DactyLoc: A minimally geo-referenced WiFi+GSM-fingerprint-based localization method for positioning in urban spaces

Kristian Cujia, Martin Wirz, Mikkel Baun Kjærgaard, Daniel Roggen, Gerhard Tröster
Wearable Computing Group, ETH Zurich, Switzerland

***Abstract*—Fingerprinting-based localization methods relying on WiFi and GSM information provide sufficient localization accuracy for many mobile phone applications. Most of the existing approaches require a training set consisting of geo-referenced fingerprints to build a reference database. We propose a collaborative, semi-supervised WiFi+GSM fingerprinting method where only a small fraction of all fingerprints needs to be geo-referenced. Our approach enables indexing of areas in the absence of GPS reception as often found in urban spaces and indoors without manual labeling of fingerprints. The method takes advantage of the characteristic that the similarity of two fingerprints correlates to the distance between their corresponding location. By applying multidimensional scaling, a topology estimation is generated and with the help of a small set of geo-referenced fingerprints anchored to physical locations. An evaluation with an urban-scale data set shows that we can locate a mobile device with a median error of $30m$. While normally all fingerprints of the training set need to be geo-referenced, with our method, only 8% require geo-referencing. We further show that the localization error decreases as new fingerprints are added and converges to an accuracy comparable to related work.**

I. Introduction

Knowing the geographical position of a person enables a large number of location-based mobile applications [1]. State-of-the-art mobile phones contain multiple technologies to provide such location information including GPS, WiFi and GSM-based approaches. GPS provides accurate positions in open sky conditions and less accurate ones or none in urban and indoor areas [2]. These, however, are places where people spend most of their time [3]. Thanks to the vast penetration of cellular and WiFi networks, exploiting existing infrastructures has found great interest. The achievable location accuracy has been found to be sufficient for many mobile phone applications. Additionally, WiFi- and GSM-based approaches have the advantage of performing well in urban areas and indoor venues [1] where GPS reception is limited. Recently, so-called fingerprinting approaches have found great interest in the research community for localization purposes. A fingerprint consists of a list of access points (APs) and their corresponding received signal strengths at a given locations. The assumption is that each fingerprint is unique across the space and thus uniquely represents a particular geographical location. A reference database is built with a training set consisting of geo-referenced fingerprints. To be localized, a mobile device gathers signal strength readings to obtain the fingerprints at the current location and with the help of a fingerprinting algorithm, the closest match in the reference database is found, revealing the location. Bahl *et al.*'s RADAR localization system [4] was a pioneer effort in that direction. In a more recent work, LaMarca *et al.* [3], achieve $20 - 30$ meters median localization accuracy in urban areas with their Place Lab system. One of the key problems with fingerprinting is that a set of fingerprint in close proximity to a location of interest must be known. Thus, training data is required to build a reference database consisting of geo-referenced fingerprints. GPS is often used to obtain the recording location of fingerprints. For collecting such GPS-referenced fingerprints, approaches like war-driving [5] and war-walking [6] became popular. War-walking tends to take more time but provides better accuracy and larger coverage in metropolitan areas as some regions in a city are only accessible by pedestrians [6]. Fingerprinting efforts can be minimized by e.g. geocoded information to bootstrap fingerprinting databases [7]. Following a collaborative approach, fingerprint databases are automatically updated when GPS and WiFi scanning is active on a user's phone [8]. Many of the existing methods assume the availability of accurate GPS signals during the recording of the training set. GPS reception, however, is not always available in many urban areas as well as indoors, limiting the possible indexing space significantly and thus the usefulness of such localization systems. Hence, to provide extensive coverage and high accuracy for urban positioning, methods are required to be able to index urban spaces also in the absence of GPS-based reference information. Existing commercial solutions for space indexation are user surveys which come at a steep price: a large office building can cost $10,000 USD with no maintenance included [9], and existing approaches require expensive equipment [10] or rely on crowdsourced, or 'organic', methods [11], based on manual labeling of the reference locations [12]. In this work we present a fingerprinting approach which does not require a reference location for each fingerprint in the training set. Only a small number of anchor points is required. Our contribution is threefold:

978-1-4673-1954-6/12/$31.00

1) We propose a collaborative, semi-supervised WiFi and GSM (termed WiFi+GSM in the following) fingerprinting method that only require geo-referencing of a fraction of the fingerprints by taking advantage of the characteristic that the similarity between two fingerprints correlates to the distance between the location of the recordings. By applying Multidimensional Scaling (MDS) [13] on the similarity information, we obtain a topology of fingerprints which can be mapped to a geographical coordinate system using a set of geo-referenced fingerprints serving as anchor points. This reference topology can then be used to locate new fingerprints. 2) The topology can be updated with new fingerprints to increase the localization accuracy and to extend the covered space and is therefore suitable for a collaborative approach. 3) We present evaluation results using an extensive urban data set that provides evidence for the feasibility of our approach.

