Empirical Evaluation of Signal-Strength Fingerprint Positioning in Wireless LANs

Dimitris Milioris, Lito Kriara, Artemis Papakonstantinou, George Tzagkarakis,
Panagiotis Tsakalides and Maria Papadopouli [*]
Computer Science Department, University of Crete and
Institute of Computer Science, Foundation for Research and Technology-Hellas (FORTH)
FORTH-ICS, N. Plastira 100, Vassilika Vouton, GR 700 13 Heraklion, Crete, Greece [†]
{milioris, kriara, artpap, gtzag, tsakalid, mgp}@ics.forth.gr

ABSTRACT

This paper proposes a novel localization technique based on a multivariate Gaussian modeling of the signal strength measurements collected from several access points (APs) at different locations. It considers a discretized grid-like form of the environment and computes a signature at each cell of the grid. At run time the system compares the signature at the unknown position with the signature of each cell using the Kullback-Leibler Divergence estimation (KLD) between their corresponding probability densities. The paper evaluates the performance of the proposed technique and compares it with other statistical fingerprint-based localization systems. The performance analysis studies were conducted at the premises of a research laboratory and an aquarium under various conditions. Furthermore, the paper evaluates the impact of the number of APs and the size of the measurement datasets.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Measurement techniques

General Terms

Algorithms, Measurement, Performance

Keywords

Multivariate Gaussian model, Kullback-Leibler divergence, percentiles, RSSI measurements

1. INTRODUCTION

Location-sensing has been impelled by the emergence of location-based services in the transportation industry, emergency situations for disaster relief, the entertainment industry, and assistive technology in the medical community.

Location-sensing systems can be classified according to their dependency on and use of: (a) specialized infrastructure and hardware, (b) signal modalities, (c) training, (d) methodology and/or use of models for estimating distances, orientation, and position, (e) coordination system (absolute or relative), scale, and location description, (f) localized or remote computation, (g) mechanisms for device identification, classification, and recognition (h) accuracy and precision requirements. The distance can be estimated using time of arrival (e.g., GPS, PinPoint [36]) or signal-strength measurements, if the velocity of the signal and a signal attenuation model for the given environment, respectively, are known. Positioning systems may employ different modalities, such as, IEEE802.11 (Radar [7, 15], Ubisense, Ekahau [2]), infrared (Active Badge [34]), ultrasonic (Cricket [26, 27], Active Bat), Bluetooth [8, 13, 28, 5, 15], 4G [29], vision (EasyLiving), and physical contact with pressure (Smart Floor), touch sensors or capacitive detectors. They may also combine multiple modalities to improve the localization, such as optical, acoustic and motion attributes (e.g., [6]).

The popularity of *IEEE802.11 infrastructures*, their low deployment cost, and the advantages of using them for both communication and positioning, make them an attractive choice. Most of the signal-strength based localization systems can be classified into the following two categories, namely *signature- or map*-based and *distance-prediction*-based techniques. The first type creates a signal-strength signature or map of the physical space during a training phase and compares it with the signature generated at runtime (at the unknown position) [7, 21, 35]. To build such signatures, signal-strength data is gathered from beacons received from APs. During a training phase, such measurements are collected at various predefined positions (of the map) and signatures are generated that associate the corresponding positions of the physical space with statistical measurements based on signal-strength values acquired at those positions. Such maps can be formed with data from different sources or signal modalities to improve location-sensing [26, 15]. The distance-prediction-based techniques use the signal-strength values and radio-propagation models to predict the distance

of a wireless client from an AP (or any landmark) or even between two wireless clients (peers) with estimated position (such as CLS [32]). In situations where a deployment of a wireless infrastructure may not be feasible, positioning mechanisms may exploit cooperation by enabling devices to share positioning estimates [30, 9, 22, 32, 14, 11, 12, 36]. A survey of positioning systems can be found in [17].

This paper builds on our earlier work on CLS [32, 14]. CLS generates statistical-based fingerprints using the collected RSSI measurements from an IEEE802.11 infrastructure and also signal-strength measurements from single-hop neighboring wireless peers. The vast majority of current fingerprint positioning methods does not take into account the interdependencies among the RSSI measurements at a certain position from the various APs. These interdependencies provide important information about the geometry of the environment and can be quantified using the second-order spatial correlations among the measurements. Hence, the employment of multi-dimensional distributions is expected to provide a more accurate representation of the RSSI profiles, leading to improved positioning performance. Simple models whose parameters (second-order statistics) can be accurately and easily estimated should be used in a practical positioning scenario. This paper designs and evaluates a novel fingerprint approach based on this observation. Specifically, it makes two distinct contributions:

1. It proposes and evaluates a novel fingerprinting approach that exploits the spatial correlations of signal-strength measurements collected from various wireless APs based on a *multivariate Gaussian* model.

2. It performs a comparative performance analysis of various signal-strength fingerprinting methods and Ekahau in the premises of a research laboratory and an aquarium under different conditions.

The multivariate Gaussian-based approach takes into consideration not only the signal strength measurements from each AP but also the interplay (covariance) of measurements collected from pairs of APs. The signature comparison and position estimation is based on the Kullback-Leibler divergence (KLD): the cell corresponding to the minimum KLD is reported as the estimated position. The paper generalizes this approach by applying it iteratively in different spatial scales. It also evaluates its accuracy using empirical measurements in the premises of FORTH under different conditions. Furthermore, we ran a comparative performance analysis of various signal-strength fingerprinting approaches and Ekahau in the premises of an aquarium.

The paper is organized as follows: Section 2 presents various statistical signal-strength signature techniques. In Section 3.1, we discuss the comparative performance study of these techniques in the premises TNL, while Section 3.2 evaluates their performance in the premises of a popular aquarium. Section 4 overviews related positioning systems for mobile computing. Finally, Section 5 summarizes our main results and provides directions for future work.

