KL-Divergence Kernel Regression for Non-Gaussian Fingerprint Based Localization

Piotr Mirowski, Harald Steck, Philip Whiting, Ravishankar Palaniappan, Michael MacDonald and Tin Kam Ho
Bell Laboratories, Alcatel-Lucent
600 Mountain Avenue, Murray Hill, NJ 07974, USA
Email: see http://ect.bell-labs.com/who/piotr

***Abstract*—Various methods have been developed for indoor localization using WLAN signals. Algorithms that fingerprint the Received Signal Strength Indication (RSSI) of WiFi for different locations can achieve tracking accuracies of the order of a few meters. RSSI fingerprinting suffers though from two main limitations: first, as the signal environment changes, so does the fingerprint database, which requires regular updates; second, it has been reported that, in practice, certain devices record more complex (e.g bimodal) distributions of WiFi signals, precluding algorithms based on the mean RSSI. In this article, we propose a simple methodology that takes into account the full distribution for computing similarities among fingerprints using Kullback-Leibler divergence, and that performs localization through kernel regression. Our method provides a natural way of smoothing over time and trajectories. Moreover, we propose unsupervised KL-divergence-based recalibration of the training fingerprints. Finally, we apply our method to work with histograms of WiFi connections to access points, ignoring RSSI distributions, and thus removing the need for recalibration. We demonstrate that our results outperform nearest neighbors or Kalman and Particle Filters, achieving up to 1m accuracy in office environments. We also show that our method generalizes to non-Gaussian RSSI distributions.**

***Index Terms*—Signal Strength, WiFi, Fingerprinting, Localization, Distributions, Kernel Methods**

I. Introduction

Indoor tracking of people and objects based on radio signal strength measurements can be performed with an accuracy of a few meters in a typical building. As a first step, localization methods require laborious human involvement in the training phase to build so-called *fingerprint* maps for each Access Point (AP). In predictive mode, the Received Signal Strength Indicators (RSSI) from visible APs are matched to the fingerprints to estimate the location of a person or object. Typical algorithms such as nearest neighbor matching [1] may involve solely the RSSI; other techniques take advantage of time-stamping and of assumptions about the motion, and resort to state-space models and dynamic system inference, such as in Kalman or particle filtering [6].

Those fingerprint maps however generally store only the mean value of RSSI [4], [6] and do not exploit information about the fluctuations of RSSI in the environment. In practice, however, we noticed that certain devices record more complex distributions, complicating the fingerprinting process and introducing errors at estimation. Moreover, frequent retraining is necessary to maintain accuracy. Finally, some APs may be no longer visible during estimation, for instance due to equipment failures or their roles in mobile ad-hoc networks.

978-1-4577-1804-5/11/$26.00 © 2011 IEEE

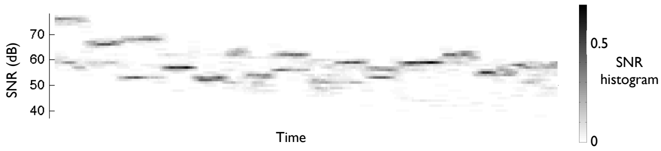

Fig. 1. Non-Gaussian Distributions of the Signal-to-Noise Ratio (SNR) of the RSSI. Data were recorded over 30min along a long corridor and for a single AP. The mobile would alternately stop for about two minutes at each location and move one meter further, repeating these steps for about 15 locations. The histograms have one bin per SNR level, and were constructed using 60s sliding windows and 10s steps.

A. The Challenge of Non-Gaussianity

A common assumption about the RSSI coming from multiple APs is that the signals are distributed as multivariate Gaussians. It has however been reported [17] that this is not always the case: the signal can be multimodal, or different recording devices can measure quite different distributions at the same location. In our experiments, we noticed that the RSSI can be distributed in a bimodal way, oscillating between two values distant by as much as 10dB, as illustrated on Fig. 1.

Presumably, the use of mean and variance of a multimodal distribution may ignore important information that is helpful for discriminating amomg different locations. We therefore look for a procedure that can provide a richer characterization of the distribution. We represent the RSSI or Signal-to-Noise Ratio (SNR) distributions by histograms. Because the RSSI values recorded by such software as NetStumbler® (http://www.netstumbler.com) are integers, we propose the natural binning scheme of one bin for each integer level[1]. In the most general case that accounts for the multi-modality of the signals we consider multinomial distributions as our model for RSSI distributions. In order to compare such multimodal distributions, we propose to use the Kullback-Leibler (KL) divergence.

B. Proposed Improvements

In this paper, we propose a probability kernel-based approach to matching fingerprints, where each fingerprint is

[1] We evaluated the trade-off between coarser binning schemes, e.g., 5dB bins, and time window lengths in the results section.

associated with a location in the fingerprint database. Matching is done by comparing distributions using the symmetrized Kullback-Leibler divergence and by constructing probability kernels that can be used either in a simple weighted regression scheme. We found that this metric on fingerprints is robust to various noise and RSSI distributions, and we provide means to estimate the location using RSSI measurements during a short time window. As an extension, we also propose an alternative approach to fingerprinting, which records only the count of successful connections to APs (rather than the RSSI levels) over a small time interval, a method similar in principle to AP coverage area estimates [9].

C. Prior Art in Probability-Based Indoor Localization

The first usage of a probabilistic approach to RSSI in indoor localization was explained in [3], [16]. They proposed to model the distribution of RSSI at each fingerprint location as a histogram, and to use that prior in a Bayesian framework, to compute the probability of having a specific histogram of RSSI at a new location using Bayesian Networks [3] or the Naive Bayes algorithm [16]. Paschalidis et al. [15] use a Kullback-Leibler-based statistical framework for Wireless Sensor Networks localization (consisting in null hypothesis testing for each fingerprint). Bargh et al. [2] use the KL divergence to find the (single) nearest neighbor in the space of multinomial counts of Bluetooth dongles. Milioris et al. [11] also use KL divergence, this time on RSSI from WiFi data, but they assume that the RSSI from multiple APs is simply a multivariate Gaussian, a hypothesis that is not always true, as pointed out in Section I-A.

None of the previous methods considered probability kernels with distance-like metrics between distributions. We do, and show that such probabilistic kernels can be used for the regression of the location, achieving up to 1m accuracy in office environments.

II. Methods

Our method can be summarized as follows: we sample the distribution p of RSSI from all visible APs for a duration τ (typically of a few seconds), and we compare it to the distributions q in the fingerprint database, using the Kullback-Leibler divergence (Section II-A) and the KL-divergence kernel (Section II-B). In the database, each fingerprint is assciated with a location. The location is estimated through kernel regression (Section II-D). Our method naturally copes with unknown RSSI (Section II-C), contains few hyperparameters, and can be trivially extended to operate merely on histograms of AP connection (i.e., binary) instead of full RSSI levels (Section II-G). We justify sampling RSSI or AP during motion of the mobile device in Section II-E.