II. Related Work

In this work we investigate the potential of MDS for generating a fingerprint topology estimation by relying on the similarity between fingerprints. MDS is a statistical method which optimizes the placement of samples in an n-dimensional space in such a way that the pairwise distances between the samples corresponds as good as possible to a measured similarity between the samples. Our assumption is that if we can come up with a similarity measure between two fingerprints that corresponds to their respective recording distance, MDS should be able to generate a topology estimation which corresponds to the physical location of the recorded fingerprints. MDS has recently found interest for localization purposes. Shang *et al.* [14] applied MDS to derive the position of nodes in a wireless sensor network (WSN) based exclusively on binary connectivity information of the sensors. The approach first calculates the shortest path between all nodes to obtain a pairwise distance matrix. Afterwards, they apply MDS to obtain a relative topology estimation of the network. This relative topology can be approximated into a real topology if the position of at least three nodes is known. The work was later extended in [15] to work without knowledge of the entire connectivity. The authors in [16] developed a localization algorithm based on MDS and signal strength measurements (RSSI) in a WSN. In their work the RSSI values among nodes with known location were used to construct a map of the network with MDS. Nodes with no location information used the map to determine their locations. The inclusion of signal strength instead of binary connectivity information to obtain a distance estimation is shown to be more effective in both simulations and experiments. Koo *et al.* in [17] apply MDS to WiFi fingerprints. They extract dissimilarities between pairs of WiFi APs from signal strength measurements. Afterwards, they analyzed the dissimilarities to estimate a geometric configuration of WiFi APs using MDS. To validate the scheme, they conducted experiments on five floors in an office building covering an area of $50m$ by $35m$ on each floor with the result that WiFi APs were located within a $10m$ error range. While they intend to locate WiFi APs, we intend to locate mobile devices based on the available WiFi and GSM information. The work that is most comparable to our effort is presented by Pulkkinen *et al.* in [18] where they generate a topology map based on the similarity of WiFi fingerprints. They achieve a median localization error of $1.5m$ by using MDS to generate the reference database and a classic fingerprinting algorithm for the positioning part. The approach in our work, on the other hand, also uses MDS for locating fingerprints. Further, their approach was only evaluated on one floor of a building. For urban-scale deployment as envisioned in our work, additional effects have to be taken into consideration mainly due to the non-linear relation between distance estimation and fingerprint similarity. In this work, we address these issues and evaluate a method for urban-scale positioning using both WiFi and GSM readings.

III. Method

Our localization approach consists of three steps: 1) Building a reference topology from a set of training fingerprints. 2) Providing a location estimation for new fingerprints using the reference topology. 3) Including the new fingerprint in the reference topology to refine and extend it. Figure 1 schematically shows the process to build the reference topology (top) and to obtain a location estimation (bottom). To generate a reference topology, WiFi and GSM fingerprints are collected ①. Among all fingerprints, a pairwise similarity measure is calculated ②. Unreliable similarity measures are removed during the pruning process ③. By applying MDS, a topology estimation can be generated ④. Hereby, MDS tries to optimally place the fingerprints into a two-dimensional configuration that retains the similarity relations between fingerprint pairs. Using a minimal set of geo-referenced fingerprints serving as anchor points, a non-linear mapping to geographical locations is determined ⑤. To obtain a location estimation of a new fingerprint, the same procedure is applied on a subset of the graph. Besides obtaining a location estimation, the fingerprint can also be added to the list of fingerprints of the reference topology which then gets refined and can grow in size. These steps are described in more detail in the following sections.

A. Generating a reference topology

a) Collecting fingerprints and generating a similarity matrix: A fingerprint contains signal strength readings of detectable APs and base stations referenced by their IDs at a given location. Hereby, the set of fingerprints should ideally have the following properties:

- Each fingerprint should be unique across the space to uniquely reference a geographical location. I.e. if two fingerprints are identical, they stem from the same location.

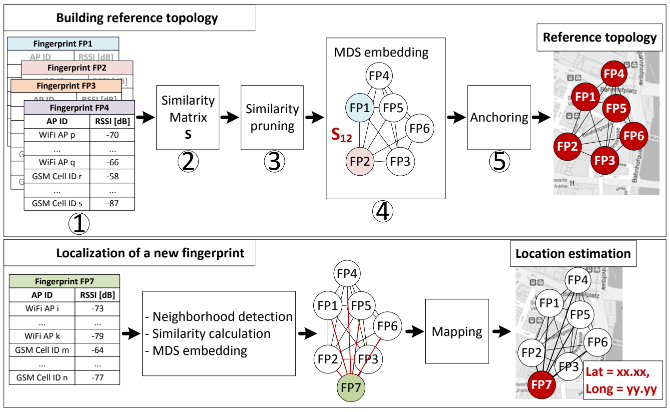


Fig. 1: Process to build the reference topology (top) and to obtain a location estimation (bottom).

- For a given location, the fingerprints should not vary over time.
- The similarity between fingerprints should correlate to the distance between their recordings. Close fingerprints should have a higher similarity compared to those far apart.

However, in practice the fingerprints are effected by both multipath and shadow fading. Our approach provides robustness to mitigate their influences.