2. FINGERPRINT METHODS

A wireless device that listens to a channel receives the beacons sent by APs (at that channel) periodically and records their RSSI values. Wireless devices that run fingerprint-based positioning systems acquire such measurements and generate statistical fingerprints for a position using these measurements. The statistical-based generation of fingerprints can take place using various methods, such as *confidence intervals, percentiles, the empirical distribution* or the parameters of a theoretical distribution. The physical space is represented as a grid of cells with fixed size and well-known coordinates. During a training phase, at known positions of the physical space such measurements are collected by a wireless client (*training measurements*). At each position, the wireless client scans all the available channels and listens for beacons from APs. During the runtime phase, the system also records the RSSI values from the received beacons (*runtime measurements*). As in the case of training, the wireless client scans all the available channels.

A statistical-based signature is constructed for each cell of the grid using the signal-strength measurements collected during the training phase (training signatures). Similarly, applying the same statistical method, at runtime, a statistical-based signature is also generated using the runtime measurements on-the-fly (*runtime signature*). The runtime signature is then compared with all the training signatures. The cell with a training signature that has the smallest distance from the runtime signature is reported as the estimated position. The *fingerprint of a cell* is a vector of training signatures. Each entry of the vector corresponds to one AP. The fingerprint of the unknown position is the corresponding vector of the runtime signatures. The next paragraphs present the various methods for generating statistical-based signatures used in this work.

2.1 Confidence interval

The signature is a vector of confidence intervals, each corresponding to an AP. Each confidence interval is generated using the RSSI values of the beacons received from the corresponding AP. Let us denote as $[T_i^-(c), T_i^+(c)]$ the confidence interval for AP i at cell c during the training phase. The fingerprint of a cell is the vector of these confidence intervals (for all APs) at that cell. Similarly, at run time, at the unknown position, the system records the RSSI values from a number of beacons sent by the APs and computes a confidence interval for each AP. For example, the runtime confidence interval for AP i is the $[R_i^-, R_i^+]$. The runtime fingerprint is a vector composed by all confidence intervals formed at runtime from all APs. This approach will compare the runtime fingerprint with the training fingerprint of each cell. An AP (e.g., i) participates in this technique by assigning a *vote for a cell* (e.g., c) that indicates the similarity of its training confidence interval ($[T_i^-(c), T_i^+(c)]$) with the runtime confidence interval ($[R_i^-, R_i^+]$). By adding these votes, the confidence-interval based approach computes a weight for that cell that indicates its likelihood to be the unknown position (at which the corresponding runtime measurements were collected).

At the start of the runtime phase, each cell has a zero weight. For each cell, the training confidence interval of each AP is compared with the corresponding (for that AP) runtime confidence interval. The algorithm assigns a weight at cell c $w(c)$ that indicates the likelihood that this cell is the position of the device. Each AP participates by assigning a vote to that cell. Specifically, the weight of that cell is increased by a specific value, indicated by the following criteria: In the case that the training confidence interval is included in the runtime confidence interval or the runtime

confidence interval is included in the training confidence interval, the weight of that cell is increased by one. In the case of partial overlap of these two confidence intervals, the value corresponds to the ratio of this overlap.

The cell with the maximum weight is reported as the estimated position.

2.2 Percentiles

This approach is similar to the confidence-interval one. However, instead of using confidence intervals for constructing the fingerprints, percentiles are employed. A set of percentiles can capture more detailed information about the signal strength distribution than confidence intervals, and thus, resulting to more accurate fingerprints. The weight of a cell c, w(c), is computed as follows:

$$w(c) = \sum_{i=1}^{N} \sqrt{\sum_{j=1}^{p}(R_j^i - T_j^i(c))^2} \qquad (1)$$

where N is the number of APs, p the number of percentiles, R_j^i the j-th percentile of runtime measurements from the i-th AP and $T_j^i(c)$ the j-th percentile using the training measurements from the i-th AP at the cell c.

As in the confidence interval case, the cell with the maximum weight is reported as the estimated position. In the case of the *top 5 weighted percentiles* approach, the centroid of the top five cells with the largest weight is reported as the estimated position.

2.3 Empirical distribution

The signature of a cell is a vector of size equal to the number of APs that appear in both the training and runtime measurements. Each entry of a training (runtime) signature corresponds to the complete set of RSSI values collected during the training (runtime) phase, respectively.

This method creates a signature based on the set of signal-strength measurements collected at each cell from all APs. At runtime, at an unknown position, each cell is assigned a *weight* which corresponds to the *average* empirical KLD distance of each AP (at that cell) from the runtime measurements collected at the unknown position from the same AP. The cell with the smallest weight is reported as the position.

2.4 Multivariate Gaussian model

Unlike other fingerprint positioning methods, this one focuses on the interdependencies among the RSSI measurements in a cell from various APs. These interdependencies provide information about the geometry/topology of the environment and can be quantified using the second-order spatial correlations among the measurements. According to this proposed approach, in the training phase, a statistical signature is extracted for each cell of the grid by modeling the acquired signal-strength measurements using a multivariate Gaussian distribution. The density function of a multivariate Gaussian in $\mathbb{R}^K$, with a mean vector $\vec{\mu}$ and covariance matrix $\mathbf{\Sigma}$, is given by:

$$p(\vec{x}|\vec{\mu}, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{K/2}|\mathbf{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^T \mathbf{\Sigma}^{-1}(\vec{x} - \vec{\mu})\right), \qquad (2)$$

where $|\mathbf{\Sigma}|$ is the determinant of $\mathbf{\Sigma}$.

