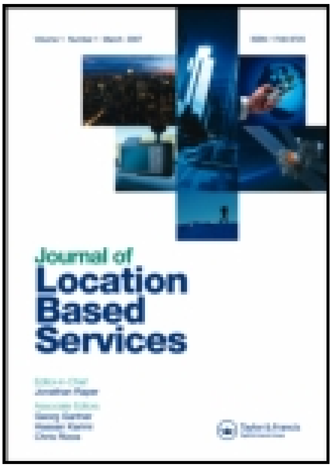

Journal of Location Based Services
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tlbs20

Probability kernel regression for WiFi localisation

Piotr Mirowski [a], Philip Whiting [a], Harald Steck [a], Ravishankar Palaniappan [a], Michael MacDonald [a], Detlef Hartmann [b] & Tin Kam Ho [a]

[a] Bell Laboratories, Alcatel-Lucent, 600 Mountain Avenue, Berkeley Heights, NJ 07974, USA

[b] Bell Laboratories, Alcatel-Lucent, Colditzstrasse 34-36, Berlin 12099, Germany

Published online: 20 Jun 2012.

To cite this article: Piotr Mirowski, Philip Whiting, Harald Steck, Ravishankar Palaniappan, Michael MacDonald, Detlef Hartmann & Tin Kam Ho (2012) Probability kernel regression for WiFi localisation, Journal of Location Based Services, 6:2, 81-100, DOI: 10.1080/17489725.2012.694723

To link to this article: http://dx.doi.org/10.1080/17489725.2012.694723

Journal of Location Based Services
Vol. 6, No. 2, June 2012, 81–100

Probability kernel regression for WiFi localisation

Piotr Mirowski[a]*, Philip Whiting[a], Harald Steck[a], Ravishankar Palaniappan[a], Michael MacDonald[a], Detlef Hartmann[b] and Tin Kam Ho[a]

[a]*Bell Laboratories, Alcatel-Lucent, 600 Mountain Avenue, Berkeley Heights, NJ 07974, USA;* [b]*Bell Laboratories, Alcatel-Lucent, Colditzstrasse 34-36, Berlin 12099, Germany*

(*Received 3 December 2011; final version received 3 April 2012; accepted 15 May 2012*)

Various methods have been developed for indoor localisation using WLAN signals. Algorithms that fingerprint the received signal strength indicators (RSSI) of WiFi for different locations can achieve tracking accuracies of the order of a few metres. RSSI fingerprinting suffers 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. Mirowski *et al*. [2011. KL-divergence kernel regression for non-Gaussian fingerprint based localization. *In*: *International conference on indoor positioning and indoor navigation*, Guimaraes, Portugal] have recently introduced a simple methodology that takes into account the full distribution for computing similarities among fingerprints using the Kullback–Leibler (KL) divergence, and then performs localisation through kernel regression. Their algorithm provides a natural way of smoothing over time and motion trajectories and can be applied directly to histograms of WiFi connections to access points, ignoring RSSI distributions, hence removing the need for fingerprint recalibration. It has been shown to outperform nearest neighbours or Kalman and particle filtres, achieving up to 1 m accuracy in office environments. In this article, we focus on the relevance of Gaussian or non-Gaussian distributions for modelling RSSI distributions by considering additional probabilistic kernels for comparing Gaussian distributions and by evaluating them on three contrasting datasets. We discuss their limitations and formulate how the KL-divergence kernel regression algorithm bridges the gap with other WiFi localisation algorithms, notably Bayesian networks, support vector machines and K nearest neighbours. Finally, we revisit the assumptions on the fingerprint maps and overview practical WiFi localisation software implementation.

Keywords: WiFi positioning; fingerprinting; localisation; distributions; kernel methods

1. Introduction

Indoor tracking of people and objects based on WiFi signal strength measurements can be performed with an accuracy of a few metres in a typical building. As a first step, localisation methods require laborious human involvement in the training

*Corresponding author. Email: piotr.mirowski@computer.org

ISSN 1748–9725 print/ISSN 1748–9733 online

http://dx.doi.org/10.1080/17489725.2012.694723
http://www.tandfonline.com

phase to build the so-called *fingerprint* maps for each access point (AP). In the 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 neighbour matching (Bahl and Padmanabhan 2000), 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 (Evennou *et al.* 2005).

Those fingerprint maps, however, generally store only the mean value of RSSI (Evennou *et al.* 2005), sometimes the mean and variance of RSSI (Chen *et al.* 2007) and do not fully 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. The specific problem we try to overcome is the non-Gaussianity of the RSSI distribution in the context of WiFi localisation (Section 1.1). Our probabilistic localisation algorithm (Section 1.2) relies on comparing RSSI distributions, and for that effect, we extend the Kullback–Leibler (KL)-divergence kernel regression (Mirowski *et al.* 2011) with additional probabilistic kernels (Section 1.3). This article focuses on the theoretical foundations, limitations and practical implementation of our method.

The issue of frequent re-training, necessary to maintain accuracy, is beyond the scope of this work. Workarounds have been designed for recalibrating fingerprints, such as automating the process with self-driving indoor mapping robots (Palaniappan *et al.* 2011) that match the position of a vehicle with recorded RSSI to automatically build a signal map. Alternatively, a server-based positioning system can be designed so that APs self-calibrate by the mutual analysis of their RSSI (Cypriani *et al.* 2011).

1.1. *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 (di Flora and Hermersdorf 2008, Vauper *et al.* 2010) that this is not always the case: the signal can be multimodal, different recording devices can measure quite different distributions at the same location and simple changes in antenna orientation can impact the RSSI by 10 dBm (Curran *et al.* 2011). In their experiments, Mirowski *et al.* (2011) reported that the RSSI of an immobile receptor can be distributed in a bimodal way and they oscillate between two extreme values distant by as much as 10 dB (Figure 1).

Because of non-Gaussianity, the use of mean and variance of a multimodal distribution may ignore important information that is helpful for discriminating among different locations. A procedure that can provide a richer characterisation of the distribution is needed. One can represent RSSI or signal-to-noise ratio (SNR) distributions by histograms, with the natural binning scheme of one bin for each integer level.[1] This is an intuitive scheme since RSSI values recorded by software such as NetStumbler® (http://www.netstumbler.com) or WiFi Scanner® (http://wlanbook.com) are integers. In the most general case that accounts for the multi-modality of the signals, Mirowski *et al.* (2011) consider multinomial

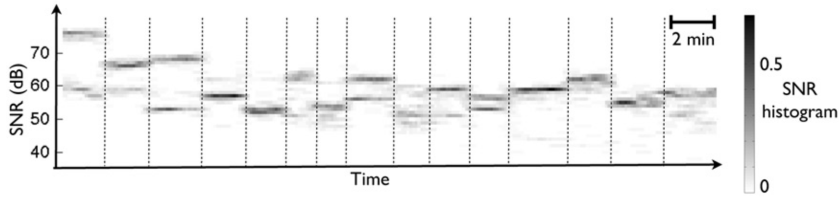

Figure 1. Non-Gaussian distributions of the SNR of the RSSI. Data were recorded over 30 min along a long corridor and for a single AP. The mobile would alternately stop for about two minutes at each location and move 1 m further, repeating these steps for 15 locations. Each vertical line corresponds to one histogram. The histograms have one bin per SNR level, and were constructed using 60s sliding windows and 10s steps. The 15 locations are clearly visible on the graph, as they correspond to stationary RSSI distributions. Source: Mirowski *et al.* (2011).

distributions as a model for RSSI distributions, and compare such multimodal distributions using the divergence metric developed by Kullback and Leibler (1951).