The last property is of great importance for our method. We used the Tanimoto coefficient [19] as a metric to determine the similarity between two fingerprints. This measure has been used in other works for the comparison of fingerprints [20]. The metric considers each fingerprint as an n-dimensional vector $\vec{F_i}$ with one dimension for each visible access point and the signal strength as the magnitude in the corresponding direction. The Tanimoto coefficient between two fingerprints $\vec{F_1}$ and $\vec{F_2}$ is then calculated according to Equation 1:

$$T(\vec{F_1}, \vec{F_2}) = \frac{\vec{F_1} \cdot \vec{F_2}}{\left|\left|\vec{F_1}\right|\right|^2 + \left|\left|\vec{F_2}\right|\right|^2 - \vec{F_1} \cdot \vec{F_2}}. \quad (1)$$

The coefficient is bounded between 0 and 1, with $T(F_1, F_2) = 0$ if F_1 and F_2 have no APs in common and $T(F_1, F_2) = 1$ if F_1 and F_2 being the same fingerprint. The similarity matrix is then given by

$$S = \begin{pmatrix} s_{11} & \cdots & s_{1m} \\ \vdots & \ddots & \vdots \\ s_{n1} & \cdots & s_{nm} \end{pmatrix}; \quad s_{ij} = T\left(F_i, F_j\right). \quad (2)$$

b) Pruning the similarity matrix to increase robustness: As we will evaluate in Section V-A, the relation between the Tanimoto similarity measure and the distance between two fingerprints is a monotone, non-linear decaying function. A characteristic of this function is its good discrimination capability within small distances between two fingerprints and its weak performance when the distance is large. To increase robustness against errors introduced by the variance of the similarity estimation from fingerprint-pairs far apart, we disregard their similarity values. To do so, we consider our similarity matrix S as a graph G with vertices for each fingerprint and edges for all pairwise similarities. Hereby, we prune the graph by removing all edges with similarity values below a threshold θ.

In a next step, as some edges of graph G have been removed, some zones of it might end up isolated. Thus, we divide it into subgraphs by finding its connected components with $n > 3$ nodes. Each subgraph is then fed into the MDS algorithm to obtain an embedding into a two-dimensional space.

c) Topology estimation using MDS: Multi dimensional scaling is a method which represents measurements of dissimilarity in a higher dimensional space among pairs of objects as distances between points in a low dimensional space. Through the analysis of dissimilarities between pairs of objects, MDS estimates a mapping into a geometric configuration in a low dimensional space by trying to keep the pairwise original dissimilarity relations [13]. The MDS method takes dissimilarity values as input. We take each subgraph from the previous step and transform all edges into dissimilarities as follows:

$$\bar{s}_{ij} = 1/s_{ij}. \quad (3)$$

By a calculating all shortest paths, we obtain a dissimilarity matrix $\bar{S}$ for each subgraph which can be fed into MDS. Supposed we have a set of n fingerprints and we are able to estimate the dissimilarity $\bar{s}_{ij}$ between all pairs of fingerprints i and j, MDS finds a configuration represented by a matrix $\mathbf{X}$ of size $n \times m$ where the entries represent the positions of the n fingerprints in m dimensions. So x_{ia} represents the relative position (or coordinate) of fingerprint i in dimension a. Hence, the output of the MDS method is $\mathbf{X}$ with

$$\mathbf{X} = \begin{bmatrix} x_{11} & \dots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nm} \end{bmatrix}_{[n \times m]}, \quad (4)$$

such that the euclidean distance d_{ij} for any two points is:

$$d_{ij} = \sqrt{\sum_{a=1}^{m} \left(x_{ia} - x_{ja}\right)^2}. \quad (5)$$

As this is an optimization problem and d_{ij} is an estimation, an error of estimation e_{ij} is introduced with

$$e_{ij}^2 = \left(d_{ij} - \bar{s}_{ij}\right)^2, \quad (6)$$

and $\bar{s}_{ij}$ being the dissimilarity, and d_{ij} the euclidean distance in the MDS representation. Averaging e_{ij}^2 over all pairs gives a measure of the error σ_r for the entire MDS representation, called *Raw Stress* [13]. MDS tries to find a configuration $\mathbf{X}$ which minimizes σ_r:

$$\sigma_r(\mathbf{X}) = \min\left(\sum_{i<j} e_{ij}^2\right) = \min\left(\sum_{i<j} (d_{ij} - \bar{s}_{ij})^2\right). \quad (7)$$

An MDS embedding is performed for every subgraph obtained during the pruning process.

d) Anchoring of the MDS output to geographical coordinates: The position of the fingerprints estimated by MDS are relative positions in an arbitrary two-dimensional space. A transformation has to be applied to map the MDS topology to geographical coordinates. Knowing the geographical position of at least three fingerprints included in the MDS topology, such a transformation can be found. Hence, the output of the MDS method is passed through an anchoring process to obtain a transformation into geographic locations. The anchoring process is a regression problem. Our method is comparable to the approach presented in [13]. When all subgraphs have been passed through the anchoring process, a global representation is obtained. We now have a reference topology where all fingerprints are assigned to a geographical location.

