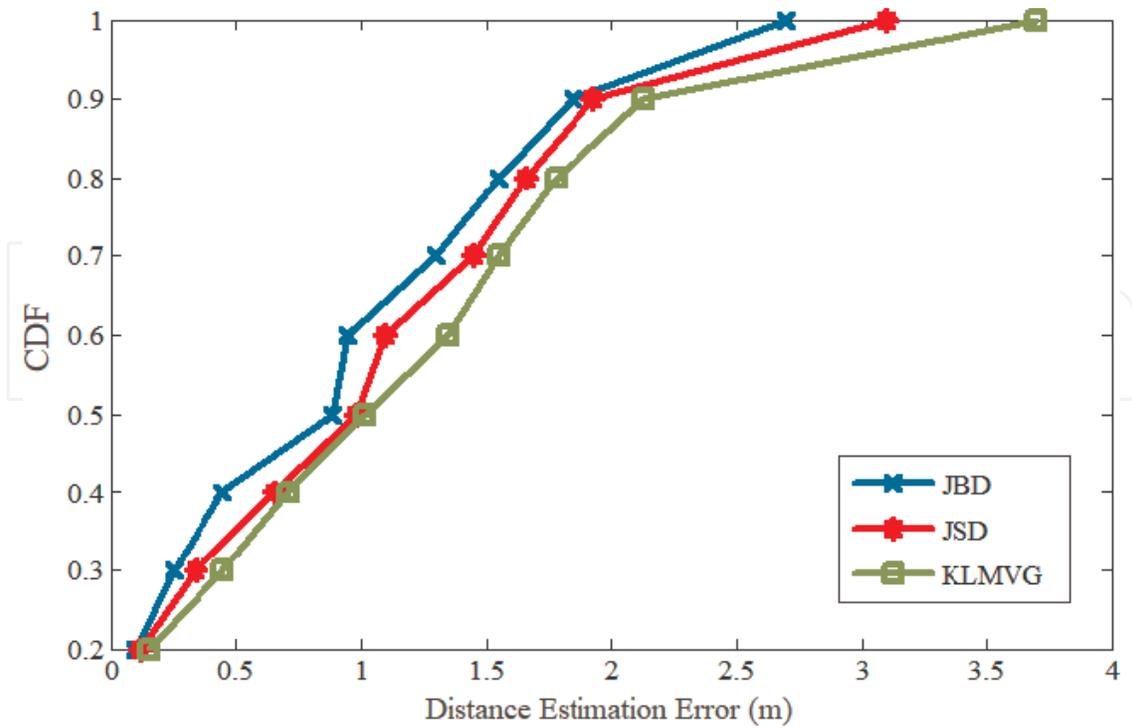


Figure 7. Experiment results: the CDF of localization error when using 80 nearest neighbors.

Technique	Median [m]	Accuracy 90% [m]
kNN [10]	1.8	3.7
Kernel-based [28]	1.6	3.6
CS-based [27]	1.5	2.7
KL _{MVG}	1.02	2.13
kJSD	0.98	1.93
kJBD	0.89	1.85

Table 1. Position error statistic.

7. Conclusion

IPS incorporates the power of GPS and indoor mapping and has many potential applications, for example, by emergency healthcare services, by people with impaired vision, and for navigating unfamiliar complex buildings where it is easy to get disoriented or lost (e.g., malls, airports, subways). A fingerprint map was created for a segment of the CEAS in order to utilize the relationship between different RSS readings. Different algorithms were used and compared using different approaches, including kNN and PNN, and their performances assessed for a number of APs. The results were quite adequate for the indoor environment with an average error of less than 1 m. The kJBD had the highest accuracy when there were 80

nearest neighbors with 22 APs. We are currently in the process of investigating position prediction error distributions and quantifying the localization variation of Wi-Fi signal distribution in space

Author details

Osamah Ali Abdullah* and Ikhlas Abdel-Qader

*Address all correspondence to: osamah.abdullah@wmich.edu

Electrical and Computer Engineering Department, Almamoon University College, Kalamazoo, Michigan USA

References

- [1] Markets & Markets. Indoor Localization Market by Positioning Systems, Map and Navigation, Location Based Analysis, Monitoring and Emergency Services-Worldwide Market Forecasts and Analysis (2014–2019). Technical Report; 2014
- [2] Torres-Sospedra J, Montoliu R, Trilles S, Belmonte O, Huerta J. Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems. *Expert Systems with Applications*. 2015;**42**(23):9263-9278
- [3] Jiang P, Zhang Y, Fu W, Liu H, Su X. Indoor mobile localization based on Wi-Fi fingerprint's important access point. *International Journal of Distributed Sensor Networks*. 2015;**11**(4):429104
- [4] Shchekotov M. Indoor localization methods based on Wi-Fi lateration and signal strength data collection. In: 2015 17th Conference of. Open Innovations Association (FRUCT); Yaroslavl; 2015
- [5] Swangmuang N, Prashant K. An effective location fingerprint model for wireless indoor localization. *Pervasive and Mobile Computing*. 2008;**4**(6):836-850
- [6] Abdullah O, Abdel-Qader I, Bazuin B. A probability neural network-Jensen-Shannon divergence for a fingerprint based localization. In: 2016 Annual Conference on Information Science and Systems (CISS). Princeton, NJ; 2016
- [7] Abdullah O, Abdel-Qader I. A PNN-Jensen-Bregman divergence symmetrization for a WLAN indoor positioning system. In: 2016 IEEE International Conference on Electro Information Technology (EIT). Grand Forks, ND; 2016
- [8] Abdullah O, Abdel-Qader I, Bazuin B, Fingerprint-based technique for indoor positioning system via machine learning and convex optimization. In: 2016 IEEE 7th Annual Ubiquitous