1.2. *Prior art in probability-based indoor localisation*

The first usage of a probabilistic approach to RSSI in indoor localisation was explained in Castro *et al.* (2001), Roos *et al.* (2002) and Youssef *et al.* (2003). They proposed to model the distribution of RSSI at each fingerprint location as a histogram, and to use it as a prior in a Bayesian framework, to compute the probability of having a specific histogram of RSSI at a new location using Bayesian networks (Castro *et al.* 2001, Roos *et al.* 2002) or the Naive Bayes algorithm (Youssef *et al.* 2003). Paschalidis *et al.* (2009) use a KL-based statistical framework for wireless sensor networks localisation (consisting in null hypothesis testing for each fingerprint). Bargh and de Groote (2008) use the KL divergence to find the (single) nearest neighbour in the space of multinomial counts of Bluetooth dongles. Milioris *et al.* (2010) also perform nearest neighbour matching by resorting to 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 1.1.

Alternatively, Del Mundo *et al.* (2011) reported that support vector machines (SVMs) with Gaussian or polynomial kernels could achieve better WiFi location classification accuracy than naive Bayes or nearest neighbours.

1.3. *Proposed improvements*

No prior method to Mirowski *et al.* (2011) considered probability kernels with distance-like metrics between distributions. They introduced a probability kernel-based approach to track location and suggested to compare distributions of RSSI using the symmetrised KL divergence and by constructing probability kernels that can be used in a simple weighted regression scheme and found that their metric on fingerprints was robust to various noise and RSSI distributions, achieving up to 1 m accuracy in office environments. They also offered 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 (Koski *et al*. 2010) or to the simple AP count-based algorithm described in di Flora and Hermersdorf (2008).

In this article, we focus on the relevance of Gaussian or non-Gaussian distributions for modelling RSSI distributions and consider additional probabilistic kernels for comparing Gaussian distributions. We then evaluate, on three contrasting datasets, several probabilistic kernel regression methods and discuss their limitations. We also formulate how the KL-divergence kernel regression algorithm bridges the gap with other WiFi localisation algorithms, notably Bayesian networks, SVMs and K nearest neighbours (KNN). Finally, we revisit the assumptions on the fingerprint maps that were made in Mirowski *et al*. (2011) and overview practical software implementation of that localisation algorithm.

2. Methods: probabilistic kernel regression for WiFi localisation

We build on the method introduced in Mirowski *et al*. (2011), which can be summarised as follows: one samples the distribution p of RSSI from all visible APs for a short duration and compares it the distributions q in the fingerprint database, using a metric such as the KL divergence and the KL-divergence kernel (Section 2.1). In the database, each fingerprint is associated with a location, and the predicted location is estimated through kernel regression. This method naturally copes with unknown RSSI, contains few hyperparametres and can be trivially extended to operate merely on histograms of AP connection (i.e. multinomial distributions) instead of full RSSI levels (Section 2.3). This article introduces several extensions to this method: comparisons with Gaussian kernels representing different distributions (Section 2.4) and with Bayesian networks (Section 2.5). We also reconsider how to sample RSSI or AP during motion of the mobile device in Section 2.2.

2.1. *KL divergence kernel regression*

The distribution p of discrete-valued RSSI is sampled from all visible APs for a duration τ (typically of a few seconds), and is compared to the distributions q in the labelled fingerprint database, using the symmetrised KL divergence:[2] $D(p,q) = KL(p\,||\,q) + KL(q\,||\,p)$. In the discrete case where the random variable S takes discrete values (e.g. integer-valued RSSI or SNR from an AP), the KL divergence is defined as: $KL(p\,||\,q) = \sum_s p(S=s)\ \log\ (p(S=s)/q(S=s))$. To avoid taking logarithms of zero-valued bins, it can be smoothed by adding a small constant term.

In the case when the discrete random vector representing RSSI values $\mathbf{S}$ is multivariate (e.g. when measuring RSSI from J multiple APs), Mirowski *et al*. (2011) made the assumption of local independence of each AP's marginal distribution and used the chain rule for relative entropy (Cover and Thomas 2006) to express the KL-divergence of a joint distribution of independent variables. One 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.

Mirowski *et al.* (2011) proposed to combine the KL-divergence with kernel[3] methods and to use kernel-based regression algorithms. Following Moreno *et al.* (2003), and for a data-dependent range of values α, it is possible to define such positive semi-definite kernels by exponentiating the symmetrised KL-divergence: $k(p, q_\ell) = e^{-\alpha \sum_{j=1}^{J} D\left(p(S_j), q(S_j | \{x_l, y_l\})\right)}$. Using that kernel function and a set of known training datapoints $\{q_\ell, \{x_\ell, y_\ell\}\}$, weighted kernel regression (WKR) (Nadaraya 1964) can produce an estimate of location using p, the sampled distribution of RSSI:

$$\{\hat{x}, \hat{y}\} = \frac{\sum_\ell \{x_\ell, y_\ell\} k(p, q_\ell)}{\sum_\ell k(p, q_\ell)} \tag{1}$$

For speed-up, Mirowski *et al.* (2011) suggested to do WKR-based regression using only the KNNs (in the KL-divergence sense), instead of the full set of known training datapoints. The two hyperparametres, namely the kernel coefficient α and the number of KNN, are optimised on the training dataset (i.e. on the fingerprints) by leave-one-out cross-validation. WKR reduces to nearest neighbour matching when $K = 1$. The impact of the choice of the length of sampling window τ, of the size of the RSSI bins, and of the number N of samples taken during fingerprinting have been investigated in Mirowski *et al.* (2011).

Finally, when the signal fingerprint at location $\{x, y\}$ does not include any RSSI from a specific AP j, its distribution can be considered as $p(S_j \leq s_{\min} | \{x, y\}) = 1$, where $s_{\min}$ is the limit of detection of the signal. One can approximate this by putting all the mass on the first ('lowest') bin of the RSSI histogram.

2.2. *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 over the course of the sampling window τ. Note that our specific sampling window scheme gives an estimate for the location before $\tau/2$s. Such an approach has the advantage of reducing the position estimate bias but suffers from the disadvantage of giving 'old' position estimates when the sampling window τ lasts for several seconds. This problem is perceived as more acute for real-time tracking of fast moving objects than for self-localisation (as the end-user typically slows down or stops to read her position). Our method could be further combined with a motion model to make position predictions.

One can make the crucial assumption that the distributions continuously change for neighbouring points. There is however a trade-off between the number of RSSI samples necessary to get a good approximation of p, and the error introduced by sampling from neighbouring locations. Throughout our experiments, we limited the sampling window τ so that it covers the spatial extent of at most 3 fingerprints. For instance, in Section 3.1, $\tau = 8$ s, corresponding to 4 m at a leisurely walking speed of 0.5 m/s, compared to a 2.5 m inter-fingerprint distance; whereas in Section 3.2, $\tau = 10$ s corresponds to 10m at 1 m/s, compared to a 5.5 m average distance between fingerprints.