A. Kullback-Leibler Divergence

In information theory, the Kullback-Leibler divergence KL is a non-symmetric measure of the difference between two probability distributions[2] p and q. In the discrete case where the random variable S takes discrete values (e.g., integer-valued RSSI or SNR from an access point), we have: $KL(p||q) = \sum_s p(S = s) \log \left(p(S = s)/q(S = s)\right)$. To avoid taking logarithms of zero-valued bins, we smooth the distribution by adding a small constant term (e.g., 10^{-6}) and re-normalizing the empirical distribution function.

The symmetrized Kullback-Leibler divergence D between two distributions p and q can be simply defined[3] as

$$D(p, q) = KL(p||q) + KL(q||p). \tag{1}$$

Note that this metric does not satisfy the triangle inequality and cannot be considered a distance measure [5].

In the case when the discrete random vector $\mathbf{S} = \{S_1, \ldots, S_J\}$ is multivariate (e.g., when measuring RSSI from multiple access points $\{1, \ldots, J\}$), we can make the assumption of local independence of each AP's distribution[4], i.e., that $p(\mathbf{S}|\{x, y\}) = \prod_{j=1}^{J} p(S_j|\{x, y\})$ at specific location $\{x, y\}$. Such a local independence assumption for multiple APs was already made in [16]. Note that we now use the shorthands $p = p(\mathbf{S}|\{x, y\})$ to express the RSSI distribution obtained during tracking and around position $\{x, y\}$, and $q_\ell = q(\mathbf{S}|\{x_\ell, y_\ell\})$ to express the RSSI distribution at the fingerprint indexed by ℓ.

Using the chain rule for relative entropy, one can prove that the KL-divergence of a joint distribution of independent variables is equal to the sum of the KL-divergences for each variable's distribution [5]. We therefore have, for any two locations $\{x, y\}$ and $\{x_\ell, y_\ell\}$ and their associated multivariate distributions p and q_ℓ, and for J access points:

$$D\left(p, q_\ell\right) = \sum_{j=1}^{J} D\left(p(S_j|\{x, y\}), q(S_j|\{x_\ell, y_\ell\})\right) \tag{2}$$

B. KL-Divergence Kernel

The Kullback-Leibler divergence is used in [15] for localization in a statistical framework: the RSSI of the mobile is compared to several fingerprints through KL-based null hypothesis testing. We propose to combine the KL-divergence with kernel methods, as has already been done for other applications [12], and to use kernel-based regression algorithms.

Briefly, a *kernel function* $k(p, q)$ is a symmetric function equal to one if $p = q$ and decaying to zero as the dissimilarity of the two inputs increases. Kernel methods such as Support Vector Regression [13] often require the kernel matrix between all training datapoints to be Positive Semi-Definite (PSD)[5]. Following [12], and for a data-dependent range of values α, it

[2] We can also write $KL(p||q) = H(p, q) - H(p)$, where $H(p)$ is the entropy of p and $H(p, q)$ the cross-entropy due to using q instead of p.

[3] Our notation $D(p, q)$ for the symmetrized KL-divergence is not to be confused with the asymmetric KL-divergence $KL(p||q)$.

[4] We can indeed argue that the software most likely queries and receives answers from the APs independently, and that the fluctuations in signal propagation for various APs happen along somewhat different paths. There is no fundamental reason why we could not work with joint distributions, but the number of bins would grow exponentially with the number of APs, while the independence assumption helps us carry out the computations efficiently.

[5] A real-valued symmetric matrix $\mathbf{K} \in \mathcal{R}^{n \times n}$ is positive semi-definite if for all vectors $\mathbf{x} \in \mathcal{R}^n$, we have $\mathbf{x}^T \mathbf{K} \mathbf{x} \geq 0$.

is possible to define such PSD kernels by exponentiating the symmetrized KL-divergence:

$$k(p, q_\ell) = e^{-\alpha \sum_{j=1}^{J} D(p(S_j|\{x,y\}), q(S_j|\{x_l,y_l\}))} \tag{3}$$

Several other kernels have been suggested for multinomial probability distributions, such as the Bhattacharyya affinity kernel [8]. Such a Bhattacharyya kernel for multinomial distributions (e.g., counts of integer-valued RSSI from an AP) has a simple form. Its only limitation is its lack of obvious extension to joint distributions over several multinomials; we would however encounter that situation when sampling RSSIs from multiple APs. By contrast, we just showed that the Kullback-Leibler divergence can be easily computed for joint distributions of multiple independent multinomials.

C. Handling Missing Data

When the signal fingerprint at location $\{x, y\}$ does not sample any RSSI from a specific AP j, the obvious choice is to set that distribution to $p(S_j = -\infty|\{x, y\}) = 1$. We can approximate this by putting all the mass on the first bin of the histogram (typically the bin below the limit of detection).

When an AP is "unknown" both to the current sample p and to training fingerprint q_ℓ, then $D(p(S_j), q_\ell(S_j)) = 0$, i.e., we ignore the j-th AP in the kernel regression. However, if that AP is sampled by p and by a fingerprint q_ℓ but not by another fingerprint q_m, then the KL-divergence for that AP is smaller between p and q_ℓ than it is between p and q_m, giving more kernel weight to the fingerprint ℓ who "knows" that AP.

An alternative approach is to consider that when one distribution is defined but not the other, then the two distributions are infinitely different (i.e., their KL-divergence should be equal to infinity). Instead of using infinite values, we use a large constant that is equal to the maximum KL-divergence that can be obtained for that number of bins and for that smoothing coefficient, multiplied by a factor. In most cases, we use a factor of 1 (again, obtaining similar numerical results as by setting $p(S_j = -\infty|\{x, y\}) = 1$), and we use factors bigger than 1 (e.g., 4) only when the area covered by the fingerprints is very large, resulting in many APs not being "heard" in different parts of the map. Specifically, we used this approach with the large dataset of the public space in Section III-D.

Finally, when it appears that an AP is down and is never sampled, it can be simply removed from the sum in the kernel function exponent (Eq. 3).

D. KL-Divergence Kernel Regression

Using the KL-divergence kernel function k and a set of known training datapoints $\{q^{\{x_l,y_l\}}\}$, we perform Weighted Kernel Regression [14] to obtain an estimate of the location using p, the sampled distribution of RSSI:

$$(\bar{x}, \bar{y}) = \frac{\sum_\ell (x_\ell, y_\ell) k(p, q_\ell)}{\sum_\ell k(p, q_\ell)} \tag{4}$$

We propose to do this regression using only the K nearest neighbors (in the KL-divergence sense), instead of the full set of known training datapoints, i.e., to keep the K fingerprints $\{q_\ell\}$ that maximize $k(p, q_\ell)$. Our method amounts to nearest neighbor matching in the case when $K = 1$. Note that the choice of the K neighbors depends on the test datapoint p, and that the kernel function still needs to be evaluated for all known fingerprints. We optimize hyperparameters α and K on the training dataset (i.e., on the fingerprints), for instance using leave-one-out cross-validation.

Kernels provide a simple way to interpolate the location estimates between fingerprint locations; some earlier methods such as [4] were using more ad-hoc Delaunay triangulation of mean values of RSSI distributions. As an obvious extension, we could consider performing Support Vector Regression to restrict the number of support vectors (fingerprints).