Let N be the number of APs from which the mobile device receives the measurements, K be the number of measurements from each AP, and $\mathbf{S}_i = [\vec{y}_1, \ldots, \vec{y}_N]$ denote the $K \times N$ matrix for the i-th cell c_i, whose j-th column $\vec{y}_j \in \mathbb{R}^K$ contains the received signal-strength values from the j-th AP. The signal-strength measurements are modeled by a multivariate Gaussian distribution due to its simplicity and the closed-form expression of the associated similarity measure (KLD). More specifically, the signature $\mathcal{S}_i$ of the i-th cell is given by:

$$c_i \mapsto \mathcal{S}_i = \{\vec{\mu}_i, \mathbf{\Sigma}_i\} , \qquad (3)$$

where $\vec{\mu}_i = [\mu_{i,1}, \ldots, \mu_{i,N}]$, with $\mu_{i,j}$ being the mean of the j-th column of the measurement matrix $\mathbf{S}_i$, and $\mathbf{\Sigma}_i$ is the corresponding covariance matrix, with its mn-th element being equal to the covariance between the m-th and n-th columns of $\mathbf{S}_i$. Hence, the mn-th element of matrix Σ_i denotes the *spatial correlation* between the RSSI measurements of the i-th cell from the $m-$th and n-th APs. Thus, if L is the number of cells in the grid representing the physical space, during the training phase, the following set of *training* signatures (T) is generated:

$$\{\mathcal{S}_{i,T}\}_{i=1}^{L} = \left\{\{\vec{\mu}_{i,T}, \mathbf{\Sigma}_{i,T}\}\right\}_{i=1}^{L} . \qquad (4)$$

In addition, the i-th cell, $c_{i,T}$, is also associated to a set of indices $I_{i,T}$ indicating its corresponding "active" APs, that is, the APs from which it acquires the measurements during the training phase.

During the *run-time* phase (R), we assume that the mobile user is placed at an unknown cell (c_R), whose location must be estimated. Following the approach used in the training phase, if $\mathbf{S}_R = [\vec{y}_{1,R}, \ldots, \vec{y}_{N',R}]$ is the $K' \times N'$ run-time signal-strength measurement matrix of c_R, a signature is generated as follows,

$$c_R \mapsto \mathcal{S}_R = \{\vec{\mu}_R, \mathbf{\Sigma}_R\} . \qquad (5)$$

Notice here that in general the dimensions of the run time measurement matrix are smaller than the dimensions of the corresponding training matrix ($K \times N$). This is due to the fact that during run time it is more difficult to collect extensive measurements than during training. Furthermore, the set of APs operating during the training phase is *not* necessarily the same with the set of APs at runtime. Let us denote as $I_R^{i,T}$ the set of APs from which signal strength measurements were collected *both at runtime and also during training* at cell i. For the run-time (c_R) and the i-th training cell ($c_{i,T}$), we extract their corresponding mean subvectors $\vec{\mu}_R^s$, $\vec{\mu}_{i,T}^s$ and covariance sub-matrices $\mathbf{\Sigma}_R^s$, $\mathbf{\Sigma}_{i,T}^s$ according to the indices of $I_R^{i,T}$. Finally, if $p_R(\vec{x}|\vec{\mu}_R^s, \mathbf{\Sigma}_R^s)$ and $p_{i,T}(\vec{x}|\vec{\mu}_{i,T}^s, \mathbf{\Sigma}_{i,T}^s)$ denote the multivariate Gaussian densities of c_R and $c_{i,T}$, respectively, their KLD is given by the following closed-form expression:

$$\begin{aligned} D(p_R||p_{i,T}) &= \frac{1}{2}\left((\vec{\mu}_{i,T}^s - \vec{\mu}_R^s)^T(\mathbf{\Sigma}_{i,T}^s)^{-1}(\vec{\mu}_{i,T}^s - \vec{\mu}_R^s)\right.\\ &\left.+ tr\big(\mathbf{\Sigma}_R^s(\mathbf{\Sigma}_{i,T}^s)^{-1} - \mathbf{I}\big) - \ln|\mathbf{\Sigma}_R^s(\mathbf{\Sigma}_{i,T}^s)^{-1}|\right), \end{aligned} \qquad (6)$$

where $tr(\cdot)$ denotes the trace of a matrix (sum of its diagonal elements) and $\mathbf{I}$ is the identity matrix. KLD is a (non-symmetric) measure of the difference between two probability distributions, well established and widely used in probability and information theory. The estimated location $[x_R^*, y_R^*]$ is given by the coordinates of the i^*-th cell, which

minimizes (6), that is,

$$i^* = \arg \min_{i=1,...,L} D(p_R||p_{i,T}) . \qquad (7)$$

Algorithm 1 The multivariate Gaussian-based positioning method (spatial scale of a cell)

1. During training phase, collect RSSI measurements from APs at each cell
   trainingAP(c): set of APs from which data are collected at cell c

2. During run time, collect RSSI measurements from each AP at the unknown position
   runtimeAPs: set of APs from data are collected
   $effectiveAP(c) : trainingAP(c) \cap runtimeAP$

3. During runtime, perform the following steps for each cell c:

   - Generate the signature for cell c using only training measurements collected from APs $\in$ effectiveAP(c) (i.e., *training signature(c))*

   - *Generate the runtime signature using only runtime measurements collected from APs in effectiveAP(c) (i.e.,* runtime signature)

   - Estimate the KLD distance of the training and runtime signatures

4. Report as the estimated position the cell c^* with the smallest KLD distance

---

We performed a preliminary evaluation of Algorithm 1 using measurements from an IEEE802.11 infrastructure. The algorithm did not always estimate correctly the cell due to the radio propagation characteristics in the physical space, affected by transient phenomena. It was not uncommon to have a cell which is located far away from the unknown position with a training fingerprint very close to the runtime fingerprint.

To improve the accuracy, we proposed a generalization of this approach: instead of applying the multi-variate Gaussian *per cell*, we apply it in an iterative fashion in multiple spatial scales (e.g., *regions*). First, the physical space is divided into overlapping regions of size larger than the size of a cell and the multivariate Gaussian algorithm is applied for each region separately. To generate the fingerprint of a region, we employ all the signal-strength measurements from all APs collected at positions within that region. This *spatial aggregation* reduces the likelihood of selecting a false region/cell (a region/cell that does not include/correspond to the actual position) over the correct one. Essentially, via this aggregation an incorrect region is eliminated (in the first iteration) while the "weight" of the correct region is enhanced by considering the signatures of the neighboring to the actual position cells. The region-based multivariate Gaussian algorithm proceeds iteratively: after it estimates the region at which the device is located, it repeats the process by dividing the selected region into sub-regions and applying the algorithm on them. In this paper, a subregion corresponds to a cell (two spatial granularities/scales).