An approximation suggested in Mirowski *et al.* (2011) is a weighted smoothing scheme that divides the sampling window (of duration τ seconds) into three segments: a 'first' quarter, a 'middle' half and a 'last' quarter, respectively, weighted by $\kappa/2$, $1-\kappa$ and $\kappa/2$, so as to give less importance to samples acquired at the beginning and at the end of the sampling window, as well as a recipe for setting the weights κ of the sampling window that exploits the fingerprint data and assumptions about the movement speed. We contend that an alternative optimisation of the κ weights could be made by looking at the localisation performance directly, in a bias and variance minimisation setting, where the bias is the systematic error that is introduced by sampling RSSI before and after the tracked location. The smaller the number of RSSI samples, the larger the variance of the location estimate.

2.3. *Extension to AP connection histograms*

The 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, obtained by counting the number of RSSI measurements made on each AP and irrespective of the RSSI values). Even though one ignores the actual RSSI levels, one can thus achieve a median accuracy of 2–3 m in an office environment, as shown in the next section.

One benefit of this 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 is to remove, from all histograms, the transient APs that do not show up during tracking. Alternatively, one 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. Another way of filtering out APs is to weed out devices with short ranges.

2.4. *Comparison with Gaussian kernel regression methods*

The multinomial KL divergence kernel introduced in Mirowski *et al*. (2011) faces two major shortcomings, which become manifest in datasets with sparse and under-sampled fingerprints.

(1) **Data scarcity**: if only few samples of varying RSSI or SNR levels are collected,[4] it is impossible to represent the fingerprint distribution well by a histogram. In that case, just using the mean (and, optionally, the variance) might be a better estimator for the empirical distribution. In this case, our assumption is that the lack of Gaussianity in the observed distribution is caused by under-sampling rather than by an intrinsic non-Gaussian distribution; therefore we resort to the assumption of Gaussianity to lessen the need for estimating many parameters.
(2) **Non-overlapping histograms**: the multinomial KL divergence kernel does not make a distinction between two non-overlapping histograms that are only 10 dB apart, and two non-overlapping histograms that are 40 dB apart.

For these reasons, we have investigated a simpler, Gaussian form of the KL divergence kernel (Section 2.4.1), its relation to the standard KNN kernel (Section 2.4.2), and propose three different kernels that may be better suited for under-sampled RSSI distributions (Section 2.4.3).

2.4.1. *Gaussian KL divergence kernel regression*

Let us denote by $\mathcal{N}(\mu_{\mathbf{p}}, \mathbf{\Sigma_p})$ and $\mathcal{N}(\mu_{\mathbf{q}}, \mathbf{\Sigma_q})$ the two multivariate normal distributions p and q that will be fitted to some RSSI measurements coming from J APs. Then, as stated in Milioris *et al*. (2010), the KL divergence between these two Gaussians can be written as follows (Equation (2)). Since we assumed that the signals received from the J APs at any given location are conditionally independent given that location, then the covariance matrices $\mathbf{\Sigma_p}$ and $\mathbf{\Sigma_q}$ are diagonal. Letting $\sigma^2_{p,j}$ and $\sigma^2_{q,j}$ be their respective j-th diagonal element, we can simplify Equation (2) to obtain Equation (3):

$$KL(p||q) = \frac{1}{2}\left(\mathrm{tr}\left(\mathbf{\Sigma_q^{-1}\Sigma_p}\right) + (\mu_{\mathbf{q}} - \mu_{\mathbf{p}})^{\mathrm{T}}\mathbf{\Sigma_q^{-1}}(\mu_{\mathbf{q}} - \mu_{\mathbf{p}}) - \log\left(\frac{\det(\mathbf{\Sigma_p})}{\det(\mathbf{\Sigma_q})}\right) - J\right) \quad (2)$$

$$KL(p||q) + KL(q||p) \approx \frac{1}{2}\sum_j\left(\frac{\sigma^2_{p,j}}{\sigma^2_{q,j}} + \frac{\sigma^2_{q,j}}{\sigma^2_{p,j}} + \left(\frac{1}{\sigma^2_{q,j}} + \frac{1}{\sigma^2_{p,j}}\right)(\mu_{q,j} - \mu_{p,j})^2\right) - 2J \quad (3)$$

Equation (3) can be exponentiated to form a kernel function. Software implementation of such kernel merely requires to estimate, individually for each AP $j \in \{1, \ldots, J\}$, the mean and variance of the AP-specific RSSI values (i.e. the first and second moments on univariate distributions p_j and q_j).

The case when an AP j is not 'heard' at a fingerprint location or during tracking (i.e. when univariate distribution p_j is undefined) could be handled as in Section 2.1 by setting the probability mass at the limit of detection, i.e. $p_j(S \leq s_{\min}) = 1$. This yields a mean of $\mu_{p,j} = s_{\min}$ but leaves the variance undefined. We simply chose to set all such variances to a minimum value $\sigma^2_{\min}$. We experimented (on several, non-reported datasets) with $\sigma^2_{\min}$ values corresponding to standard deviations of 1, 2, 3, 5 and 9 dB, and consistently obtained the best performance with $\sigma_{\min} = 5$ dB. We therefore used that value throughout the experiments reported in Section 3 (*a posteriori*, $\sigma_{\min} = 5$ dB turned out to be optimal on those three datasets as well).

2.4.2. *Limit case: distance of mean vectors and time-averaged KNN algorithm*

In the special case when the variance of the RSSI is assumed to be the same for each AP, regardless of fingerprint location or tracking, i.e. $\forall \ell, \forall j, \sigma^2_{\ell,j} = \sigma^2$, a constant then Equation (3) becomes the simple Euclidian distance of mean vectors: $KL(p||q) + KL(q||p) \approx d(p,q) = \frac{1}{\sigma^2}\sum_j (\mu_{p,j} - \mu_{q,j})^2$. The weighted kernel regression with KNN simplifies subsequently to the standard KNN algorithm (albeit with time-averaging) employed for instance in Bahl and Padmanabhan (2000) and Chen *et al*. (2007).

2.4.3. *Three additional kernels for comparing RSSI distributions*

Let us note 'KLgauss' the kernel based on the KL divergence between Gaussian distributions (Section 2.4.1), 'DistMean' the kernel based on distance between mean RSSI vectors (Section 2.4.2) and 'KL' the kernel introduced in Mirowski *et al*. (2011) and explained in Section 2.1. A last, fourth, kernel can be constructed by combining the KL and the DistMean kernels: $k(p, q) = \exp(-\alpha\, D(p, q) - \beta\, d(p, q))$. In the latter expression, the two hyperparametres α and β (as well as the number of nearest neighbours for WKR) can be optimised in a leave-one-out cross-validation scheme.

It is to be expected (and can be seen Section 3) that the hybrid kernel should have a localisation accuracy about as good as the better among the KL and the DistMean kernels.