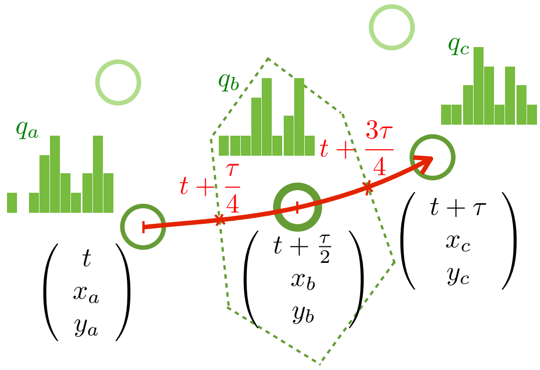

Fig. 2. Possible trajectory (red line) traversing three adjacent fingerprints located at $\{x_a, y_a\}$, $\{x_b, y_b\}$ and $\{x_c, y_c\}$ at times t, $t + \frac{\tau}{2}$ and $t + \tau$.

E. Evaluating the Distribution During Motion Tracking

In realistic scenarios, the distribution p for which one wishes to estimate the location is going to be sampled during motion, as the mobile moves through areas with different RSSI distributions. The crucial assumption that we make for estimating the location is that the PDFs continuously change for neighboring points[6]. In other words, for two close positions $\{x_a, y_a\}$ and $\{x_b, y_b\}$:

$$q(\mathbf{S}|\lambda\{x_a, y_a\} + (1-\lambda)\{x_b, y_b\}) \approx \lambda q_a + (1-\lambda) q_b \tag{5}$$

There is a trade-off between the number of RSSI samples necessary to get a good approximation of p (i.e., the time required τ and the distance travelled), and the error introduced by sampling from neighboring locations. The latter can be controlled by knowing how adjacent fingerprints are spaced, how frequently APs are queried, and having a prior idea on the speed of motion can however help. For instance, in some of our experiments, we used a time window with $\tau = 8s$, while the motion speed was 0.5m/s, adjacent training fingerprints were spaced every 2-2.5m, and APs were probed at 5Hz: this means that our sampling windows covered roughly 2 to 3 training fingerprints and up to 40 RSSI samples, as illustrated on Fig. 2.

[6]We plan on verifying that assumption quantitatively for specific datasets.

For comparison, each training fingerprint would have up to 130 samples. Let us assume from now on that τ is always adjusted to cover 3 fingerprints during tracking. We propose a weighting scheme that involves a smaller weight $\frac{\kappa}{2}$ for samples from q_a collected at the beginning $[t, t+\frac{\tau}{4})$ of the sampling window, and for samples from q_c at the end $[t+\frac{3\tau}{4}, t+\tau)$ of that window, and $1-\kappa$ for samples from q_b in the middle window $[t+\frac{\tau}{4}, t+\frac{3\tau}{4})$. κ can be determined by cross-validation using a multinomial sampler on the training dataset from three adjacent fingerprints for total duration τ, to be the value that minimizes the KL-divergence between the sampled $\frac{\kappa}{2}p\left(S|\{x_a, y_a\}\right)+(1-\kappa)p\left(S|\{x_b, y_b\}\right)+\frac{\kappa}{2}p\left(S|\{x_c, y_c\}\right)$ and the actual q_b. Note that our specific sampling window scheme gives an estimate for the location at $\frac{\tau}{2}=4s$ ago, which is acceptable for practical usage[7].

The impact of the choice of the sampling window τ, of the size of the RSSI bins, and of the number N of samples taken during fingerprinting are all investigated in Sections III-A3 and IV-B.

F. Unsupervised Recalibration of RSSI Histograms

The KL divergence can also be used as a metric to compare two global distributions. We propose to correct for some variations in the signal strength maps by shifting the test data RSSI histogram collected during tracking[8], so as to minimize its KL divergence with the distribution of RSSI from that same AP but for all training fingerprints.

G. Extension to Access Point Connection Histograms

Our KL-divergence kernel regression can be trivially extended to accommodate AP connection histograms (i.e., multinomials of the number of connections for each AP during time window τ). Even though we ignore the actual RSSI levels, we can thus achieve a median accuracy of 2 to 3m in an office environment, as shown in the next section.

One benefit of our approach is that it foregoes RSSI recalibration completely: what APs are seen might be similar across devices, even if the RSSI levels change. The only trick that we suggest is to remove, from all histograms, the APs that do not show up during tracking. Alternatively, we can know through software and at training time if the AP is ad-hoc or part of the infrastructure and use this information to filter out mobile phones acting as hot spots. Other ways of filtering out APs is to weed out devices with short ranges.

III. Results

In this section we report the localization results we obtained in various buildings featuring different training and tracking scenarios. Section III-A describes a series of experiments in an office building, where dense and repeated RSSI measurements were obtained during training. This high quality database allows for experimentation with different options and hyper-parameters in the tracking algorithm. Then we report results in progressively more difficult scenarios, including a trial in a long corridor (Section III-B) where single RSSI measurements were obtained while walking without stops, a trial involving a large, open space (Section III-C) and another trial in a public indoor environment with mixed layouts and heavy pedestrian traffic (Section III-D). These trials demonstrate the robustness of our algorithm under different settings.

TABLE I
Results on the 2D Office Dataset Using 4 APs

Technique	Median accuracy	Accuracy at 90%
Kalman filter [6]	2.0m	-
Voronoi particle filter [6]	1.6m	-
Model-free tracking [4]	1.3m	2.5m
KL divergence, 1 NN	1.25m	3.18m
KL divergence, 3 NN WKR	**1.06m**	**2.34m**
KL divergence, 88 NN WKR	**1.14m**	**1.98m**

A. 2D Office Space with Dense Fingerprinting

For our first set of experiments, we used a 2D office dataset investigated in [4], [6], consisting of a 40m×40m area, depicted in Fig. 3. The training data consisted of 88 fingerprints recorded for 22 APs[9]; some APs had 130 samples for each location. 4 APs only were used in the published experiments. Tracking data in that dataset were acquired a few days later,

Given the high density and large number of repeated RSSI samples, we were able to evaluate the results with different options of our tracking algorithm, and compare them to the results reported by previous researchers using the same data.

1) Localization Based on RSSI Distributions: Using leave-out-last cross-validation on the training data, we selected the optimal coefficient α in the KL-divergence kernel function (Eq. 3) and the optimal number of nearest neighbor fingerprints K for kernel regression, both when using 4 APs and when using 22 APs. We also selected the optimal α when using all fingerprints for regression for both numbers of APs.

Tracking data were re-calibrated as explained in Section II-F. As we report in Table I, we achieved a median accuracy of 1.06m, when using the optimal number of nearest neighbors ($K=3$) for kernel regression. This result is considerably better than previously published Kalman filters (2m) and Voronoi particle filters (1.6m) [6] or model-free tracking (1.3m) [4]. As we show on Fig. 3, the estimated trajectory is reasonably smooth. Interestingly, using the location of only one nearest neighbor (based on the KL-divergence) still yields good tracking performance at 1.25m.