The original area of interest is discretized in G regions, each of N cells. Let A_i be a $GK \times N$ matrix whose j-th column ($\in \mathbb{R}^{GK}$) contains the received signal-strength values from the j-th AP collected at cells of region i during training.

Let us denote with $A_{i,T}(\vec{x}|\vec{\mu}_{i,T}^s, \mathbf{\Sigma}_{i,T}^s)$ the multivariate Gaussian density of region i and $p_R(\vec{x}|\vec{\mu}_R^s, \mathbf{\Sigma}_R^s)$ the multivariate Gaussian density of the unknown position (runtime signature). The KLD distance can be computed as in (6) and the region closest to the unknown position is given by

$$i_A^* = \arg \min_{i=1,...,G} D(p_R||A_{i,T}) \qquad (8)$$

After the estimation of the correct region, the process is repeated (using the Algorithm 1) to compute the cell in that region that corresponds to the unknown position (considering only the cells of that region).

3. PERFORMANCE ANALYSIS

3.1 Evaluation at FORTH

The empirical-based evaluation at FORTH took place in the Telecommunication and Networks Lab (TNL), an area of 7m×12m, which was discretized in a grid structure with cells of 55cm×55cm.

During the training phase, we collected two data sets: one during a relatively busy period and another one during a quiet period. The busy period corresponds to a period around 3pm on typical weekdays, during which there were at least five people in the laboratory, and several others walking in the hallways outside. The quiet period was at around 11pm on the same weekdays as the busy period dataset. During that time period, there was only one person in the laboratory. The *busy (quiet) period dataset* included measurements from 108 (104) different cells and 13 (12) APs, respectively. On average 6 APs were detected at a given cell and more than 300 RSSI values were collected at each cell per AP. To generate the training signatures, signal-strength values at various cells of the grid were collected. The trainer remained still for approximately 90s and 30s to collect beacons at each position during training and runtime, respectively. To capture signal strength values, *iwlist*, which polls each channel and acquires the MAC address and RSSI measurements from each AP (in dBm), and the *tcpdump*, a passive scanner relying on libpcap, for the retrieval of each packet were used. A Sony Vaio and Fujitsu Siemens Tablet PC with the same wireless adapter were used for both the training and run-time experiments.

In order to evaluate the performance of the various fingerprinting methods, we computed the *localization error*, measured as the Euclidean distance between the centers of the reported cell and the cell at which the mobile user was actually located at run time. We ran 30 measurements at different positions (run time cells).

Figures 1(a) and 1(b) illustrate the localization error of the different signature-based approaches during busy and quiet period, respectively. The multivariate Gaussian model (MvGs) outperforms the percentiles, confidence interval (90%), and empirical distribution approaches. Specifically, for the quiet period datatrace, the median error is 4.16m and 2.91m for the confidence interval, and percentiles, respectively, while the MvGs results in a median error of 1.72m. For the busy period dataset, the median error of the MvGs is 1.60m while the others report a median error of 2.82m and 2.65m, for the confidence interval (90%) and percentiles, respectively.

In general, it is expected that as the number of APs
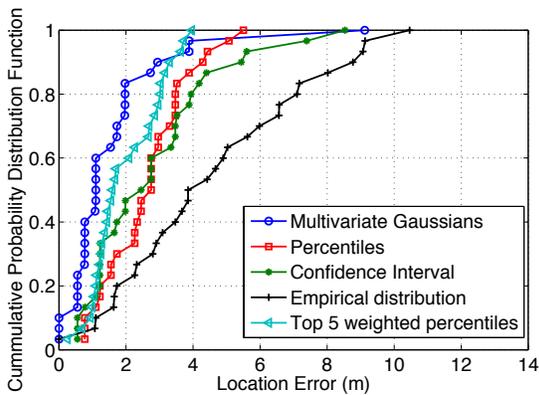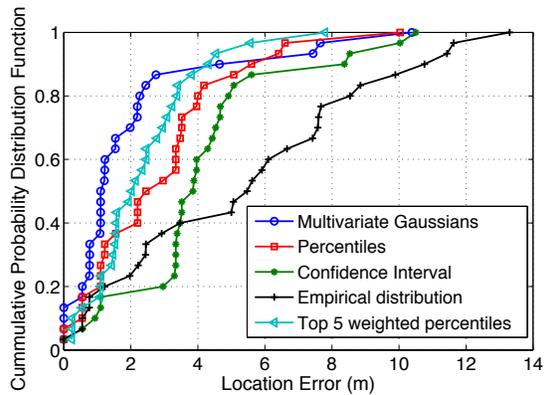
(a) Busy period dataset   (b) Quiet period dataset

Figure 1: The performance of various fingerprint positioning methods at FORTH.

that participate in the signature generation increases, it will become "easier" to distinguish the correct cell from other further-away cells, which may have similar training fingerprint with the runtime one due to transient phenomena or radio propagation characteristics in the given environment. To measure the impact of the number of APs on the localization accuracy, we associate each AP with a *popularity index* that indicates the number of cells at which there were measurements (from that AP) at both training and at runtime. For example, the popularity index of AP i is $|\{c|APi \in effectiveAP(c)\}|$. The APs were sorted in a decreasing order based on their popularity index. The analysis was repeated using the top k most popular APs for the busy period and the quiet period datasets.