B. Fingerprint localization and updating the reference topology

The obtained reference topology can be used for locating new fingerprints, i.e. fingerprints which were not located in the initial topology e.g. coming from new positioning inquiries. To locate such a new fingerprint, the similarity between a new fingerprint and each fingerprint in the current topology is determined. The subset of all fingerprints which yield a similarity value greater than θ with the new fingerprint is selected and the same process as described previously of calculating the dissimilarity matrix, applying MDS and anchoring the map is performed for this subset. The result is a location estimation of the fingerprint. Additionally, this fingerprint can now be included into the reference database and help the topology map to grow. However, we only add the new fingerprint to our reference database if the distortion of the topology is low. To evaluate this, we determine the fingerprint's influence on the existing topology by calculating the average displacement of the nodes in the subgraph before and after the insertion of the new fingerprint using the *Haversine Formula* [21] with Equation 8:

$$\overline{e} = \frac{1}{n}\sum_{i=1}^{n} distance\left([lat_i, long_i], [lat_{i'}, long_{i'}]\right). \quad (8)$$

Hereby $[lat_i, long_i]$ and $[lat_{i'}, long_{i'}]$ are the current and proposed locations of fingerprint i in the subgraph, respectively. If $\overline{e} < \lambda$ with λ a given error threshold in meters, the current topology map is updated adding the newly located fingerprint to it. Figure 2 summarizes the process.

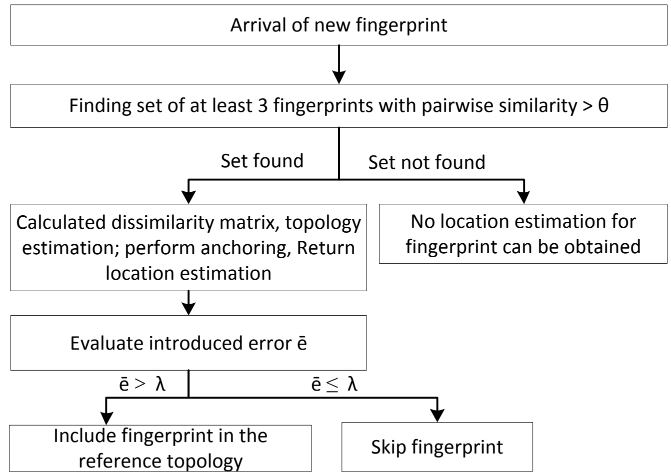


Fig. 2: Operation scheme to locate a new fingerprint and to consider its inclusion into the reference topology.

IV. Data Set

A real-world data set recorded in the city center of Zurich was used to verify our approach. Three *Nexus One* Android phones worn in a users' pocket were used. Each phone periodically recorded the ID and received signal strength of all visible WiFi hotspots and GSM cell towers together with a timestamp. Additionally, the phones recorded GPS location information which will serve as ground truth in our evaluation. A sampling rate of $0.2Hz$ was selected which resulted in the successful recording of 2576 fingerprints. The experiment was conducted during day time and hence, normal conditions as the influence of non-permanent obstacles such as pedestrians, and trams or buses passing by were present. Figure 3 shows a heat map generated from the locations where the fingerprints were recorded (based on GPS information). The recording space covers an area of $0.795km^2$. Our data set has a WiFi space of 2028 dimensions i.e. APs, and a GSM space of 66 dimensions i.e. GSM cells. That means that the density of WiFi APs is 30.73 APs per GSM cell. In terms of area units, and considering GSM cells as circular areas of $200m^2$ size, this would correspond to 0.15 APs per square meter. As we will compare the MDS-based location estimation against the GPS samples to evaluate the approach, we have to understand the accuracy of the GPS-provided location estimations. The median accuracy of the GPS samples is $12.5m$. Hereby, the accuracy value defines the 95% confidence circle.

V. Evaluation

In this section we first evaluate the relationship between the similarity measure and the recording distance between two fingerprints. Afterwards, we investigate the accuracy of our MDS-based topology estimation. Hereby, we investigate the influence of different pruning thresholds θ. For each evaluation step, we compare the approach with fingerprints generated i) with WiFi information, ii)