We observed a further decrease in the median tracking error when using 22 APs rather than 4 APs, contrary to what was suggested in [9]; as shown in Table II, the 90% quantile error was reduced to around 1.7m-1.9m from 2m-2.3m, and the

[7]We could consider different weightings κ, perhaps with a continuous κ-parameterized curve (exponential smoothing).

[8]We are limiting ourselves to changes in RSSI or SNR comes from variations in the ambient noise. Our method does not consider local changes in the environment, such as furniture or people movements.

[9]It is common to observe hundreds of unique MAC addresses in office environments, coming from various floors and individual offices.

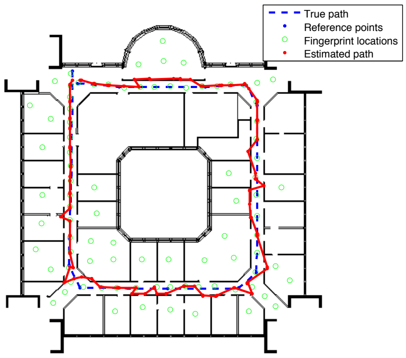

Fig. 3. Tracking results on the 2D office dataset using 4 APs. The true path is in dashed blue, the estimated path in solid red line, and the 88 fingerprint locations appear as green circles. We used a KL-divergence kernel with weight $\alpha = 0.024$, $\tau = 8s$ windows and performed kernel regression on the $N = 3$ nearest neighbors. Median error was 1.06m and 2.34m at the 90% percentile. RSSI was sampled at 5Hz, yielding up to 40 samples per sampling window for each AP.

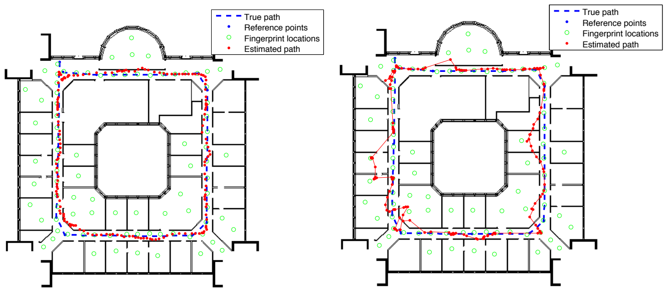

Fig. 4. Tracking results on the 2D office dataset using 22 APs. The left figure shows the results obtained using a KL-divergence kernel with weight $\alpha = 0.041$, $\tau = 8s$ windows and kernel regression on the $N = 88$ fingerprints, with a very low median localization error of 0.86m and 1.72m at the 90% percentile. The right figure shows the results obtained using a KL-divergence kernel only on the AP connection counts (ignoring the RSSI values); $\alpha = 77$, $\tau = 8s$ windows and kernel regression on the $N = 88$ fingerprints. Despite not using the signal values, we obtained a decent localization error of 1.94m and 4.95m at the 90% percentile.

median error was slightly reduced to 0.95m from 1m after including all the available 22 access points. Note that those 18 additional APs were part of the ambient RF "noise"; unlike the 4 APs that were specifically set up for the experiment, those APs may have been placed in different parts of the building, on different floors, or in individual offices.

2) Localization Based on Access Point Visibility: In a second series of experiments on the same office dataset, we ignored the RSSI from the APs, and used only multinomials of AP connections to build the KL-divergence kernels. As shown in Table II and on Fig. 4, the tracking accuracy remained decent, at about 2m median error.

3) Effects of Fingerprinting and Tracking Hyper-Parameters: Four different questions pertaining to parameters related to fingerprints and to tracking might be of interest to the users of our method:

- How many fingerprinting locations should be chosen?
- How many RSSI samples N should be measured to estimate the location-specific fingerprint distributions $q_\ell(\mathbf{S})$?

TABLE II
RESULTS ON THE 2D OFFICE DATASET USING THE KL-DIVERGENCE KERNEL ON 22 APS, WITH OR WITHOUT RSSI

Technique	Median accuracy	Accuracy at 90%
With RSSI, 1 NN	1.16m	2.84m
With RSSI, 6 NN WKR	**0.96m**	**1.88m**
With RSSI, 88 NN WKR	**0.86m**	**1.72m**
No RSSI, 1 NN	1.94m	4.95m
No RSSI, 27 NN WKR	1.90m	4.31m
No RSSI, 88 NN WKR	1.90m	4.40m

- During tracking, how many RSSI samples n should be used in the localization algorithm, assuming that the sampling frequency f is given and that the motion speed cannot be controlled? In other words, how long should be the sampling window $\tau = n/f$?
- How wide should be the histogram bins used to encode the RSSI distributions?

We quantified the effects of each of these four hyper-parameters in terms of tracking accuracy with the 2D office data. In particular, we measured the impact of:

- Reducing the number of fingerprinting locations by subsampling them in space (see Fig. 5), using inter-fingerprint subsampling factors of 2, 3, 4, 5, 7 and 14.
- Reducing the number of RSSI samples used to estimate the distributions at each location, using subsampling factors of 2, 3, 6 and 10. There were at most $N = 130$ samples per fingerprint location per AP; at a subsampling factor of 10, that number was reduced to at most $N = 13$.
- Changing the sampling window length τ during tracking, taking values of 1s, 2s, 4s, 8s and 12s. Given that the motion speed was about 0.5m/s and the sampling frequency $f = 5$Hz, this sampling window corresponded to $n = 5, 10, 20, 40, 60$ samples and approximately 0.5m, 1m, 2m, 4m and 6m, respectively.
- Changing the histogram bin size from 1dB (finest unit, since the RSSI are recorded as integers), to 2dB, 5dB and 10dB. RSSI values ranged from -100dB to -30dB.

Fig. 6 shows the tracking errors (at 10%, 50% and 90% quantile) for each variable, using the optimal result selected among all combinations of the remaining 3 variables. An immediate conclusion is that the more samples N per fingerprint, the more fingerprint locations, the longer the sampling window during tracking and the finer the histogram bins for fingerprints, the better the tracking accuracy, although the hyperparameters chosen for this dataset appear to have reached a plateau.

From the inspection of the full results, it appears that the 2D office data could be "downgraded" a little in terms of fewer fingerprinting samples N, fewer fingerprint locations, shorter tracking sampling windows than 4s and coarser fingerprint histogram bins, without much detriment to the tracking accuracy. For instance, while the optimal tracking accuracy is 0.83m (median) and 1.65m (at 90%) for up to $N = 130$ RSSI samples per fingerprint, 88 fingerprints, 1dB bins and