2.5. *Relationship between KL-divergence kernels and Bayesian methods*

We conclude this methodology section by proving the relationship between KL divergence kernel regression for WiFi localisation (Mirowski *et al.* 2011) and previous Bayesian probabilistic approaches to that problem (Castro *et al.* 2001, Roos *et al.* 2002). Assuming that we know the true fingerprint distributions q_ℓ at every fingerprint location $\{x_\ell, y_\ell\}$, we can express the probability of observing a sequence of discrete, integer RSSI measurements $\mathbf{S}$ (expressed as a histogram $\{h_1, h_2, \ldots, h_I\}$) by a multinomial distribution[5] that is conditional on the location of any such fingerprint ℓ, and that can be expressed using the true signal distribution q_ℓ.

$$p(\mathbf{S}|\{x_\ell, y_\ell\}) = p(\{h_1, h_2, , h_I\}|\{x_\ell, y_\ell\}) \equiv \prod_{s=1}^{I} q_\ell(s)^{h_s} \tag{4}$$

If we normalise the histogram counts h_s by the total number α of RSSI measurements, we obtain the empirical distribution $p(s) \equiv h_s/\alpha$. Then, taking the negative log of the likelihood expressed in Equation (4), normalising it by α and subtracting the Shannon entropy $H(p) = -\sum_s p(s) \log p(s)$ from that quantity, we get the KL divergence between the empirical (tracking) and fingerprint distributions:

$$-\frac{1}{\alpha} \log p(\mathbf{S}|\{x_\ell, y_\ell\}) - H(p) = KL(p||q_\ell) \tag{5}$$

In a Bayesian framework and assuming that each fingerprint location ℓ can be given a uniform prior distribution p ($\{x_\ell,\ y_\ell\}$), we use the Bayes rule to derive the location probability p ($\{x_\ell,\ y_\ell\}|\mathbf{S}$) in Equation (6). We can normalise the right-hand side of Equation (6) using a partition function $Z = \sum_{\ell'} p(\mathbf{S}|\{x_{\ell'}, y_{\ell'}\})$, and then compute the expected value $E[\{x,\ y\}]$ of the location having measured the signal, as in Roos *et al.* (2002). Using Equation (5) and the observation that p $(\mathbf{S}|\{x_\ell, y_\ell\}, q_\ell) = e^{-\alpha\ [KL\ (p\,||\,q_\ell) + H(p)]}$, we can express, in Equation (7), the expected location in terms of exponentiated KL divergence.

$$p(\{x_\ell, y_\ell\}|\mathbf{S}) = \frac{p(\mathbf{S}|\{x_\ell, y_\ell\})\, p(\{x_\ell, y_\ell\})}{p(\mathbf{S})} \propto p(\mathbf{S}|\{x_\ell, y_\ell\}) \tag{6}$$

$$E[\{x, y\}] = \sum_\ell \{x_\ell, y_\ell\} \frac{p(\mathbf{S}|\{x_\ell, y_\ell\})}{Z} = \frac{\sum_\ell \{x_\ell, y_\ell\} e^{-\alpha[KL(p||q_\ell) + H(p)]}}{\sum_\ell e^{-\alpha[KL(p||q_\ell) + H(p)]}} \tag{7}$$

In the above equation, the entropy term $H(p)$ appears both in the numerator and denominator and can be simplified. Yet, and although similar to Equation (1), Equation (7) is asymmetric, hence it precludes many kernel methods such as SVMs. Moreover, the normalising coefficient α that we obtain in our experiments is typically orders of magnitude smaller than the total number of RSSI observations during the sampling window.

Ultimately, kernelising the similarity between distributions, be it through nonparametric or parametric KL divergence as in Mirowski *et al.* (2011) or simply through the distance between mean vectors, proves to be a flexible method. Because of its relation to the probabilistic Bayesian networks (Roos *et al.* 2002) and naive Bayes classifiers (Youssef *et al.* 2003), to SVMs and to KNN (Bahl *et al.* 2000), our kernel-based method essentially encompasses most of the state-of-the-art machine-learning algorithms used in WiFi localisation (Del Mundo *et al.* 2011).

3. Results

In this section we refine the localisation results obtained by Mirowski *et al.* (2011) by employing the three additional kernels that we just introduced. We compare various buildings featuring different training and tracking scenarios, including a high-quality dataset in an office building with dense and repeated RSSI fingerprints (Section 3.1), a trial involving a large, open space (Section 3.2) and another trial in a public indoor environment with mixed layouts and heavy pedestrian traffic (Section 3.3). We thus re-analyse the robustness of Gaussian and non-Gaussian distribution kernel-based localisation algorithms under different settings.

All our experiments mimic a scenario where the location is predicted only from the last few seconds of RSSI measurements and using a pre-computed fingerprint database that can be stored locally on a hand-held device (Section 4.4). The duration τ of the time window during which we collect RSSI samples is typically equal to 8 or 10 s; in order not to bias the predicted locations towards the past, we evaluate the position error by comparing the prediction to the actual position $\tau/2$ s earlier.

3.1. *Office space with dense fingerprinting*

The first dataset is taken in an office (Evennou *et al.* 2005, Chen *et al.* 2007) consisting of a 40 m × 40 m area, depicted in Figure 2. The training data consisted of 88 fingerprints recorded for 22 APs[6] and spaced every 2.5 m on average; some APs had 130 samples for each location. Tracking data in that dataset were acquired a few days later using the same WiFi equipment,[7] sampling RSSI at 5 Hz frequency while moving at a speed of 0.5 m/s. Our tracking algorithm collected samples in windows of duration $\tau = 8$ s, resulting in up to 40 samples during tracking time and covering a distance of 4 m. Because only 4 APs were used in the experiments published in Chen *et al.* (2007) and Evennou *et al.* (2005) (each AP was placed at one of the 4 corners of the square-shaped corridor, so as to maximise the line-of-sight coverage), we evaluated tracking performance using only those 4 APs (for fair comparison), as well as using all 22 APs.

3.1.1. *Localisation based on RSSI*

Table 1 recapitulates the performance of multinomial KL divergence kernel regression with 4 APs (achieving a median accuracy of 1.06 m), which outperforms state-of-the-art algorithms such as Kalman filtres (2 m), Voronoi particle filtres (1.6 m) (Evennou *et al.* 2005) and model-free tracking (1.3 m) (Chen *et al.* 2007), while estimating a reasonably smooth trajectory.

