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 needs recalibration; 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. We propose in this article a simple methodology that takes into account the full distribution in computing similarities with 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 an unsupervised KL-divergence based recalibration of the training fingerprints. Finally, we apply our method to work with histograms of WiFi connections with 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 1m accuracy in office environments, and we show that our method generalizes to non-Gaussian RSSI distributions.**

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

I. Introduction

Tracking people and objects indoors from radio signal strength measurements can be performed with an accuracy of a few meters for 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. 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 dynamical system inference, such as in Kalman or particle filtering [6].

Those fingerprint maps however generally store only the mean value of the Received Signal Strength Indicators (RSSI) [4], [6] and do not exploit information about the fluctuations of the RSSI in the environment. And yet, we noticed that in practice, certain devices record more complex distributions, complicating the fingerprinting process and introducing errors at estimation. Moreover, frequent re-training 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.

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

The common assumption about the RSSI coming from multiple APs is that the signals are distributed as multivariate Gaussians. It has however been reported [14] 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 10dB, as illustrated on Fig. 1.

Presumably, if we use mean and variance methods with a multimodal distribution, then we are less discriminating than we could otherwise be. 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, and in order to compare such multimodal distributions, we propose to use the Kullback-Leibler (KL) divergence.

B. Proposed Improvements

We propose in this research a probability kernel-based approach to matching position-labeled fingerprints. We compare distributions using the symmetrized Kullback-Leibler divergence and construct probability kernels that can be used either in a simple weighted regression scheme. We contend that this metric on fingerprints is robust to various noise and RSSI distributions, and we provide means to estimate the location using a short-term history of RSSI measurements. As an extension, we also propose an alternative approach

[1]We are planning on evaluating the trade-off between coarser binning schemes, e.g. 5dB bins, and time window lengths.

to fingerprinting, that records only the count of successful connections to Access Points (AP) over a small time interval, a method similar in principle to AP coverage area estimates [7].

C. Prior Art in Probability-Based Indoor Localization

The first usage of a probabilistic approach to RSSI in indoor localization was explained in [3], [13]. 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 [13]. [12] use a Kullback-Leibler-based statistical framework for Wireless Sensor Networks localization (consisting in null hypothesis testing at each fingerprint). [2] use the KL divergence to find the one nearest neighbor in the space of multinomial counts of Bluetooth dongles. [9] also use KL divergence, this time on RSSI from WiFi data, but they assume that the RSSI from multiple AP is simply a multivariate Gaussian, a hypothesis that is not always true, as we showed 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 1m accuracies in office environments.

II. METHODS

Our method can be summarized as following: 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 labeled fingerprint database, using the Kullback-Leibler divergence (Section II-A) and the KL-divergence kernel (Section II-B). The location is estimated through kernel regression (Section II-D). Our method naturally copes with unknown RSSI (Section II-C), contains few hyper-parameters, and can be trivially extended to operate merely on AP connection histograms instead of full RSSI (Section II-G). We justify sampling RSSI or AP during motion 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 $D(p,q)$ in Eq. (1). However, this metric does not satisfy the triangle inequality and cannot be considered a distance measure.

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

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 marginal distribution[4], i.e. that $p(\mathbf{S}|\{x,y\}) = \prod_{j=1}^{J} p(S_j|\{x,y\})$ at specific location $\{x,y\}$. Note that we now use the shorthands $p = p(\mathbf{S}|\{x,y\})$ and $q^{\{x,y\}} = q(\mathbf{S}|\{x,y\})$. Such a local independence assumption for multiple APs was already made in [13].

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 marginal distribution [5]. We therefore have, for any two locations $\{x,y\}$ and $\{x',y'\}$ and their associated multivariate distributions p and q:

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

B. KL-Divergence Kernel

The Kullback-Leibler divergence is used in [12] 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 [10], 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 often require the kernel matrix between all training datapoints to be positive semi-definite. Following [10], and for a data-dependent range of values α, it is possible to define such PSD kernels by exponentiating the symmetrized KL-divergence:

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

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 $q^{\{x,y\}}$, then $D(p(S_j), q(S_j|\{x,y\})) = 0$, 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 but not by another fingerprint q', then the KL-divergence for that AP is smaller between p and q than it is between p and q', giving more kernel weight to the fingerprint who "knows" that AP.

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 $\left\{q^{\{x_l,y_l\}}\right\}$, we perform Weighted

[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 doing the computation efficiently.

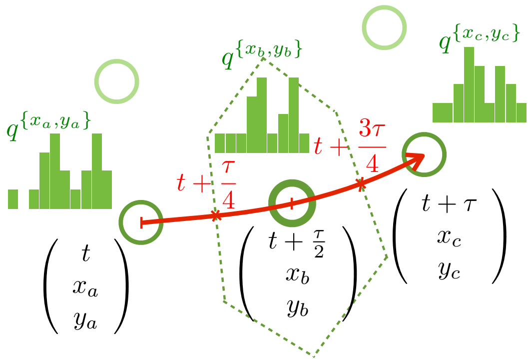

Fig. 2. Possible trajectory (red line) traversing three adjacent fingerprints located at $\{x_a, y_a\}$, $\{x_b, y_b\}$ and $\{x_c, y_c\}$ at times t, $t + \frac{\tau}{2}$ and $t + \tau$.