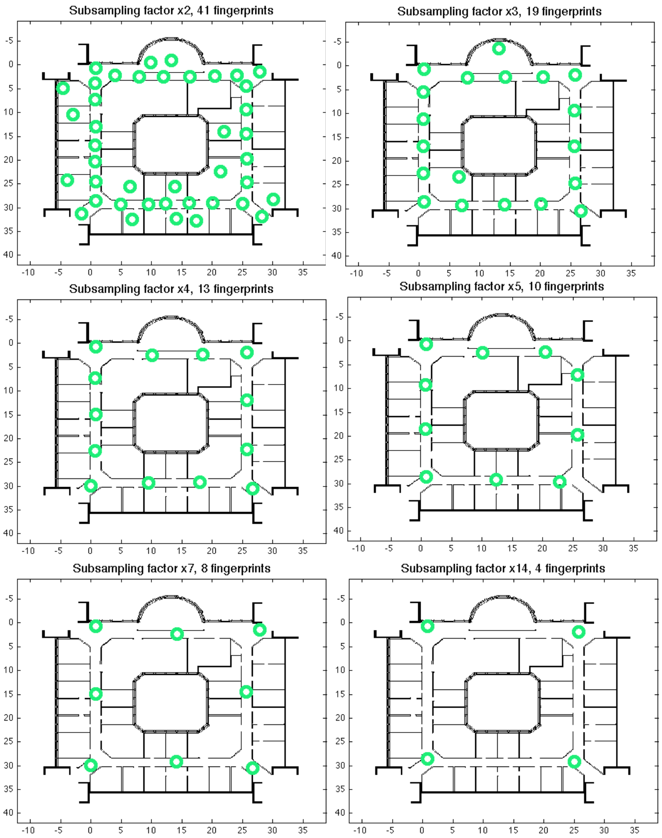

Fig. 5. Subsampling the fingerprint locations, (i.e., reducing the number of fingerprints) on 2D office dataset, for 6 different inter-fingerprint space subsampling factors (2, 3, 4, 5, 7 and 14).

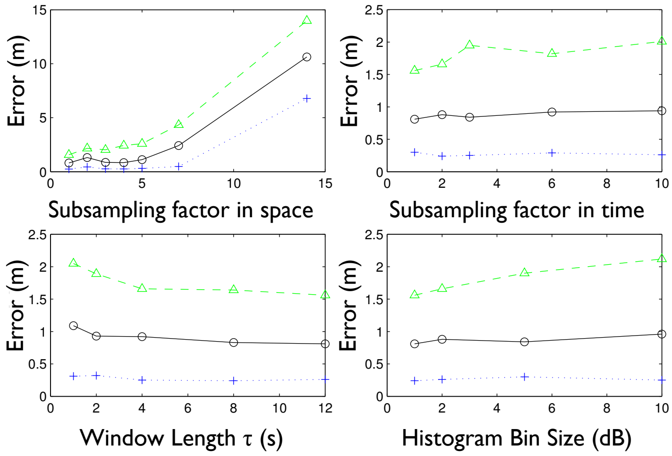

Fig. 6. Hyperparameter analysis on the 2D office dataset using 22 APs. We investigated a) reducing the number of fingerprints (i.e., subsampling in space, see Fig. 5 and see top left graph), b) reducing the number of samples N used to build the fingerprint distributions (i.e., subsampling the RSSI measurements in time, the top-right graph), c) varying the number of samples n recorded during tracking while in motion (i.e., changing the sampling window length τ, bottom-left graph) and d) increasing the bin size in the fingerprint histograms (bottom-right graph). Black solid lines represent median errors, green dashed lines errors at 90% and blue dotted lines errors at 10%.

8s tracking windows, we can still obtain 1.28m (median) and 2.59m (at 90%) tracking error using only 19 fingerprints, up to $N = 22$ fingerprint samples, 5dB histogram bins and 1s-long tracking windows. Another observation is that the worst cases are when the number of fingerprints, the number of fingerprinting samples N and the number of tracking samples n are all low and the histogram bins are narrow. Finally, when there are few fingerprint locations, then coarse histogram bins and long tracking sampling windows can help reduce the error rate.

In summary, these results suggest that the spatial density of the fingerprints is the most important performance impacting factor. In comparison, repeated measurements at each location are less important – the advantage of multiple measurements at the same location flattens beyond about 20 samples. Optimal bin sizes for the histograms vary with the length of the sampling window during tracking – more refined bins are useful only with longer tracking windows that accumulate more samples to estimate the RSSI distribution.

B. *Fingerprinting “On The Fly” While Walking*

A less favorable training scenario is when fingerprinting is done “on the fly” while walking. This allows for dense spatial coverage if the RSSI queries can be made sufficiently frequent, if the walk is slow, and if the trajectory covers the space evenly. However, only one sample is acquired for each location. The lack of repeated measurements means that the RSSI distribution at each location cannot be reliably estimated unless multiple measurements are pooled from neighboring locations. But the spread of the pooled locations introduces more variability in the RSSI values. Localization using only AP visibility provides a more robust option. A simple method would use the binary vector of AP visibility at each location as fingerprints, whereas during tracking, position would be determined by nearest-neighbor matching to those binary vectors. To apply KL-divergence regression to this data, we pooled AP visibility vectors from consecutive locations covered by the walk over a small temporal window, and used those to estimate a distribution of AP connections for the location at the center of the window.

We tested this scenario in a walk at constant speed (around 1.4m/s) along a corridor that is about 2m wide but extends to 260m in length. NetStumbler would query APs only at 1Hz. We used 8s-long sampling windows to create 55 fingerprints (that are AP connection histograms) spaced every 4s (i.e., every 5.5m) for the 130 APs discovered “on the fly”. We completely ignored RSSI values and just recorded multinomial histograms of AP visibility. When we used those fingerprints to localize ourselves later on the same day (while in motion at 1.4m/s), we achieved 3.3m median accuracy (9m at 90%), which compares with 5.2m median accuracy (15m at 90%) for 3-NN on 1s-long binary vector fingerprints. Keeping the same AP fingerprints, we repeated the tracking test one week later: we still achieved a 4m median accuracy (7.6m at 90%), in spite of some APs that had disappeared in the meantime. We illustrated those tracking results on Fig. 7. These results are upper bounds: more careful (slower) fingerprinting and

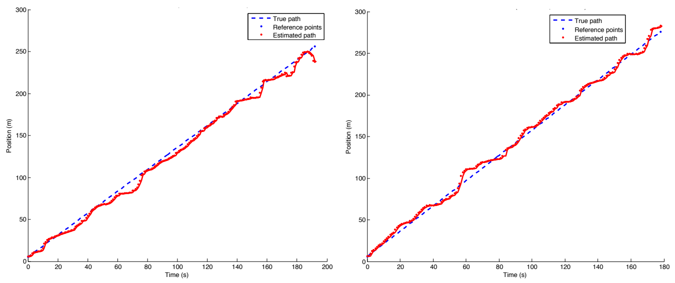

Fig. 7. Tracking results along a 260m-long corridor using 55 fingerprints acquired "on the fly" and 130 APs captured while walking at 1.4m/s, not using RSSI values but only statistics on their presence/absence. The left figure shows the results obtained the same day (median accuracy of 3.3m and 9m at 90%) and the right figure results obtained one week later (median accuracy of 4m and 7.6m at 90%).