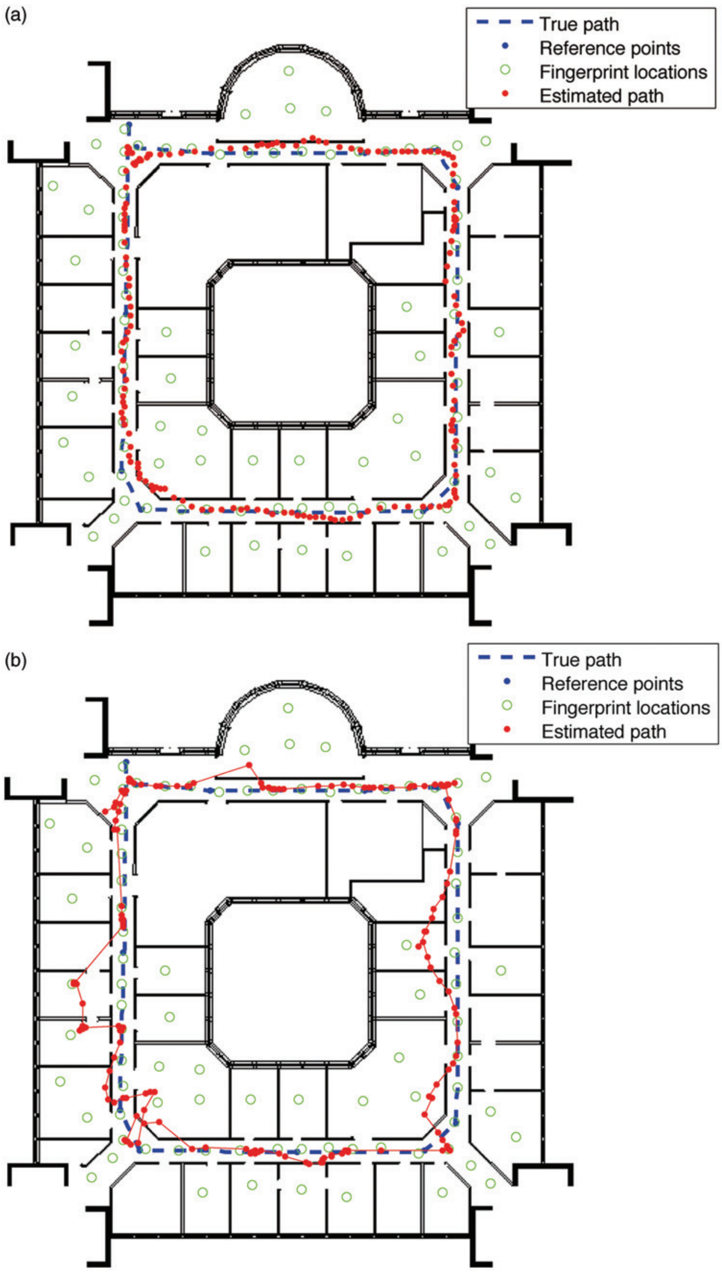

Figure 2. Tracking results on the office dataset: the true path is in dashed blue, the estimated path in solid red line, and the 88 fingerprint locations appear as green circles. RSSI was sampled at 5 Hz, yielding up to 40 samples per sampling window for each AP (there were 22 APs). (a): results obtained using a KL-divergence kernel with weight $\alpha = 0.041$, $\tau = 8$ s windows and kernel regression on the $N = 88$ fingerprints, with a very low median localisation error of 0.86 m and 1.72 m at the 90% percentile. (b): results obtained using a KL-divergence kernel only on the AP connection counts (ignoring the RSSI values); $\alpha = 77$, $\tau = 8$ s windows and kernel regression on the $N = 88$ fingerprints. Despite not using the signal values, we obtained a satisfactory localisation error of 1.94 m and 4.95 m at the 90% percentile.

Table 1. Tracking accuracy on the office dataset using 4 APs.

Technique	Median (m)	at 90% (m)
Kalman filtre (Evennou *et al*. 2005)	2.0	–
Voronoi particle filtre (Evennou *et al*. 2005)	1.6	–
Model-free tracking (Chen *et al*. 2007)	1.3	2.5
KL divergence, 1 nearest neighbour (NN)	1.25	3.18
KL divergence, 3 NN WKR	1.06	2.34

Table 2. Tracking accuracy on the office dataset using the KL-divergence kernel on 22 APs, with or without RSSI.

Technique	Kernel	Median (m)	at 90% (m)
With RSSI, 1 NN	KL	1.16	2.84
With RSSI, 6 NN WKR	KL	0.96	1.88
With RSSI, 7 NN WKR	KLgauss	1.29	2.18
With RSSI, 5 NN WKR	DistMean	1.24	2.32
With RSSI, 7 NN WKR	Hybrid (KL + DistMean)	0.93	1.72
No RSSI, 1 NN	KL	1.94	4.95
No RSSI, 27 NN WKR	KL	1.90	4.31

A further decrease in the median tracking error was observed when using 22 APs rather than 4 APs, contrary to what was suggested in Koski *et al*. (2010); as shown in Table 2, the 90% quantile error was reduced to around 1.9 m from 2.3 m, and the median error was slightly reduced to 0.95 m from 1 m after including all the available 22 APs. 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.

As given in Table 2, other kernels such as the KL divergence on Gaussian or the distance between mean RSSI vectors worsened the localisation error (going over 1.25 m median and 2.2 m at 90%), whereas using a hybrid kernel (combining the multinomial KL divergence and the distance of means) slightly reduced the error down to 0.93 m median and 1.72 m at 90%. This hints that, on this dense and well-sampled dataset (where up to 130 measurements were recorded at each fingerprint, and where fingerprints were spaced by 2.5 m), the multinomial KL divergence makes best use of the RSSI distributions.[8] We now hypothesise that dense, well-sampled fingerprint maps might be better modelled by multinomials rather than by Gaussians.

3.1.2. *Localisation based on AP visibility*

In a second series of experiments on the same office dataset, the RSSI from the APs were ignored, and only multinomials of connections to 22 APs were used to build the KL-divergence kernels. As shown in Table 2 and Figure 2(b), the tracking accuracy remained satisfactory, at about 2 m median error.

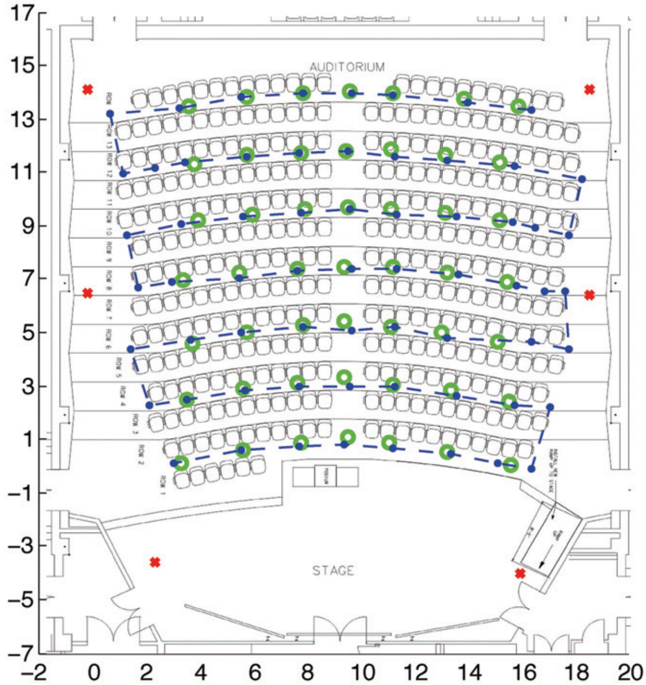

Figure 3. Floorplan of the open-space environment (auditorium), with 49 fingerprint locations (green circles) and 6 APs (red crosses). Next day's tracking path is indicated in the dashed blue line.
Source: Mirowski *et al.* (2011).

The difference between this method and that given by di Flora and Hermersdorf (2008) is that here, AP visibility vectors were pooled from consecutive locations covered by the walk over a small temporal window, and were used to estimate a multinomial distribution of AP connections for the location at the centre of the window. Localisation using only AP visibility can provide a more robust option, in particular when fingerprinting is done 'on the fly' while walking or driving a robot (Palaniappan *et al.* 2011), with a large spatial spread of the pooled RSSI measurements. For instance, Mirowski *et al.* (2011) reported 4 m median accuracy (7.6 m at 90%) during tracking, 1 week after having fingerprinted (every 5 m) a 300-m-long corridor by walking at 1.4 m/s and querying 130 APs at 1 Hz in overlapping windows of duration $\tau = 10$ s.