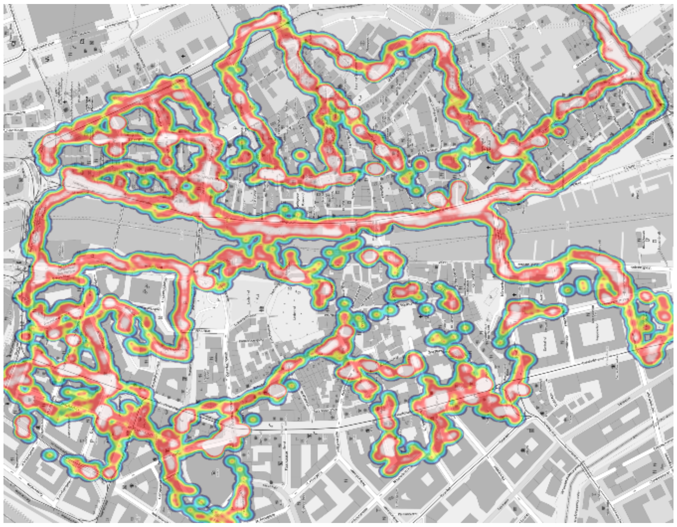


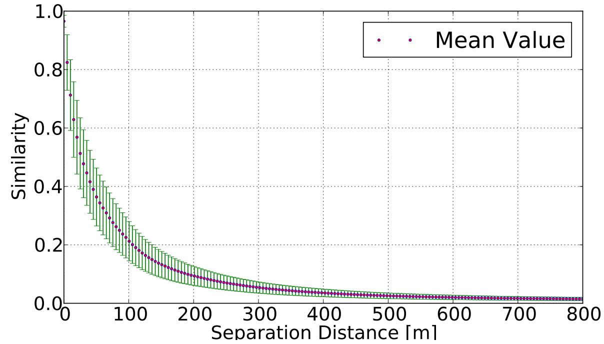
Fig. 3: Distribution of the recorded GPS-referenced WiFi+GSM fingerprints data set in the city center of Zurich.

GSM information, and iii) WiFi+GSM information. As our approach is designed for a collaborative system which gradually grows as people are using it, we further evaluate the evolution of the location accuracy as new fingerprints are added. We use GPS information as ground truth for the evaluation.

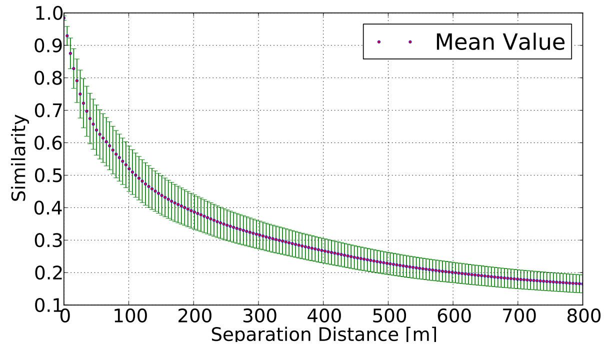
A. Fingerprint Similarity vs. Distance

We evaluate the relationship between the Tanimoto similarity measure of two fingerprints and the distance between their recordings. To do so, we calculate similarity values between all fingerprint pairs in our data set The relation between similarity measure and distance is illustrated in Figure 4 for the three different fingerprint sets WiFi, GSM and WiFi+GSM. The plots show for each distance value (obtained from GPS information) the mean similarity value together with the variance. The relationship follows a non-linear, monotonic decaying curve. The flattening for smaller similarities or larger distances, respectively, causes an increased error rate in the distance estimation by given similarity due to the non-negligible influence of the variance. For example a similarity below 0.1 can be found at any point beyond $200m$. Thus, no clear discrimination of distances is possible in the low similarity range. This effect is less influential for large similarities or small distances, respectively.

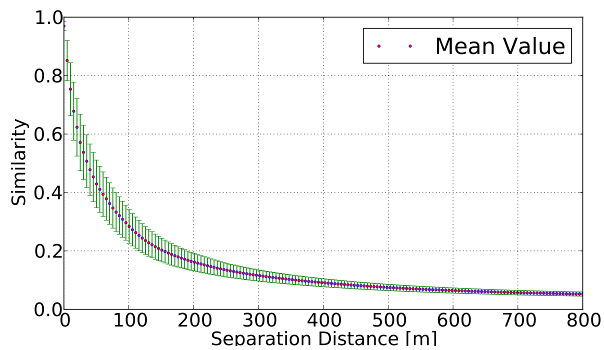
By comparing the three relations, the figures show that GSM has the highest variance. WiFi presents a steeper slope in the low distance range than WiFi+GSM which is required for good discrimination. However, WiFi+GSM has a lower variance by providing a larger discrimination range for distances than WiFi and is thus favored.



(a)



(b)



(c)

Fig. 4: Mean and variance valuess for Tanimoto Similarity Measures. a) WiFi, b) GSM, c) WiFi+GSM

B. Topology Estimation

We are now going to evaluate the localization accuracy of our approach. Table II to Table IV list the localization accuracy together with additional parameters. Table I gives a description of the parameters. θ is the pruning parameter as introduced previously. For each threshold, we run the localization process 100 times with random starting configurations. $\widetilde{e}$ is the median localization error in comparison to the GPS ground truth information. α and β are the 25% and 75% error quantiles in meters, respectively. σ is the variance of the median error for these iterations. δ represents the percentage of fingerprints out

TABLE I: Overview of evaluated parameters

Parameter	Description
θ	Pruning parameter, $0 \leq \theta \leq 1$
$\widetilde{e}$	median localization error [m]
α	25% error quantiles [m]
β	75% error quantiles [m]
σ	Variance of the median error
δ	ratio of localized fingerprints [%]
ρ	ratio of anchor points [%]

TABLE II: Summary of the algorithm performance for different thresholds θ using WiFi-based fingerprints.

θ	$\widetilde{e}\,[m]$	$\alpha\,[m]$	$\beta\,[m]$	σ	$\delta\,[\%]$	$\rho\,[\%]$
0.6	17	9	31	0.6	29	2.9
0.5	42	18	78	0.4	40	4.2
0.4	33	13	54	0.7	55	8.3
0.3	88	28	175	0.4	72	9.2
0.2	377	190	390	0.6	93	2.2
0.1	383	256	525	0.8	96	0.7
0.0	431	324	615	0.6	100	0.3

TABLE III: Summary of the algorithm performance for different thresholds θ using GSM-based fingerprints.