Figure 2 shows the impact of the number of APs on location error. The higher the number of APs, the lower the location error. However the impact of the number of APs diminishes after a certain threshold. The busy period dataset is subject to a larger number of transient phenomena than the quiet period, affecting the performance of fingerprinting. Thus, the impact of the number of APs is more prominent in the busy period dataset than in the quiet period one. For example, in the busy period dataset, the improvement in the location error when the number of APs becomes six is about 80cm, while when the number of APs increases from six to 13, the location error is reduced by only 20cm. Similar results were also observed in the case of percentiles [19].

Figures 3(a) and 3(b) illustrate the impact of the measurement size on the accuracy of the multivariate Gaussian-based method. The % indicates the percentage of measurements considered in both training and runtime datasets out of the corresponding original datasets (used in the other plots). In general, the larger the measurement set, the more accurate the position estimation.

3.2 Evaluation at the Aquarium

Cretaquarium is the largest and most popular aquarium in Greece, covering an area of $1760m^2$. During the period of our study, it included 30 tanks, while another 25 were being installed. There was an wireless infrastructure of eight IEEE802.11 APs.

The physical space was represented as a grid with cells of $1m \times 1m$. Training and runtime signal-strength measurements were collected for the entire testbed during two
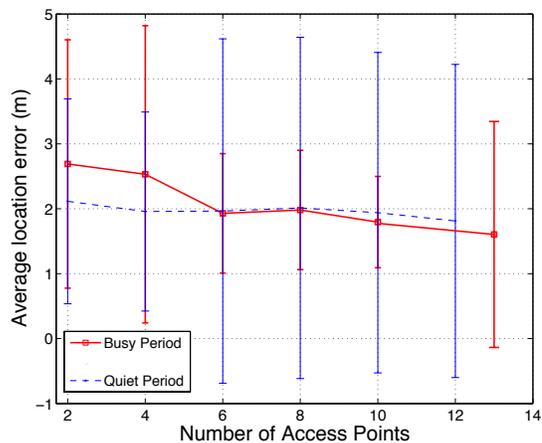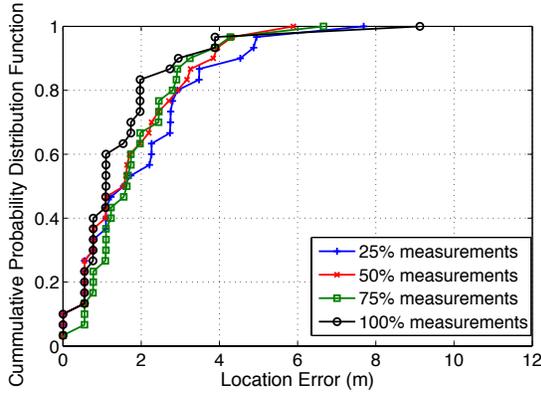

Figure 2: Impact of the number of APs on location error. The x-axis indicates the number of the top x APs considered in both training and runtime datasets.

different periods (normal and busy periods). At each cell, we collected measurements from an average number of 5.7 APs. During the normal and busy periods when the data sets were collected, there were about 100 and 250 visitors present in the aquarium, respectively.

Under normal conditions, the median location error using the percentiles method was about 2m, while the confidence interval reported a median location error of 3.6m. Under the busy period, the median location error for the confidence-interval method was about 4.3m. Figure 5 illustrates the performance of the various fingerprint methods during a busy period in the aquarium. As we mentioned before, the largest the measurement set, the more accurate the position estimation. The relatively small signal-strength measurement data set has a noticeable impact on the performance of the multivariate Gaussian method.

A guiding application was designed for the aquarium to provide personalized information to visitors about the habitats in the tank in front of them. For that, the physical
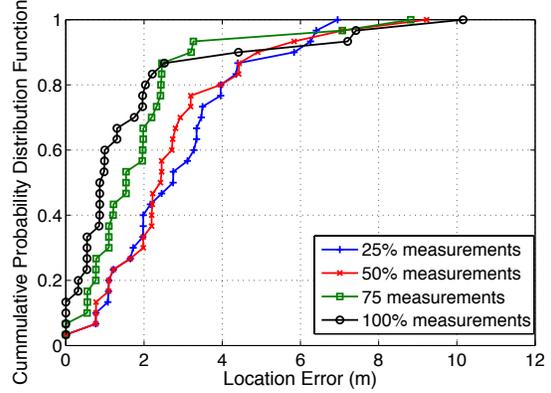
(a) Busy period dataset   (b) Quiet period dataset

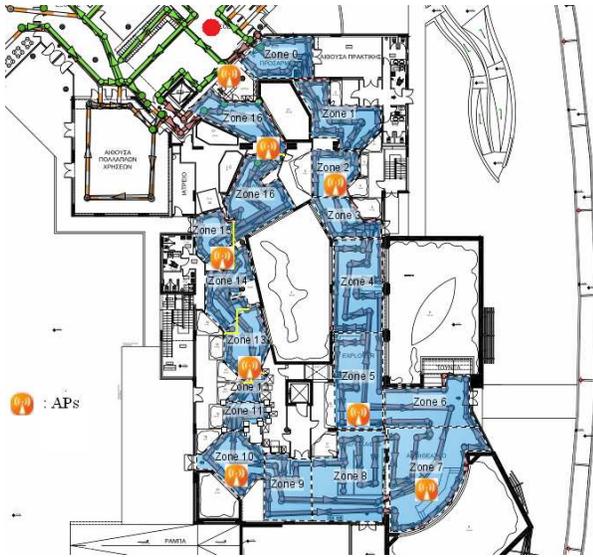**Figure 3: Impact of the number of signal-strength measurements on location error.**
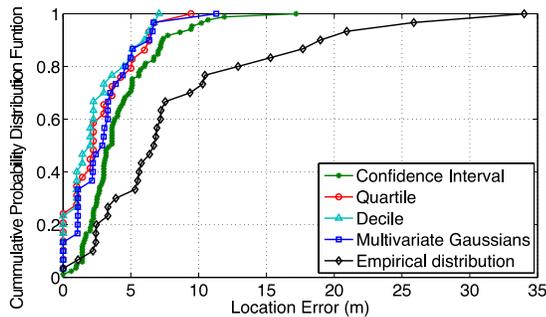

Figure 4: CreteAquarium divided into zones.