3.2. *Open-space localisation in an auditorium*

It can be argued that a narrow and long corridor is an ideal layout for localisation. Mirowski *et al.* (2011) also evaluated a large, open indoor space (an auditorium with over 200 seats) with fingerprints collected at locations spread evenly over the space and with repeated measurements made at each location (Figure 3). During training, over 60 RSSI values from each of the 6 APs were recorded at 49 fingerprint locations separated by about 2 m, using NetStumbler at the frequency of 1 Hz. As visible in

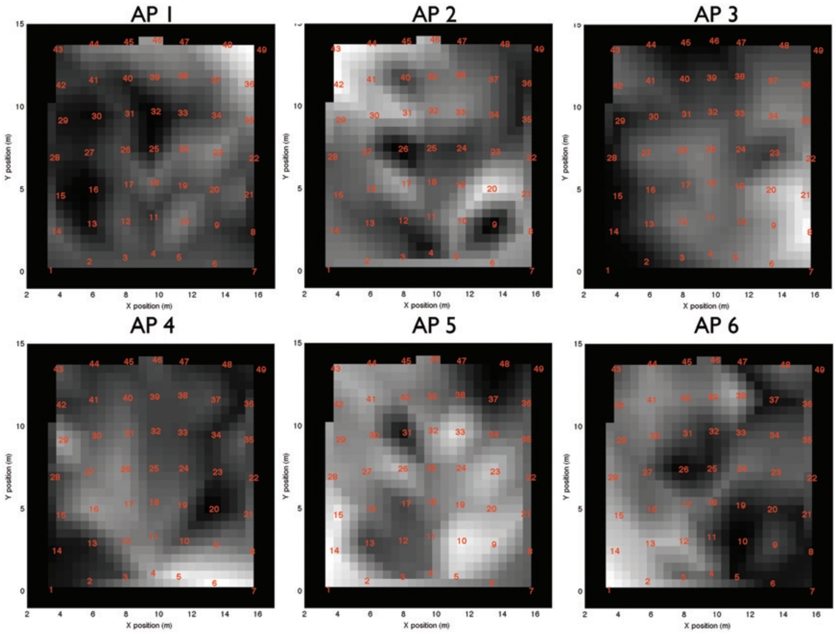

Figure 4. Mean values of the WiFI SNR interpolated every 0.5 m between the 49 fingerprints in the open-space environment (auditorium). Signal 'heat maps' from all 6 APs are shown.

Figure 4, the mean value of the WiFi signal did not monotonically change across the space (X, Y) dimensions: there were multiple adjacent points sharing the same mean RSSI for a given AP channel, e.g. fingerprints 4, 11, 18, 25, 26 and 27 for AP 3. Tracking RSSI was recorded the next day on a path going through all the fingerprints and moving slowly at 0.17 m/s. Figure 3 shows the locations of the fingerprints and the path for tracking.

We evaluated all the four kernels discussed in this article. For the multinomial KL-divergence kernel in particular, we compared the 1-nearest neighbour with weighted kernel regression on multiple nearest neighbours (the number of neighbours K was optimised on the fingerprints) and also compared different bin sizes (1, 2, 5 and 10 dB). Table 3 compares the performance on the tracking data using $\tau = 10$ s long tracking windows. Best results obtained were 4.7 m median accuracy (8.3 m at 90%) for 5 dB-bin KL-divergence 7-neighbour kernel regression (with coefficient $\alpha = 0.39$), followed by 4.9 m median accuracy (8.2 m at 90%) for a Gaussian KL-divergence kernel regression with $K = 4$ and $\alpha = 3.93$. We did notice a significant improvement or worsening of the results when switching between the kernels, provided that bins wider than 1 dB were used, and yet, as illustrated in Figure 5, the WiFi signals need to be represented by multimodal fingerprint distributions.[9] Even more, as shown in the probabilistic analysis based on Chernoff bounds that was reported in Mirowski *et al.* (2011), it is theoretically possible to distinguish between the 49 fingerprints with less than 1% of error using only 5 RSSI measurements.

Table 3. Accuracy results in the auditorium dataset.

Technique	Kernel	Bin size (dB)	Median (m)	at 90% (m)
With RSSI, 1 NN	KL	1	5.3	11.1
With RSSI, 1 NN	KL	2	5.3	11.1
With RSSI, 1 NN	KL	5	4.8	11.2
With RSSI, 1 NN	KL	10	4.9	10.2
With RSSI, 4 NN WKR	KL	1	5.3	11
With RSSI, 4 NN WKR	KL	2	5.1	9.3
With RSSI, 7 NN WKR	KL	5	4.7	8.3
With RSSI, 4 NN WKR	KL	10	5	8.6
With RSSI, 3 NN WKR	DistMean	–	5	9.1
With RSSI, 3 NN WKR	Hybrid (KL + DistMean)	1	5	9.3
With RSSI, 3 NN WKR	KLgauss	–	4.9	8.2

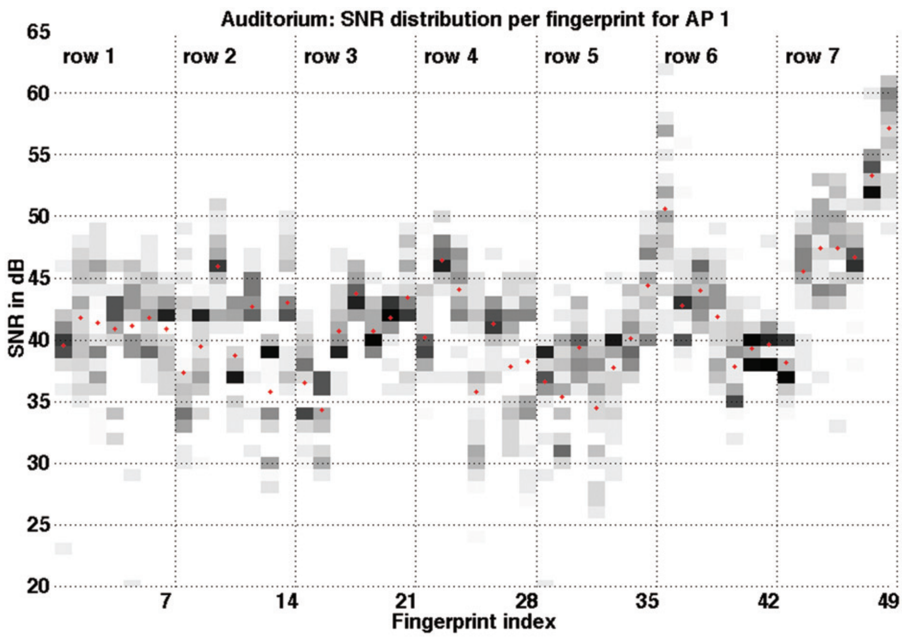

Figure 5. WiFi SNR distribution of the 49 fingerprints in the open-space environment (auditorium). Only access point AP 1 is shown here. Each vertical line corresponds to one fingerprint; the darker a pixel, the 'higher' the count in the corresponding bin of the signal histogram. Each fingerprint was sampled at last 60 times, and the multimodal nature of the signal is apparent for some fingerprints.