θ	$\widetilde{e}\,[m]$	$\alpha\,[m]$	$\beta\,[m]$	σ	$\delta\,[\%]$	$\rho\,[\%]$
0.6	286	161	708	0.16	94	1.8
0.5	201	130	309	0.06	98	0.3
0.4	273	160	397	0.08	99	0.3
0.3	300	191	605	0.12	99	0.3
0.2	466	259	859	0.11	100	0.3
0.1	577	389	733	0.09	100	0.3
0.0	640	483	893	0.13	100	0.3

TABLE IV: Summary of the algorithm performance for different thresholds θ using WiFi+GSM-based fingerprints.

θ	$\widetilde{e}\,[m]$	$\alpha\,[m]$	$\beta\,[m]$	σ	$\delta\,[\%]$	$\rho\,[\%]$
0.6	26	10	44	1.6	34	7.1
0.5	**30**	**14**	**57**	**1.4**	**60**	**7.6**
0.4	56	18	114	1.7	80	10.6
0.3	201	84	316	1.4	95	5.5
0.2	366	211	511	1.6	99	0.3
0.1	264	158	396	1.8	100	0.3
0.0	574	457	640	1.5	100	0.3

of the data set that were localized (and hence not pruned). ρ represents the percentage of fingerprints out of the data set that were used as anchor points.

Let us now have a closer look at some of the obtained results. Generally, we obtain better results (= lower $\widetilde{e}$) by considering WiFi+GSM fingerprints compared to using only WiFi or only GSM. By setting $\theta = 0$, the similarity graph is not pruned. Figure 5a shows the reference topology results from the WiFi+GSM fingerprints in blue together with the GPS ground truth in red. Ideally, the two graphs completely overlap. This, however, is not the case. Figure 5b shows a histogram of the corresponding error distribution. The median localization accuracy is $574m$. By increasing θ, the similarity graph is being pruned. Figure 6 shows the MDS-based topology reconstruction by applying a pruning threshold $\theta = 0.5$ on the WiFi+GSM fingerprints. As listed in Table IV, of WiFi+GSM, only $\delta = 60\%$ of the fingerprints can be used for the reference topology while the rest of the fingerprints do not fulfill the required similarity criteria. However, the median accuracy is now $30m$. With a pruning threshold $\theta = 0.6$, we achieve a median accuracy of $26m$ while being able to localize 34% of the fingerprints. With this, we see that by removing low similarity values we are not able to locate all fingerprints anymore but, on the other hand, the localization accuracy increases significantly. Hence, our method can automatically detect fingerprints which can not reliably be located and for the remaining provide a location estimation with an accuracy in a similar range as related work [3]. Only 8% of all fingerprints in the reference topology need to be geo-referenced. This is far less than the 100% required in state-of-the-art systems.

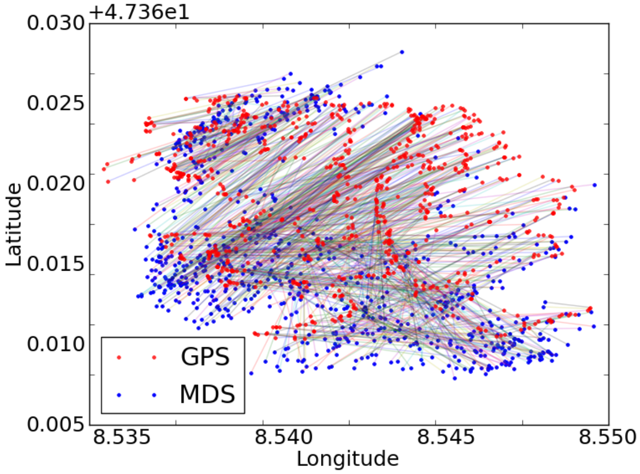
C. Evolution of the reference topology

Our approach fits a collaborative approach where the localization estimation starts with a few fingerprints and gradually grows by adding new ones. Hereby, at the beginning, when only a few data points are present, the provided localization is expected to be rather inaccurate or a localization is not possible at all as the majority of similarities stem from long distance measures and hence get pruned. However, gradually, we expect a denser sampling of the region resulting in smaller distances between fingerprints and thus larger similarity values can be expected which remain during the pruning step. With this, we expect the localization method to provide more accurate results over time. To investigate this behavior, we observe the relation between median error rate and the number of considered samples by adding samples. We start with a minimal set of three fingerprints and gradually add new ones. Only fingerprints that can be localized are considered. Figure 7 shows that the obtained result follows the expected trend that the localization error decreases by gradually adding new fingerprints arrive. The dotted red line represents the median location accuracy obtained by Place Lab [3]. We see a convergence towards a comparable error rate.