Figure 5: The performance of various fingerprint positioning methods at the aquarium during a busy period.

space was divided into 17 zones, according to the application requirement. The positioning system reported the zone in which the visitor was located. As shown in Figure 4, the

entrance is located at the top left corner of the floor plan (red dot). Zone 0 is the first area that people visit. The tanks are located at the center and at each side of the main area of the aquarium. Visitors follow a "clockwise" path, bounded by tanks, and finally arrive at the last zone (exit), located at the top left corner of the floor plan. Furthermore, we conducted experiments using the confidence interval and also the Ekahau, a commercial positioning system that also employs signal-strength-based fingerprints. The tests took place at the same period and for the same runtime cells. During that period, about 200-250 visitors were present in the aquarium (e.g., corresponding to a busy period dataset).

In each zone, the systems were tested at three different positions. The confidence-interval-based method reported the correct zone 91% of the times, while Ekahau in 80% of the times and had a median location error of 4.6m. A more detailed performance analysis of CLS using percentiles and confidence intervals at the premises of FORTH and the Aquarium can be found in [19].

4. RELATED WORK

Recently significant work has been published in the area of location-sensing using RF signals. Like CLS, Radar [7] employs signal-strength maps that integrate signal-strength measurements acquired during the training phase from APs at different positions with the physical coordinates of each position. Each measured signal-strength vector is compared against the reference map and the coordinates of the best match will be reported as the estimated position. Bahl *et al.* [24] improved Radar to alleviate side effects that are inherent properties of the signal-strength nature, such as aliasing and multipath. Ladd *et al.* [21] proposed another location-sensing algorithm that utilizes the IEEE802.11 infrastructure. In its first step, a host employs a probabilistic model to compute the conditional probability of its location for a number of different locations, based on the received signal-strength measurements from nine APs. The second step exploits the limited maximum speed of mobile users to refine the results and reject solutions with a significant change in the location of the mobile host. Kung *et al.* [20] propose a method for evaluating the impact of the IEEE802.11 APs on positioning in order to strengthen the role/contribution of a "good" AP while "de-emphasizing"

the role of the "bad" APs. The "goodness" of an AP indicates the capability of that AP to estimate accurately its distance from the others. Youssef *et al.* Horus [35] substantially improved the accuracy (e.g., an 1.3m error in 90% of their experiments) by employing an autoregressive model that captures the autocorrelation in signal strength measurements of the same AP at a particular location. Specifically, the time series generated from signal-strength measurements collected from an AP is represented by a first-order autoregressive model. The fingerprints are formed for each cell and AP based on the degree of autocorrelation, the mean, and the variance of the empirical measurements collected from that AP at that cell. Finally, an interesting approach proposes fingerprints based on attributes that characterize the effects of multipath (e.g., channel response) in order to detect changes of the positions of wireless hosts were presented in [25, 37].

Niculescu and Badri Nath [23] designed and evaluated a cooperative location-sensing system that uses specialized hardware for calculating the angle between two hosts in an ad-hoc network. This can be done through antenna arrays or ultrasound receivers. Hosts gather data, estimate their position, and propagate them throughout the network. Previously, these authors [22] introduced a cooperative location-sensing system in which position information of landmarks is propagated towards hosts that are further away, while during this process, hosts may further enrich this information by determining their own location. Another location-sensing system in ad-hoc networks performs positioning without the use of landmarks or GPS and presents the tradeoffs among internal parameters of the system [9]. The location-sensing systems presented in [30] and [16] are the closest to CLS and are compared in detail in [14].

Active Badge [33] uses diffuse infrared technology and requires each person to wear a small infrared badge that emits a globally unique identifier every ten seconds or on demand. A central server collects this data from fixed infrared sensors around the building, aggregates it and provides an application programming interface for using the data. The system suffers in the case of fluorescent lighting and direct sunlight, because of the spurious infrared emissions these light sources generate. A different approach, SmartFloor [3], employs a pressure sensor grid installed in all floors to determine presence information. In a building without requiring users to wear tags or carry devices, albeit without being able to identify individuals.

Examples of localization systems that combine multiple technologies are UbiSense [4], Active Bats [1] and Surround-Sense [6]. UbiSense can provide a high accuracy using a network of ultra wide band (UWB) sensors installed and connected into a building existing network. The UWB sensors use Ethernet for timing and synchronization. They detect and react to the position of tags based on time difference of arrival and angle of arrival. An RFtag is a silicon chip that emits an electronic signal in the presence of the energy field created by a reader device in proximity. Location can be deduced by considering the last reader to see the card. RFID proximity cards are in widespread use, especially in access control systems. The Active Bats architecture consists of a controller that sends a radio signal and a synchronized reset signal simultaneously to the ceiling sensors using a wired serial network. Bats respond to the radio request with an ultrasonic beacon. Ceiling sensors measure time-of-flight from

reset to ultrasonic pulse. Active Bat applies statistical pruning to eliminate erroneous sensor measurements caused by a sensor hearing a reflected pulse instead of one that travelled along the direct path from the Bat to the sensor. A relatively dense deployment of ultrasound sensors in the ceiling can provide within 9 cm of the true position for 95% of the measurements. SurroundSense runs on a mobile phone to provide logical localization by generating fingerprints using sound, accelerometers, cameras and IEEE802.11. Tesoriero *et al.* [31] propose a passive RFID-based indoor location system that is able to accurately locate autonomous entities, such as robots and people, within a physical-space.

Ariadne [18] is an automated location determination system. It uses a two dimensional construction floor plan and only a single actual strength measurement. It generates an estimated signal strength map comparable to those generated manually by actual measurements. Given the signal measurements for a mobile, a proposed clustering algorithm searches that signal strength map to determine the current mobile's location.