We conclude that probabilistic kernel regression algorithms (including the limit-case KNN) do somewhat work in open space areas, though with an accuracy that is lower than in corridors. Low accuracy and the lack of differentiation between various kernels might be attributed to changes in the environment due to moving

Table 4. Accuracy results in the complex public space dataset.

Technique	Sampler	Floor dataset	Kernel	Median (m)	at 90% (m)	floor (%)
With RSSI	NetStumbler	Lower	KL	8.2	16.9	96.2
		Upper	KL	9	17.1	83.8
With RSSI	NetStumbler	Lower	KLgauss	6.7	14.8	96
			DistMean	6.9	16.4	97.1
			Hybrid (KL + DistMean)	7	16.7	96.6
		Upper	KLgauss	8	13.4	84.4
			DistMean	8.4	16.6	88.8
			Hybrid (KL + DistMean)	8.6	16.7	88.8
No RSSI	WiFi Scanner	Lower	KL	10.3	24.3	89
		Upper	KL	9.1	17	92.6

people (the experimenters acquiring the data), multipath propagation and multiple sources of signal diffraction, such as seats. For comparison, the same equipment with dense and well-sampled fingerprinting of multimodal WiFi signals could obtain a median tracking error well under 1 (median) and 2 m at 90% in a corridor environment (Mirowski *et al.* 2011).

3.3. *Localisation 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 wide 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 5.5 m apart from each other on average. During fingerprinting, location errors of the order of 5 m were systematic. 10–15 repeated measurements were obtained at each location. During tracking, samples were pooled over a window of 10 s at a walking speed of 1 m/s.

Mirowski *et al.* (2011) 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 4. In our experiments with the Gaussian kernel regression, we noticed that there was a marked advantage in using the KL divergence kernels on Gaussian distributions for the localisation accuracy (and the simple distance of mean RSSI vectors for floor accuracy) over the multinomial KL divergence kernels. Specifically, Gaussian KL divergence kernel regression would reduce the median error by 1 m and the error at 90% by over 2 m. In this dataset, the fingerprints were under-sampled, being located at least every 5 m and with barely 10 samples per fingerprint.

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.

4. 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 discuss the difficulty of building fingerprint maps. We begin by recalling the assumptions that we employed throughout this article (Section 4.1) and explain the impact of various hyper-parametres of the localisation algorithms (Section 4.2). We provide pointers to several approaches for quantifying the quality of fingerprint maps (Section 4.3), and conclude by computational and software deployment considerations (Section 4.4).

4.1. *Assumptions behind the probabilistic model of fingerprints*

Mirowski *et al.* (2011) defined a *fingerprint* as a set of probability distributions (one for each AP) that are specific to (i.e. conditional on) a location indexed by ℓ. This article focussed on two examples of measurements taken from WiFi-enabled devices which can communicate with the so-called APs (AP): RSSI measurements and AP counts. We followed a few key assumptions in our probability kernel-based location algorithms:

(1) **Device-independent measurements**, even if different WiFi cards on different laptops could record different sets of RSSI values at identical locations, so that one builds device-specific fingerprint maps.
(2) **Conditional independence of the RSSI from a single AP given the location**, which means that at immobility, the measurements are theoretically 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, it becomes impractical during tracking, but one can assume that RSSI acquired during motion over short time intervals are somewhat constant, provided that the speed is relatively slow and the sampling window is relatively short.
(3) **Conditional independence of APs**, as one can 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.
(4) **Time-invariance of fingerprints** (a weak assumption). In spite of Mirowski *et al.* (2011) contending that they can correct for shifts in signal level, ignore new APs and not use removed APs in the localisation algorithm, they cannot easily deal with local changes in the environment, such as furniture or people movements. The best solution might lie in automated fingerprint recalibration or even automatic data acquisition with associated location using a self-localising robotic platform (Palaniappan *et al.* 2011).

4.2. *Impact of fingerprinting and tracking data scarcity on localisation accuracy*

Mirowski *et al.* (2011) investigated four different questions pertaining to some parameters related to fingerprints and to tracking using probability distribution kernels, three of which pertain to all signal-based localisation algorithm:

- How many fingerprinting locations should be chosen?
- How many RSSI samples N should be measured to estimate the fingerprint distributions $q_\ell(\mathbf{S})$?
- During tracking, how many RSSI samples n should be used in the localisation algorithm?
- How wide should be the histogram bins used to encode the RSSI distributions?

They quantified the effects of each of these four hyper-parametres in terms of tracking accuracy with the office data from Section 3.1, by subsampling that data (e.g. fewer fingerprints, fewer samples or wider histogram bins). Their immediate conclusion was 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. After analysis, they suggested that the spatial density of the fingerprints is the most important performance impacting factor. In comparison, repeated measurements at each location were less important – the advantage of multiple measurements at the same location flattened beyond about 20 samples. The optimal bin sizes for the histograms vary with the length of the sampling window during tracking, with more refined bins useful only with longer tracking windows that accumulated more samples to estimate the RSSI distribution.

Our Gaussian versus non-Gaussian distribution kernel results further suggest that datasets with denser fingerprints and with more samples N are more suitable for the multinomial (non-Gaussian) KL-divergence kernels (cf. the office dataset from Section 3.1). On the other hand, datasets with coarser fingerprints (fewer locations and fewer samples) tend to perform slightly better with the Gaussian KL-divergence kernel. We contend that those latter datasets do not yield enough datapoints to correctly estimate the RSSI distributions, hence simpler methods (i.e., with fewer degrees of freedom) work better.

Finally, we noticed (results not reported) that on some under-sampled datasets, increasing the size of bins (e.g. 5 or 10 dB) for the multinomial KL divergence kernel would somewhat increase the accuracy of the tracking algorithm, but not as much as by directly switching to Gaussian KL-divergence kernels or as by using the hybrid kernel (combining 1 dB-bin multinomial KL-divergence kernel with a distance of means kernel).

4.3. *Assessing the quality of fingerprint maps*

Mirowski *et al.* (2011) provided theoretical and computational insights about the number of samples n that RSSI distribution-based localisation methods needed during tracking. Using a probabilistic definition of fingerprints, the actual fingerprint database, sampling and Chernoff bounds on the probability of mistaking a fingerprint for another, they derived an indicator of the 'quality' (i.e. the separability or dissimilarity) of a non-Gaussian fingerprint map. Such a metric is perfect for discovering similar fingerprints that could confuse the localisation algorithm. Their metric would for instance indicate that the office dataset from Section 3.1 (theoretically) needed only or two samples at tracking time to choose the closest fingerprint, whereas the auditorium dataset (in Section 3.2) required five samples (i.e. those fingerprints were more 'confusing'). This Chernoff metric is of course only a theoretical lower bound on the number of samples needed to differentiate any two fingerprints with a 99% probability, not the number of samples needed to build a good histogram that accurately approximates the probability distribution function of the RF equipment (the latter number would be much higher). The Chernoff bound was derived under the assumptions made in Section 4.1 and it does not take into account trivial sources of errors such as the noise in data acquisition or the time-variance of the RSSI. In practice, many more tracking samples n would be required to distinguish between fingerprints, and it is more relevant to directly quantify the localisation accuracy as a function of n, as we suggested in the previous section.

An alternative approach (Beder *et al.* 2011) has been recently investigated to predict the expected accuracy of WiFi maps, including away from the fingerprints, by computing the covariance of the gradients of the fingerprint map in a Bayesian Gaussian setting.