Kernel Regression [11] to obtain an estimate of the location using p, the sampled distribution of RSSI:

$$(\bar{x}, \bar{y})^T = \frac{\sum_l (x_l, y_l)^T k(p, q^{\{x_l,y_l\}})}{\sum_l k(p, q^{\{x_l,y_l\}})} \tag{4}$$

We propose to do this regression using only the N nearest neighbors (in the KL-divergence sense), instead of the full set of known training datapoints, i.e. to keep the N fingerprints $\left\{q^{\{x_l,y_l\}}\right\}$ that maximize $k\left(p, q^{\{x_l,y_l\}}\right)$. Our methods amounts to nearest neighbor matching in the case when $N = 1$. Note that the choice of the N 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 N 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.

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 goes 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[5]. In other words, for two close positions $\{x, y\}$ and $\{x', y'\}$:

$$q(\mathbf{S}|\lambda\{x,y\}+(1-\lambda)\{x',y'\}) \approx \lambda q^{\{x,y\}}+(1-\lambda)q^{\{x',y'\}} \tag{5}$$

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. 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 $\tau = 8s$ time windows, while the motion speed was 0.5m/s, adjacent training fingerprints were spaced every 2-2.5m, and APs queried 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. For comparison, each training fingerprint would have up to 100 samples. Let us from now on assume that τ is always adjusted to cover 3 fingerprints during tracking. We propose a weighting scheme that involves giving a smaller weight $\frac{\kappa}{2}$ to samples from $q^{\{x_a,y_a\}}$ collected at the beginning $[t, t+\frac{\tau}{4})$ of the sampling window, and to samples from $q^{\{x_c,y_c\}}$ at the end $[t+\frac{3\tau}{4}, t+\tau)$ of that window, and $1-\kappa$ to samples from $q^{\{x_b,y_b\}}$ in the middle window $[t+\frac{\tau}{4}, t+\frac{3\tau}{4})$. κ can be cross-validated 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 $\frac{\kappa}{2}p^{\{x_a,y_a\}} + (1-\kappa)p^{\{x_b,y_b\}} + \frac{\kappa}{2}p^{\{x_c,y_c\}}$ and the actual $q^{\{x_b,y_b\}}$[6]. 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.

We are currently investigating the impact of the choice of the sampling window τ and the size of the RSSI bins.

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[7], 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). As we show in the next section, we can thus achieve a median accuracy of 2 to 3m in an office environment, even though we ignore the actual RSSI levels.

One benefit from 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

For our first set of experiments, we used a 2D office dataset used in [6], [4], consisting of a 40m×40m area, illustrated on Fig. 3. The training data consisted of 88 fingerprints recorded for 22 APs[8]); some APs had over 100 samples for each location. 4 APs only were used in the published experiments.

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 N 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.

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

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

[7]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.

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

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	1.34m	3.11m
KL divergence, 3 NN WKR	**1.06m**	**2.29m**
KL divergence, 88 NN WKR	1.26m	2.70m

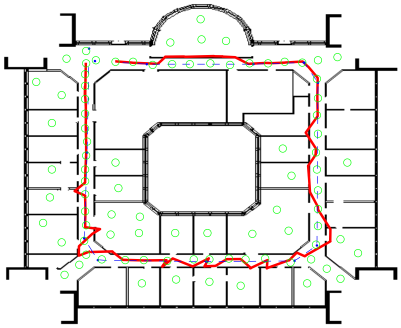

Fig. 3. Tracking results on the 2D office dataset using 4 APs. The true path is in dashed blue, the estimated path in solid red line, and the 88 fingerprint locations appear as green circles. We used a KL-divergence kernel with weight $\alpha = 0.011$, $\tau = 8s$ windows and performed kernel regression on the $N = 3$ nearest neighbors. Median error was 1.06m and 2.3m at the 90% percentile. RSSI was sampled at 5Hz, yielding up to 40 samples per sampling window for each AP.

Tracking data in that dataset were acquired a few days later, and 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 ($N = 3$) 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 1.34m.

We did not observe a decrease in the median accuracy when using 22 APs rather than 4 APs, similarly to what was suggested in [7], but as shown in Table II, the 90% quantile error was reduced to around 2m.

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

We acquired another dataset by walking at constant speed (around 1.4m/s) along a 320m corridor. NetStumbler would query APs only at 1Hz. We used 8s-long sampling windows to create 55 fingerprints (AP connection histograms only) spaced every 4s (i.e. every 5.5m) for 130 APs. When we used those fingerprints to localize ourselves later on the same day (while in motion at 1.4m/s), we achieved 3.4m median accuracy (7.9m 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 3.8m median accuracy (7.8m at 90%), in spite of some missing APs. These results are upper bounds: more careful (slower) fingerprinting and accounting for speed fluctuations should bring the errors down.

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	1.28m	3.56m
With RSSI, 6 NN WKR	1.12m	2.10m
With RSSI, 88 NN WKR	**1.05m**	**1.81m**
No RSSI, 1 NN	2.27m	5.05m
No RSSI, 14 NN WKR	1.93m	4.41m
No RSSI, 88 NN WKR	1.93m	4.41m

IV. 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. Since corridors may be idealized environments for signal propagation and fingerprinting, we are currently also experimenting with various open-space indoor environments.

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] L. Koski, T. Perala and R. Piche, "Indoor Positioning Using WLAN Coverage Area Estimates", *IPIN*, 2010.

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

[9] 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.

[10] 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.

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

[12] 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.

[13] 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.

[14] T. Vauper, J. Seitz, F. Kiefer, S. Haimerl and J. Thielecke, "Wi-Fi Positioning: System Considerations and Device Calibration", *IPIN*, 2010.