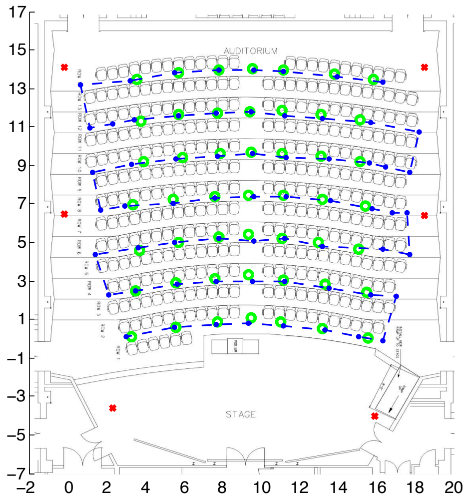

Fig. 8. Floorplan of the open-space environment (auditorium), with 49 fingerprint locations (green circles) and 6 access points (red crosses). Next day's tracking path is indicated in the dashed blue line.

accounting for speed fluctuations should bring the errors down.

C. Open-Space Localization

It can be argued that a narrow and long corridor is an ideal layout for localization. In this section we compare the results of two trials, one made in a large, open indoor space (an auditorium with over 200 seats), and another one in a narrow hallway, using the same equipment and training strategy. In both trials, fingerprints were collected at locations spread evenly over the space, and repeated measurements were made at each location.

1) Stop-and-Go Fingerprinting in an Auditorium: During training in the auditorium experiment, RSSI values from 6 APs were recorded at 49 fingerprint locations using NetStumbler at the frequency of 1Hz. We recorded tracking RSSI on a path going through all the fingerprints on the next day, moving slowly at 0.17m/s. Fig. 8 shows the locations of the fingerprints and the path for tracking. Best results obtained are 4.65m median accuracy (10.23m at 90%), using $\tau = 10$s long tracking windows, $K = 3$ nearest neighbor WKR and a kernel

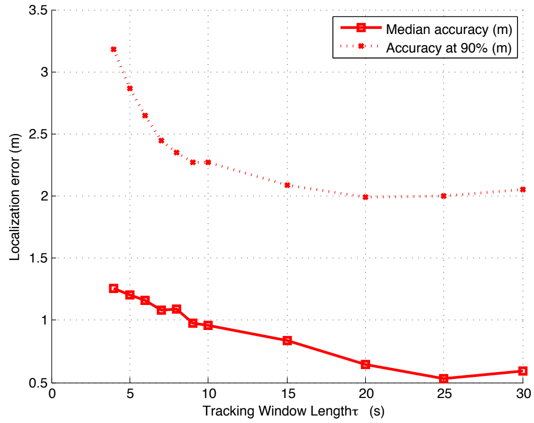

Fig. 9. Performance of KL-divergence kernel regression on the 15-long corridor (median error and error at 90%).

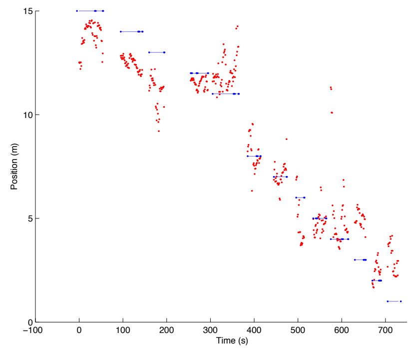

Fig. 10. Predictions versus true locations in the 15m-long corridor.

coefficient $\alpha = 0.51$.

To determine whether the localization results are sufficiently informative, we compared those with two random localization schemes, one consisting in randomly permuting the tracking locations every 2s, another in generated random walks within the tracking area and at the same speed, and their median accuracy would be between 8m and 9m, and the error at 90% would be 14 to 15m. We conclude that our algorithm does somewhat work in open space areas, though with an accuracy that is lower than in corridors.

2) Stop-and-Go Fingerprinting in a Corridor: The results in the auditorium can be compared to a control experiment using the same equipment in a 18m-long corridor. In that control experiment, we set up 4 APs (taken among the 6 APs used in the auditorium), and defined 15 fingerprint locations spaced by 1m, and fingerprinted each location for 120s. We then tracked twice using RSSI by moving between the fingerprint positions and staying there for 120s, once in the forward direction, then backwards. Our KL-divergence-based localization algorithm used $K = 3$ nearest-neighbor regression, kernel coefficients respectively equal to $\alpha = 0.12$ and $\alpha = 0.22$, with tracking sampling windows of length $\tau = 4$s through $\tau = 30$s. Fig. 9 and 10 show that the median tracking error can go well below 1m (median) and reach 2m (at 90%) in a corridor environment.

TABLE III
ACCURACY RESULTS IN A COMPLEX PUBLIC SPACE.

Technique	Sampler	Floor	Median	At 90%	Floor
With RSSI	NetStumbler	lower	8.2m	16.9m	96.2%
		upper	9m	17.1m	83.8%
No RSSI	WiFi Scanner	lower	10.3,	24.3m	89%
		upper	9.1m	17m	92.6%

D. Localization with Sparse Fingerprinting in a Complex Public Space

The last experiment we report involves a realistic, almost worst-case scenario, where the building layout includes both corridors and open spaces on two floors, and there is continuous pedestrian traffic throughout the space during both fingerprinting and tracking. Fingerprints were collected at 162 locations covering both floors, and the locations were 5.5m apart from each other on average. Location errors of the order of 5m were frequent. 10-15 repeated measurements were obtained at each location. During tracking, samples were pooled over a window of 10s.

We experimented with two options of the algorithm: using RSSI (on a PC running NetStumbler) and AP visibility only (on a Mac running WiFi Scanner). The results are detailed in Table III. It can be seen that the experimental conditions in this scenario are stretching the limits of the algorithm. They also represent opportunities for further improvements by carefully designed sampling strategies and dense, repeated data collection.

IV. DISCUSSION

The results reported in the previous section show a wide range of performance under different experimental scenarios. For real-world deployment, proper expectations need to be set in consideration of the difficulty of the particular scenario. Also, sampling strategies need to be designed to adapt to such difficulties.

In this section, we present an analysis on characterizing the difficulty of an experimental scenario given a collection of fingerprints. We begin by revisiting the definition of fingerprints and the assumptions that we employed throughout the paper (Section IV-A). We then examine the separability of unique locations in the collection, according to the similarities between RSSI distributions associated with each location. Then we investigate theoretical bounds on the localization accuracy as a function of the number n of samples acquired during tracking (Section IV-B).

A. A Probabilistic Definition of Fingerprints

Suppose we have a finite set of locations (a *location* being either a point or a "small" area) and a set of possible discrete measurement values (scalar or vector) from some finite set. We then have the following definition:

Definition 1. *Given a finite set of locations $\mathcal{L}$ and a finite and discrete measurement set $\mathcal{S}$, a fingerprint is defined as a set of probability distributions specific to a location indexed by ℓ:*

$$p\left(S|\{x_\ell, y_\ell\}\right), \quad S \in \mathcal{S},\ \forall\{x_\ell, y_\ell\} \in \mathcal{L}$$

For ease of notation, we write $p_\ell(S) = p\left(S|\{x_\ell, y_\ell\}\right)$. Fingerprints determine the probability outcomes of measurements, in that if $S_1, \ldots, S_n$ are measurements at an arbitrary sequence of locations $\{x_1, y_1\}, \ldots, \{x_n, y_n\}$, then

$$p\left(S_1 \ldots, S_n|\{x_1, y_1\}, \ldots, \{x_n, y_n\}\right) = \prod_{\ell=1}^{n} p_\ell\left(S_\ell\right). \quad (6)$$

In this paper, we considered two examples of measurements taken from WiFi enabled devices which can communicate with so called access points (AP). In the first example, RSSI measurements are used, and there are J access points and $\mathbf{S} \in \mathcal{S} \doteq \{s_L, \ldots, s_H\}^J$, where s_L and s_H are the lowest and highest RSSI values, respectively, that can be recorded by the WiFi device and software. In the second example, we supposed that it is possible to determine whether or not an access point is in or out of range, as above, but with $\mathcal{S} \doteq \{0, 1\}^J$.