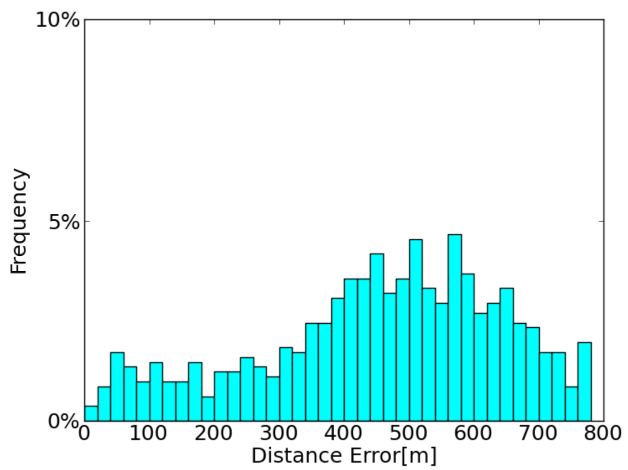
VI. Discussion

This work presents a fingerprinting method for localizing mobile devices in urban spaces using MDS-based embedding of WiFi+GSM fingerprints to obtain a reference topology. The novelty of our method is threefold:

- Only a fraction of the training set's fingerprints needs to be geo-referenced. This allows to include fingerprints into reference databases also in the absence

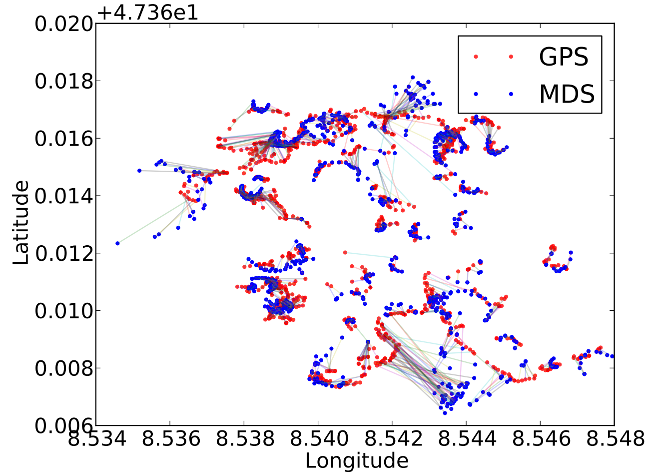


(a)

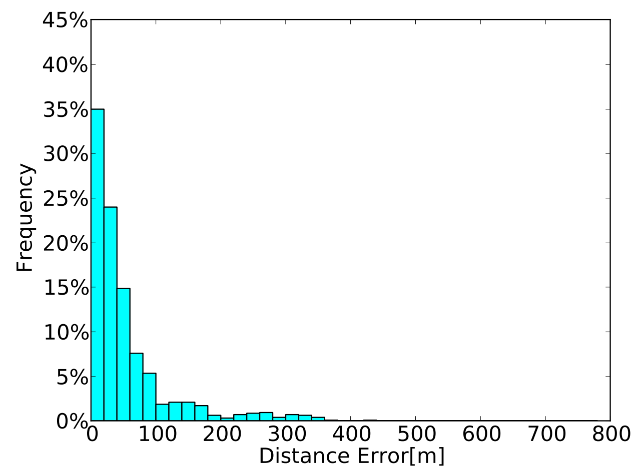


(b)

Fig. 5: Results from the topology reconstruction with no pruning ($\theta = 0.0$). a) map of the reconstructed topology using WiFi+GSM fingerprints (blue) and their actual location of recording from GPS information (red). b) Histogram of the localization error by comparing the estimated location of fingerprints to their actual position.



(a)



(b)

Fig. 6: Results from the topology reconstruction with pruning ($\theta = 0.5$). a) map of the reconstructed topology using WiFi+GSM fingerprints (blue) and their actual location of recording from GPS information (red). b) Histogram of the localization error by comparing the estimated location of fingerprints to their actual position.

of GPS reception and does not require a manual labeling.

- By removing low similarity values, increased robustness against multipath, shadow fading and other influences that affect similarity estimations can be provided.
- The method is ideal for a collaborative approach: Users provide a fingerprint to receive a location estimation. Simultaneously, this fingerprint can be used to refine and extend the topology estimation. Hence, we can gradually increase the covered space without requiring further efforts by the users.

Our evaluation shows that by increasing the pruning threshold θ, more fingerprints are discarded and cannot be located. However, for the remaining fingerprints, the accuracy of the localization increases. For $\theta = 0.5$, our method could locate 70% of the fingerprints with a median error of $30m$. Only 8% of the fingerprints were geo-referenced and the rest could be positioned witout any corresponding location informat but only considering their similarity. We further show with our data set that the localization error decreases as new fingerprints are added and converges to an accuracy comparable to related work. The reason that a fingerprint cannot be localized is that there are not enough similar fingerprints to be found. A dense, uniform sampling

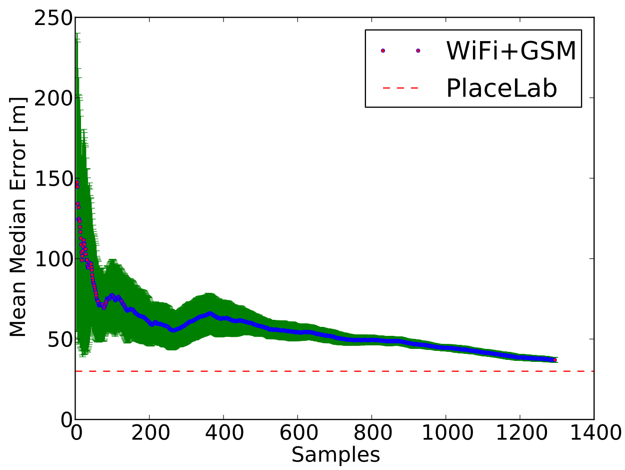


Fig. 7: Evolution of the localization error by gradually adding new fingerprints to the reference topology. Threshold $\theta = 0.5$

TABLE V: Summary of various performances of localization algorithms.