In a recent work [10], the problem of indoor location estimation is also treated in a probabilistic framework. In particular, a reduced number of locations sampled to construct a radio map is employed in conjunction with an interpolation method, which is developed to effectively patch the radio map. Furthermore, a Hidden Markov Model (HMM) that exploits the user traces to compensate for the loss of accuracy is employed to achieve further improvement of the radio map due to motion constraints, which could confine possible location changes. Both the proposed multivariate Gaussian model-based algorithm and the HMM-based approach belong to the class of the probabilistic localization techniques. Usually, a probabilistic localization method is characterized by an increased performance when compared with a deterministic one, since it provides not only a point estimate of the user's position but also gives a confidence interval for the quality of this estimate. This can be used to improve further the estimation accuracy with the goal of reducing the uncertainty. However, a first key observation is the highly reduced complexity of our method compared to the HMM-based approach. In particular, it is a one-iteration method, where in each iteration only the simple estimate of a mean vector, a covariance matrix, and the computation of the Kullback-Leibler divergence between multivariate Gaussians (given in closed form) are required. On the other hand, the HMM-based localization technique requires several iterations to converge, while in each iteration several model parameters have to be estimated (approximately of the same dimensions as the parameters of our proposed method). However, the reduced computational complexity of the Gaussian-based technique comes at the cost of a potentially degraded location estimate under certain circumstances. For instance, in the case of corrupted measurements (e.g., due to an access-point failure or the presence of an obstacle) our method is much more sensitive, since it is based on measurements collected instantaneously. In contrast, the HMM-based approach could provide a more accurate estimate via the prior knowledge of a transition-probability matrix, which is preserved and re-estimated in each iteration in conjunction with the refinement achieved by an Expectation Maximization algorithm. In conclusion, the major benefit of our proposed algorithm , when compared with the HMM-based approach, is the significantly reduced computational

complexity and implementation simplicity, as well as the high accuracy in several specific environments (obstacle-free, robust measurements) as it was revealed by the experimental evaluation. On the other hand, the HMM-based approach can be proved to be more robust in the case of system failures, but at the cost of requiring increased computational resources.

5. CONCLUSIONS

This paper introduced a novel localization method that creates signal-strength fingerprints using multivariate Gaussian distributions. It estimates the position of the device by computing the region with the training fingerprint that has the closest KLD distance from the runtime fingerprint. The empirical-based evaluation revealed that the multivariate Gaussian method outperforms other signal-strength fingerprint approaches. We performed an evaluation of various fingerprint methods in the premises of FORTH and an aquarium. The median position error varies from 1.7m to 4.6m, depending on the premise and the conditions. The presence of people as well as the density and placement of APs have a prominent impact on positioning. Furthermore, in the case of the multivariate Gaussian-based algorithm we experimented with a multiple spatial scale iterative approach in which, we applied the algorithm on larger regions, to select the correct one, and then within the selected region to estimate the correct cell. Something similar was performed in the case of percentiles by selecting of the top 5 best candidate cells. We showed that it improves the accuracy by eliminating the distant incorrect cells and taking also into consideration the "impact" of neighboring cells around the correct one. Other related work has also shown that the integration of user mobility models can further improve the accuracy. In the context of the aquarium, in which mobility patterns do exist, the integration of user mobility models could be helpful.

We have been also experimenting with other modalities, such as infrared, cameras and QR codes to improve the location estimation. Specifically, in front of each landmark (e.g., tank of the aquarium or office in TNL), a unique QR code can be placed along with three infrared sensors (e.g., WII bar). The camera of the mobile device of a visitor may capture the QR code, recognize it, and thus identify the landmark, in front of which this visitor is standing. Similarly, when the camera captures the infrared light from at least two infrared sources, it can estimate its distance from the landmark by measuring the distance of the two infrared sources on the recorded image. We plan to extend our localization system by incorporating these multi-modalities measurements.

There is a growing interest in statistical methods that exploit various spatio-temporal statistical properties of the received signal to form robust fingerprints. In general, a channel exhibits very transient phenomena and is highly time-varying. At the same time, the collection of signal measurements is subject to inaccuracies due to various issues, such as hardware mis-configurations, limitations, time synchronization, fine-grain data sampling, incomplete information, and vendor-specific dependencies (often not publicly available). Thus, the general problem of building a theoretical framework to analyze these fingerprint techniques taking into consideration the above limitations opens up exciting research opportunities.

6. REFERENCES

[1] The bat ultrasonic location system. http://www.cl.cam.ac.uk/research/dtg/attarchive/bat/.

[2] Ekahau v.3.1. (http://www.ekahau.com).

[3] Future computing environments. http://www.cc.gatech.edu/fce/smartfloor/.

[4] Precise Real-time Location. http://www.ubisense.net/.

[5] ASTHANA, S., AND KALOFONOS, D. The problem of bluetooth pollution and accelerating connectivity in bluetooth ad-hoc networks. In *IEEE PerCom* (New York, NY, USA, 2005).

[6] AZIZYAN, M., CONSTANDACHE, I., AND CHOUDHURY, R. R. SurroundSense: Mobile phone localization via ambience fingerprinting. *In ACM MobiCom* (September 20-25 2009).

[7] BAHL, P., AND RADMANABHAN, V. Radar: An in-building rf-based user location and tracking system. *In IEEE InfoCom* (March 2000).

[8] BANDARA, U., HASEGAWA, M., INOUE, M., MORIKAWA, H., AND AOYAMA, T. Design and implementation of a bluetooth signal strength based location sensing system. In *IEEE (RAWCON)* (Atlanta, GA, USA, Sept. 2004).

[9] CAPKUN, S., HAMDI, M., AND HUBAUX, J.-P. GPS-Free Positioning in Mobile Ad-Hoc Networks. In *Proceedings of Hawaii International Conference On System Sciences* (Hawaii, Jan. 2001).