Both examples necessitate evaluation of the probabilities, either of measuring a specific RSSI value s for a specific AP, or of reporting an access point as being in range. Intuitively, those probabilities will become more precise if we increase the number of samples N at fingerprinting time (the samples which are used to estimate the distributions $p_\ell(S)$), and we confirm the impact of N on the tracking accuracy in Section III-A3.

Let us now make refinements of the above definition.

1) Device Independent Measurements: First, we suppose that location outcomes are device independent. In practice, we observed that different WiFi cards on different laptops recorded different sets of RSSI values at identical locations. We nevertheless assume that an appropriate rescaling can be applied to the distribution of measurements from an RF device relative to another one, in order to compensate for manufacturing differences between the two RF measuring devices.

2) Motion and Conditional Independence Given the Location: Second, the above definition states that the measurements are conditionally independent given the location, which implies that the fingerprints determine the measurement statistics given the sequence of locations at which they were recorded. If the location does not change for an interval of time, then the measurements are theoretically i.i.d. (independently and identically distributed), therefore interchangeable, provided that no other phenomena occur that might disturb the radio-frequency field, such as people passing by or electrical equipment being turned on or off.

While it is easy to enforce immobility during fingerprinting (i.e., when building the database of fingerprints), this becomes impractical during tracking, and consecutive measurements might be acquired at slightly different locations. Recall that our algorithm needs to acquire n samples within a time window of duration τ to estimate the location at the middle of that window. We can nevertheless assume that the scale at which RF values change is of the same order as the distance covered by the tracked person or object during tracking time τ, and we explained in Section II-E a way to smooth over adjacent locations. Our assumptions imply that the probability of location error goes to 0 with increasing numbers of tracking measurements n "around" a location; it has been experimentally proved, during motion, in Section III-A3, and we give in Section IV-B a theoretical bound in the case of fixed location.

3) Conditional Independence of Access Points: Note that the fingerprint definition and the conditional independence given location ℓ does not necessarily imply, in the case of vector measurements $\mathbf{S}_\ell = \{S_{\ell,1}, \ldots, S_{\ell,J}\}$ and of a set $\mathcal{S}$ of J-dimensional vectors, that the following assumption holds:

$$p_\ell\left(\mathbf{S}_\ell\right) = \prod_{j=1}^{J} p_\ell\left(S_{\ell,j}\right) \tag{7}$$

While we do not need that assumption for the Chernoff bound on localization error (Section IV-B), this becomes practical for the evaluation of the Kullback-Leibler divergence between fingerprints of J access points (Section II-A). We can indeed argue that the WiFi software most likely queries and receives answers from the APs independently, and that the fluctuations in signal propagation for various APs happen along somewhat different paths. There is no fundamental reason why we could not work with joint distributions, but the number of bins would grow exponentially with the number of APs, while the independence assumption helps us doing the computation efficiently.

4) Time-Invariance of Fingerprints: The time-invariance assumption that information contained in fingerprints remains valid at tracking time is arguably the weakest hypothesis in this field of research, since there are a number of perturbations affecting the fingerprints that are outside of our control and which would be difficult to model. We will only contend that we can correct for shifts in signal level (Section II-F), ignore new APs and not use removed APs in the localization algorithm, but that we cannot easily deal with local changes in the environment, such as furniture or people movements.

B. Theoretical Bounds on Localization Error and Number of Tracking Samples n

The following sub-section tries to provide theoretical insight about the number of samples n that our probability-based localization method needs during tracking. For a specific RSSI sampling frequency f (e.g., 5Hz on one set of experiments, 1Hz on another), it specifically answers the question about the choice of the sampling duration $\tau = f \times n$.

As a key disclaimer, we are assuming here that the distributions of RSSI at the fingerprints are perfectly known[10], that there is no drift in RSSI between fingerprinting and tracking, and that there is no motion at all: we are simply hypothesizing that we are exactly at one of the fingerprints' locations and that we observe n measurements of RSSI coming from J access points.

Denote by $\{x_\ell, y_\ell\}$ the actual location and $\{\bar{x}, \bar{y}\}$ the estimated location of the user, and $\bar{\mathbf{S}} = \{\mathbf{S}_1, \ldots, \mathbf{S}_n\}$ a sequence of n J-variate measurements from J access points. We can compute a bound on the pairwise error probability of "preferring" one location over another, using maximum likelihood, as follows. The Chernoff bound on preferring m to ℓ when the user is actually at location ℓ, i.e., that $p_m(\hat{\mathbf{S}}) > p_\ell(\hat{\mathbf{S}})$, can be written as:

$$p\left(\frac{p_m(\hat{\mathbf{S}})}{p_\ell(\hat{\mathbf{S}})} > 1\right) \leq p_{\ell,m} \tag{8}$$

$$p_{\ell,m} \doteq \inf_{t\in[0,1]} \sum_{\hat{\mathbf{S}}} p_m^t(\hat{\mathbf{S}}) p_\ell^{1-t}(\hat{\mathbf{S}}) \tag{9}$$

The bound in Eq. 9 is obtained from the moment generating function for the log-likelihood as, for all $t \geq 0$,

$$\begin{aligned} p\left(\log\frac{p_m(\hat{\mathbf{S}})}{p_\ell(\hat{\mathbf{S}})} > 0\right) &\leq E\left(e^{t\log\frac{p_m(\hat{\mathbf{S}})}{p_\ell(\hat{\mathbf{S}})}}\right) \\ &= \sum_{\hat{\mathbf{S}}} p_m^t(\hat{\mathbf{S}}) p_\ell^{1-t}(\hat{\mathbf{S}}) \end{aligned}$$

using Markov's inequality (an alternative derivation could be provided in [7]). It can be shown that the optimum choice of t is within interval $(0,1)$, so nothing is lost by the above constraint[11].

Using Eq. 9, let us provide a simple bound on the overall probability that some location $\{\bar{x}, \bar{y}\}$ other than the true location $\{x_\ell, y_\ell\}$ is selected, supposing the locations are selected uniformly at random. Eq. 10 essentially estimates the probability that the "wrong" fingerprint (not ℓ) is chosen, if we were sampling n samples from that fingerprint ℓ and its associated distribution $p_\ell(\mathbf{S})$. This definition ignores how bad the confusion is (whether we choose an adjacent fingerprint, or a fingerprint that is dozens of meters away).

$$p\left(\{\bar{x}, \bar{y}\} \neq \{x_\ell, y_\ell\} | \{x_\ell, y_\ell\}\right) \leq \sum_{m\neq\ell} p_{\ell,m} \tag{10}$$

The construction of the above bound assumes an ideal knowledge of the fingerprints. In reality, those fingerprints depend on the number of measurements N used to obtain the multinomial histograms of $\mathbf{S}$. If the fingerprint distributions are i.i.d. as supposed, then the probability of pairwise error goes to 0 with the number of tracking measurements n. Of course with some fingerprint methods these assumptions may not hold.