Computing, Electronics & Mobile Communication Conference (UEMCON). New York, NY; 2016

- [9] Abdullah O, Abdel-Qader I, Bazuin B. K-means-Jensen-Shannon divergence for a WLAN indoor positioning system. In: 2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). New York, NY; 2016
- [10] Abdullah O, Abdel-Qader I, Bazuin B. Convex optimization via Jensen-Bregman divergence for WLAN indoor positioning system. *International Journal of Handheld Computing Research (IJHCR)*. 2017;8(1):29-41. DOI: 10.4018/IJHCR.201701010
- [11] Sharma P, Chakraborty D, Banerjee N, Banerjee D, Agarwal SK, Mittal S. KARMA: Improving WiFi-based indoor localization with dynamic causality calibration. In: Eleventh Annual IEEE International Conference on Sensing, Communication, and Networking (SECON). Singapore; 2014
- [12] Hähnel B, Dirk B, Fox D. Gaussian processes for signal strength-based location estimation. In: *Proceedings of Robotics: Science and Systems*; 2006
- [13] Mirowski P, Steck H, Whiting P, Palaniappan R, MacDonald M, Ho TK. KL-divergence kernel regression for non-Gaussian fingerprint based localization. In: *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation*; 2011
- [14] Bahl P, Padmanabhan VN. RADAR: An in-building RF-based user location and tracking system. In: *Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE. Tel Aviv, INFOCOM*; 2000
- [15] Yang Z, Wu C, Liu Y. Locating in fingerprint space wireless indoor localization with little. In: *Mobicom*; 2012
- [16] Youssef M, Agrawala A. The Horus WLAN location determination system. In: *MobiSys*; 2005
- [17] Ni LM, Liu Y, Lau YC, Patil AP. LANDMARC: Indoor location sensing using active RFID. *Wireless Networks*. 2004;10(6):701-710
- [18] Sen S, Radunović B, Choudhury RR, Minka T. You are facing the Mona Lisa: Spot localization using PHY layer information. In: *MobiSys'12—Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*; 2012
- [19] Yim J. Introducing a decision tree-based indoor positioning technique. *Expert Systems with Applications: An International Journal*. 2008;34(2):1296-1302
- [20] Farshad A et al. A microscopic look at WiFi fingerprinting for indoor mobile phone localization in diverse environments. In: *Indoor Positioning and Indoor Navigation (IPIN)*; 2013
- [21] Castro P et al. A probabilistic location service for wireless network environments. In: *Ubiquitous Computing*; September 2001

- [22] Chen C, Chen Y, Yin L, Hwang R. A modified probability neural network indoor positioning technique. In: Information Security and Intelligence Control (ISIC), International Conference; 2012
- [23] Roos T et al. A probabilistic approach to WLAN user location estimation. *International Journal of Wireless Information Networks*. 2002;**3**:7
- [24] Wen Tsui A, Chuang Y, Chu H. Unsupervised learning for solving RSS hardware variance problem in WiFi localization. *Mobile Networks and Applications*. 2009;**14**(5):677-691
- [25] Milioris D, Kriara L, Papakonstantinou A, Tzagkarakis G. Empirical evaluation of signal-strength fingerprint positioning in wireless LANs. In: ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems; 2010
- [26] Mirowski P et al. KL-Divergence Kernel Regression for Non-Gaussian Fingerprint Based Localization. In: 2011 International Conference on Indoor Positioning and Indoor Navigation; 2011
- [27] Mirowski P et al. Probability kernel regression for WiFi localisation. *Journal of Location Based Services*. 2013;**6**(2):81-100
- [28] Feng C et al. Received-signal-strength-based indoor positioning using compressive sensing. *IEEE Transactions on Mobile Computing*. 2012;**11**(12)
- [29] Dimitris M et al. Low-dimensional signal-strength fingerprint-based positioning in wireless LANs. *Ad Hoc Networks*. 2014;**12**:100-114
- [30] Nielsen F, Nock R. Skew Jensen-Bregman Voronoi Diagrams. In: Gavrilova ML, Tan CJK, Mostafavi MA, editors. *Transactions on Computational Science XIV. Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer; 2011;6970
- [31] Nielsen, F. A family of statistical symmetric divergences based on Jensen's inequality. *Clinical Orthopaedics and Related Research*, vol. abs/1009.4004, (2010)
- [32] Lin J. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*. 1991;**37**(1):145-151
- [33] Kushki A, Plataniotis KN, Venetsanopoulos AN. Kernel-based positioning in wireless local area networks. *IEEE Transactions on Mobile Computing*. 2007;**6**(6):689-705