[10] CHAI, X., AND YANG, Q. Reducing the calibration effort for probabilistic indoor location estimation. *In IEEE Transactions on Mobile Computing* (June 2007).

[11] CHINTALAPUDI, K., GOVINDAN, R., SUKHATME, G., AND DHARIWAL, A. Ad-hoc localization using ranging and sectoring. In *IEEE InfoCom* (Hong Kong, March 2004).

[12] FANG, L., DU, W., AND NING, P. A beacon-less location discovery scheme for wireless sensor networks. In *IEEE InfoCom* (Miami, Florida, March 2005), pp. 161–171.

[13] FELDMANN, S., KYAMAKYA, K., ZAPATER, A., AND LUE, Z. An indoor Bluetooth-based positioning system: concept, implementation and experimental evaluation. In *International Conference on Wireless Networks* (Las Vegas, Nevada, USA, 2003), pp. 109–113.

[14] FRETZAGIAS, C., AND PAPADOPOULI, M. Cooperative Location Sensing for Wireless Networks. In *IEEE PerCom* (Orlando, Florida, March 2004).

[15] GWON, Y., JAIN, R., AND KAWAHARA, T. Robust indoor location estimation of stationary and mobile users. In *In IEEE InfoCom* (Hong Kong, March 2004).

[16] HE, T., HUANG, C., BLUM, B. M., STANKOVIC, J. A., AND ABDELZAHER, T. Range-free localization schemes for large-scale sensor networks. In *ACM MobiCom* (San Diego, CA, USA, Sep 2003).

[17] HIGHTOWER, J., AND BORRIELLO, G. A Survey and Taxonomy of Location Sensing Systems for Ubiquitous Computing. Technical Report, University of Washington, Department of Computer Science and Engineering UW CSE 01-08-03, Seattle, WA, Aug. 2001.

[18] JI, Y., BIAZ, S., PANDEY, S., AND AGRAWAL, P.

Ariadne: A dynamic indoor signal map construction and localization system. *In ACM MobiSys* (June 19-22 2006).

[19] KRIARA, L. Experimenting with the fingerprinting method using signal-based measurements for providing positioning information to location-based applications. Master's thesis, Heraklion, Crete, Greece, July 2009.

[20] KUNG, H. T., LIN, C.-K., LIN, T.-H., AND VLAH, D. Localization with snap-inducing shaped residuals (sisr): Coping with errors in measurement. *In ACM MobiCom* (September 20-25 2009).

[21] LADD, A. M., BEKRIS, K., RUDYS, A., MARCEAU, G., KAVRAKI, L. E., AND WALLACH, D. Robotics-Based Location Sensing using Wireless Ethernet. In *ACM MobiCom* (Atlanta, GE, USA, Sep 2002).

[22] NICULESCU, D., AND NATH, B. Ad Hoc Positioning System (APS). In *IEEE GlobeCom* (San Antonio, TX, Nov 2001).

[23] NICULESCU, D., AND NATH, B. Ad Hoc Positioning System (APS) using AoA. In *In IEEE InfoCom* (San Francisco,CA, Apr 2003).

[24] PARAMVIR BAHL, V. N. P., AND BALACHANDRAN, A. Enhancements to the radar user location and tracking system. *Technical Report, Microsoft Research* (February 2000).

[25] PATWARI, N., AND KASERA, S. K. Robust location distinction using temporal link signatures. *In ACM MobiCom* (September 9-14 2007).

[26] PRIYANTHA, N. B., CHAKRABORTY, A., AND BALAKRISHNAN, H. The cricket location-support system. *In ACM MobiCom* (August 2000).

[27] PRIYANTHA, N. B., MIU, A. K. L., BALAKRISHNAN, H., AND TELLER, S. The Cricket compass for context-aware mobile applications. In *ACM MobiCom* (Rome, Italy, July 2001), pp. 1–14.

[28] RODRIGUEZ, M., PECE, J. P., AND ESCUDERO, C. J. In-building location using bluetooth. In *International Workshop on Wireless Ad-hoc Networks* (Coruna, Spain, May 2005).

[29] ROY, A., MISRA, A., AND DAS, S. K. An information theoretic framework for optimal location tracking in multi-system 4G wireless networks. In *IEEE InfoCom* (Hong Kong, March 2004).

[30] SAVARESE, C., RABAEY, J., AND LANGENDOEN, K. Robust positioning algorithms for distributed ad-hoc wireless sensor networks. In *USENIX* (Monterey, CA, June 2002).

[31] TESORIERO, M., TEBARA, R., GALLUD, J., LOZANOA, M., AND PENICHET, V. Improving location awareness in indoor spaces using RFID technology. *In Expert System Application*.

[32] VANDIKAS, K., KRIARA, L., PAPAKONSTANTINOU, T., KATRANIDOU, A., BALTZAKIS, H., AND PAPADOPOULI, M. Empirical-based analysis of a cooperative location-sensing system. In *ACM First International Conference on Autonomic Computing and Communication Systems (Autonomics)* (Rome, Italy, Oct. 2007).

[33] WANT, R., AND HOPPER, A. Active badges and personal interactive computing objects. *Technical Report ORL 92-2, Olivetti Research, also in IEEE Transactions on Consumer Electronics* (February 1992).

[34] WANT, R., HOPPER, A., FALCAO, V., AND GIBBONS, J. The active badge location system, 10(1), 91-102, January 1992.

[35] YOUSSEF, M., AND AGRAWALA, A. The horus WLAN location determination system. *In ACM MobiSys* (June 6-8 2005).

[36] YOUSSEF, M., YOUSSEF, A., RIEGER, C., SHANKAR, U., AND AGRAWALA, A. PinPoint: An asynchronous time-based location determination system. In *ACM MobiSys* (Uppsala, Sweden, June 2006), pp. 165–176.

[37] ZHANG, J., FIROOZ, M. H., PATWARI, N., AND KASERA, S. K. Advancing wireless link signatures for location distinction. *In ACM MobiCom* (September 14-19 2008).