Keeping in mind that our Chernoff bounds are designed for an optimal strategy as well as supposing that the fingerprints are perfectly known and that there is no fingerprint drift or problems with calibration, which would all lead to poorer

[10]In this section, our definition of fingerprint here is that we model (for each access point) RSSI distributions as multinomials with 1dB-wide bins.

[11]More specifically, the bound in Eq. 9 is convex in t and decreasing at 0. Since the bound is an increasing function at $t = 1$ the minimum is in $(0, 1)$.

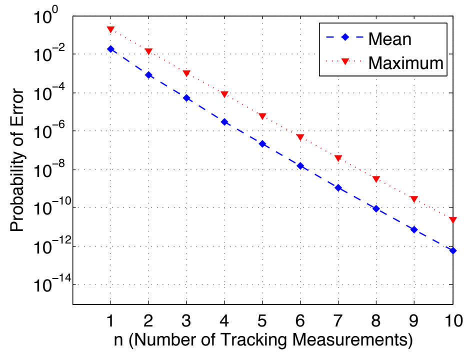

Fig. 11. Chernoff bound-based probability of choosing the wrong fingerprint location (among the 88 fingerprints of the office dataset) if we sampled n measures during "tracking". Fingerprint drift and calibration are ignored and simulated "tracking" data are used (i.e., sampled from the true $q_\ell(\mathbf{S})$).

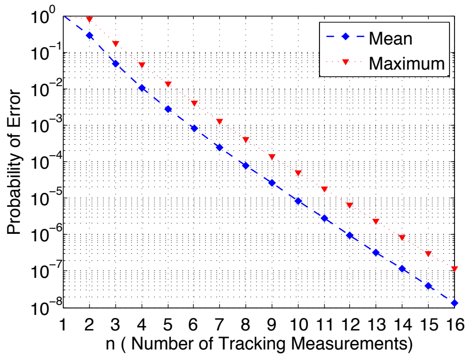

Fig. 12. Chernoff bound-based probability of choosing the wrong fingerprint location (among the 49 fingerprints of the auditorium dataset) if we sampled n measures during "tracking". As in Fig. 11, fingerprint drift and calibration are ignored and simulated "tracking" data are used.

performance, we provide in Fig. 11 and in Fig. 12 two graphs for the mean and maximum probability of error (i.e., of choosing the wrong fingerprint), respectively for the 88-fingerprint and 22-access point office dataset from Section III-A, and for the 49-fingerprint and 6-access point open-space (auditorium) dataset from Section III-C. As we explained earlier, the bound on the probability of predicting the wrong fingerprint location decreases exponentially with the number of samples n at tracking (i.e., with the tracking window length τ). There clearly is a difference between the 2D office dataset (with few open spaces and mostly corridors), where $n = 2$ already gives a theoretical misclassification probability of at most 1%, and the open-space auditorium dataset, where we need $n = 5$ measurements to achieve the same theoretical performance. The latter results also indicate that a single measurement of RSSI will be clearly insufficient for localization, which makes a strong case for probabilistic or smoothing-based methods.

In summary, the pairwise Chernoff bounds which give rive to these results might be taken as indicative of the "quality" (i.e., the separability or dissimiliarity) of the fingerprint map.

V. Conclusions

We designed a simple probabilistic algorithm for WLAN fingerprint-based tracking, relying on location regression with KL-divergence kernels. Our time-window based sampling approach is a very simple way to account both for the motion and for the complex non-Gaussian distributions of RSSI. Moreover, the structure of our model is such that we can further investigate the distributions of location prediction error and to quantify the localization uncertainty due to how the WiFi signal distribution varies in space.

Acknowledgment

The authors would like to thank Benjamin Lowe and Detlef Hartmann for their assistance in data collection and helpful feedback.

References

[1] P. Bahl and V.N. Padmanabhan, "An In-Building RF-based User Location and Tracking System", *IEEE Infocom*, vol.2, pp.775–784, 2000.

[2] M.S. Bargh and R. de Groote, "Indoor Localization Based on Response Rate of Bluetooth Inquiries", *International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments*, pp.48–54, 2008.

[3] P. Castro, P. Chiu, T. Kremenek and R. Muntz, "A Probabilistic Location Service for Wireless Network Environments", *Ubiquitous Computing*, September 2001.

[4] A. Chen, C. Harko, D. Lambert and P. Whiting, "An algorithm for fast, model-free tracking indoors", *ACM SIGMOBILE Mobile Computing and Communications Review*, vol.11, n.3, 2007.

[5] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, 2nd edition, Wiley Interscience, Hoboken, NJ, 2006.

[6] F. Evennou, F. Marx and E. Novakov, "Map-aided indoor mobile positioning system using particle filter", *IEEE Conference on Wireless Communications and Networking*, pp.2490–2494, 2005.

[7] R.G. Gallger, *Information Theory and Reliable Communications*, Wiley, 1968.

[8] T. Jebara, R. Kondor and A. Howard, "Probability Product Kernels", *Journal of Machine Learning Research*, vol.5, pp.819–844, 2004.

[9] L. Koski, T. Perala and R. Piche, "Indoor Positioning Using WLAN Coverage Area Estimates", *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2010.

[10] S. Kullback and R.A. Leibler, "On Information and Sufficiency", *Annals of Mathematical Statistics*, vol.22, n.1, pp.79–86, 1951.

[11] D. Milioris, L. Kriara, A. Papakonstantinou and G. Tzagkarakis, "Empirical Evaluation of Signal-Strength Fingerprint Positioning in Wireless LANs", *ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, 2010.

[12] P.J. Moreno, P.P. Ho and N. Vasconcelos, "A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications", *Neural Information Processing Systems*, 2002.

[13] A.J. Smola and B. Scholkopf, "A tutorial on support vector regression", *Statistics and Computing*, vol.14, n.3, pp.199–222, 2004.

[14] E. Nadaraya, "On estimating regression", *Theory of Probability and Applications*, vol.9, pp.141–142, 1964.

[15] I.C. Paschalidis, K. Li and D. Guo, "Model-Free Probabilistic Localization of Wireless Sensor Network Nodes in Indoor Environments", *Second International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments*, pp.68-78, 2009.

[16] T. Roos, P. Myllymaki, H. Tirri, P. Misikangas and J. Sievanen, "A Probabilistic Approach to WLAN User Location Estimation", *International Journal of Wireless Information Networks*, vol.7, n.3, 2002.

[17] T. Vauper, J. Seitz, F. Kiefer, S. Haimerl and J. Thielecke, "Wi-Fi Positioning: System Considerations and Device Calibration", *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2010.