4.4. *Software implementation of probabilistic kernel regression*

We conducted our WiFi localisation research mostly using Matlab, but we also implemented all the kernel regression algorithms mentioned in this article in Java under the Android® software development toolkit and operating system, and tested the algorithm in real time on an Android Smartphone. We exploited the sparsity in the AP visibility and in the RSSI histograms to encode the fingerprint database as a small-size text file and to enable the algorithm to run locally on the smartphone and in acceptable time. As an illustration, the 503-AP and 162-fingerprint public space dataset from Section 3.3 held under 200 kB and we could run one location estimation in less than 1 s. Such a performance enables privacy-preserving localisation applications that run locally on the individual's device instead of a server.

5. Conclusions

We analysed a simple probabilistic algorithm for WLAN fingerprint-based tracking, relying on location regression with KL-divergence kernels. Its time-window-based sampling approach is a very simple way to account both for the motion and for the complex distributions of RSSI. Depending on the quality of the signal map, which may or may not be Gaussian, we discussed optimal kernel choices. Our kernel-based algorithm proves to be very flexible and generalises several existing WiFi localisation algorithms, including KNN and Bayesian networks. Moreover, the structure of our model is such that, exploiting an automated setup for dense fingerprinting, we can further investigate the distributions of location prediction error and thus quantify the localisation uncertainty due to how the WiFi signal distribution varies in space. Finally, although the experiments in this work relied solely on active WiFi probing (in order to accommodate existing hand-held devices such as smartphones), we would like to stress that our algorithm could easily be applied to passive WiFi scanning (if hand-held devices had software to process such large amounts of data) or to any other radio-frequency signals such as those coming from the global system for mobile communications (GSM) or long-term evolution (LTE).

Acknowledgements

The authors would like to thank Benjamin Lowe for his assistance in data collection and helpful feedback, as well as Jin Cao for her very useful input on probability kernels.

Notes

1. Mirowski *et al.* (2011) evaluated the trade-off between coarser binning schemes, e.g. 5 dB bins, and time window lengths.
2. In information theory, the KL divergence is a non-symmetric measure of the difference between probability distributions.
3. A *kernel* is a symmetric function equal to one if $p=q$ and decaying to zero as the dissimilarity of the two inputs increases.
4. Data scarcity would typically be an issue for online tracking data collection. It might also happen during fingerprinting, when the experimenter cannot afford resources to collect more than a dozen RSSI samples per fingerprint location.
5. For simplicity, we limit ourselves to a single AP, but an extension to multiple independent APs is possible.
6. It is common to observe hundreds of unique MAC addresses in office environments, coming from various floors and individual offices.
7. Information about movements of people during tracking or fingerprinting was not available.
8. Mirowski *et al.* (2011) report that on this dataset, there is no significant localisation improvement beyond 20 samples per fingerprint, but that fewer fingerprints are detrimental (see also Section 4.2).
9. We also notice in Figure 5 that some fingerprints seem to have identical RSSI distributions for AP channel 1, e.g. fingerprints 41 and 42: this will confuse the KL-divergence algorithm.

References

Bahl, P. and Padmanabhan, V.N., 2000. RADAR: an in-building RF-based user location and tracking system. *IEEE Infocom*, 2, 775–784.

Bargh, M.S. and de Groote, R., 2008. Indoor localization based on response rate of bluetooth inquiries. *In*: *International workshop on mobile entity localization and tracking in GPS-less environments*, San Francisco, CA, 48–54.

Beder, C., McGibney, A., and Klepal, M., 2011. Predicting the expected accuracy for fingerprinting based WiFi localisation systems. *In*: *International conference on indoor positioning and indoor navigation*, Guimaraes, Portugal.

Castro, P., *et al.*, 2001. A probabilistic location service for wireless network environments. *In*: *Proceedings of ubiquitous computing*, Atlanta.

Chen, A., *et al.*, 2007. An algorithm for fast, model-free tracking indoors. *ACM SIGMOBILE Mobile Computing and Communications Review*, 11 (3), 48–58.

Cover, T.M. and Thomas, J.A., 2006. *Elements of information theory*. 2nd ed. Hoboken, NJ: Wiley Interscience.

Curran, K., *et al.*, 2011. An evaluation of indoor location determination technologies. *Journal of Location Based Services*, 5 (2), 61–78.

Cypriani, M., Canalda, P., and Spies, F., 2011. Performance evaluation of a self calibrated 3D Wi-Fi finger-printing positioning system. *In*: *International conference on indoor positioning and indoor navigation*, Guimaraes, Portugal.

Del Mundo, *et al.*, 2011. A comparison of wireless fidelity (Wi-Fi) fingerprinting techniques. *In*: *International conference on ICT convergence*, Seoul, Korea, 20–25.

di Flora, C. and Hermersdorf, M., 2008. A practical implementation of indoor location-based services using simple WiFi positioning. *Journal of Location Based Services*, 2 (2), 87–111.

Evennou, F., Marx, F., and Novakov, E., 2005. Map-aided indoor mobile positioning system using particle filter. *In*: *IEEE conference on wireless communications and networking*, New Orleans, LA, 2490–2494.

Koski, L., Perala, T., and Piche, R., 2010. Indoor positioning using WLAN coverage area estimates. *In*: *International conference on indoor positioning and indoor navigation*, Hoenggerberg, Switzerland.

Kullback, S. and Leibler, R.A., 1951. On information and sufficiency. *Annals of Mathematical Statistics*, 22 (1), 79–86.

Milioris, D., *et al.*, 2010. Empirical evaluation of signal-strength fingerprint positioning in wireless LANs. *In*: *ACM international conference on modeling, analysis and simulation of wireless and mobile systems*, Bodrum, Turkey.

Mirowski, P., *et al.*, 2011. KL-divergence kernel regression for non-Gaussian fingerprint based localization. *In*: *International conference on indoor positioning and indoor navigation*, Guimaraes, Portugal.

Moreno, P.J., Ho, P.P., and Vasconcelos, N., 2003. A Kullback–Leibler divergence based kernel for SVM classification in multimedia applications. *Neural Information Processing Systems*, 16, 17.

Nadaraya, E., 1964. On estimating regression. *Theory of Probability and Applications*, 9, 141–142.

Palaniappan, R., *et al.*, 2011. Autonomous RF surveying robot for indoor localization and tracking. *In*: *International conference on indoor positioning and indoor navigation*, Guimaraes, Portugal.

Paschalidis, I.C., Li, K., and Guo, D., 2009. Model-free probabilistic localization of wireless sensor network nodes in indoor environments. *In*: *Second international workshop on mobile entity localization and tracking in GPS-less environments*, Orlando, FL, 68–78.

Roos, T., *et al.*, 2002. A probabilistic approach to WLAN user location estimation. *International Journal of Wireless Information Networks*, 7 (3), 155–163.

Vauper, T., *et al.*, 2010. Wi-Fi positioning: system considerations and device calibration. *In*: *International conference on indoor positioning and indoor navigation*, Hoenggerberg, Switzerland.

Youssef, M.A., Agrawala, A., and Shankar, U., 2003. WLAN location determination via clustering and probability distributions. *In*: *First international conference on pervasive computing and communications*, Dallas, TX, 143–150.