Ref	ω	$\widetilde{e}\,[m]$	$\varphi\,[\%]$	τ
DactyLoc	WiFi+GSM	20-30	7	Urban
2011 [18]	WiFi	1.5	8.7	Indoor
2010 [8]	GSM	SoA	100	Indoor
2010 [20]	WiFi+acc	P.L.	100	Urban
2008 [12]	WiFi	R.L.	100	Indoor
2007 [22]	GSM	5	100	Indoor
2005 [3]	WiFi	20-30	100	Urban
2000 [4]	WiFi	8	100	Indoor

of the space would increase the ratio of fingerprints that can be localized. Our system is designed in a way that this ratio increases while using the system. Further, a minimal density of WiFi access points in an urban area is required so that signal from different networks overlap. The density of access points in our experiment area was on average 1.5 access points every $10m^2$ circular area. We expect this number to be reasonable for many urban areas and indoor venues and hence, comparable results can be expected.

Table V provides a comparison of our method to different existing related positioning methods. Columns present following information: year and reference to the related work, type of used information ω, reported median accuracy $\widetilde{e}$, percentage of geo-referenced fingerprints in the training set φ, area of application τ. SoA states for State-Of-Art comparable accuracy, $R.L.$ for Room-Level, and $P.L.$ for Place-Level accuracy.

As Table V shows, related literature reports that the achievable median accuracy of fingerprinting approaches yields under $10m$ for indoors, and $20 - 30m$ for urban spaces. However, the accuracy of WiFi positioning tends to vary from indoor to urban environments. While the average accuracy is typically within few meters of the actual position for indoors, sudden jumps in the estimated positions are possible [23]. In section V we presented an overview of the trade-off between the number of geo-referenced fingerprints (and thus the deployment cost), and the achievable accuracy, which according to [18] is an urgent topic of research for such approaches. The introduction of the pruning threshold θ makes possible to handle this trade-off.

For a real-time implementation, the computational complexity is a key factor: At the core of MDS is an eigen-decomposition on an $n \times n$ symmetric matrix which for classic MDS takes $O\left(n^3\right)$ time. However, it can be reduced to $O\left(n \lg n\right)$ steps and easily parallelized for use with large datasets [24]. When a new fingerprint arrives, it takes $O\left(C \cdot m \cdot n + m(n+1)^3\right)$ time for our algorithm to generate location results. Hereby, n is the number of closest fingerprints, m the number of anchor points, usually $m = 3$, and C the cost of computing and accessing each entry of the dissimilarity matrix built with the new and the closest fingerprints. We expect the method to be scalable to also work with large data sets.

Our approach faces similar limitations as traditional fingerprinting methods have: If reference data in an area is missing, our method is not able to determine a location. However, our approach has the advantage to be able to provide location estimations in regions where GPS information is either unreliable or not present at all and hence ideal for urban spaces and indoor venues. We see a promising application of our method by combining it with existing systems such as Place Lab [3] to extend their functionality into areas where a GPS-based indexing is not possible. GPS-referenced fingerprints obtained in regions with good reception can serve as anchor points. With our method the covered space can gradually grow as people are using the system without the requirement of manual labeling of fingerprints.

VII. Acknowledgment

This work is supported under the FP7 ICT FET Programme, grant agreement No 231288 (SOCIONICAL).

References

[1] A. Küpper, *Location–based Services — Fundamentals and Operation*. John Wiley & Sons, 2005.

[2] M. B. Kjærgaard, H. Blunck *et al.*, “Indoor positioning using gps revisited,” in *Proc. of the 8th International Conference on Pervasive Computing*, 2010.

[3] A. LaMarca, Y. Chawathe *et al.*, “Place lab: Device positioning using radio beacons in the wild,” *Pervasive Computing*, 2005.

[4] P. Bahl and V. Padmanabhan, “Radar: An in-building rf-based user location and tracking system,” in *Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2. IEEE, 2000.

[5] Y. Cheng, Y. Chawathe *et al.*, “Accuracy characterization for metropolitan-scale wi-fi localization,” in *Proc. of the 3rd international conference on Mobile systems, applications, and services*. ACM, 2005.

[6] A. W. Tsui, W.-C. Lin, W.-J. Chen, P. Huang, and H.-H. Chu, “Accuracy performance analysis between war driving and war walking in metropolitan wi-fi localization,” *Trans. Mob. Comput.*, vol. 9, no. 11, 2010.

[7] G. Chandrasekaran, M. A. Ergin, M. Gruteser, and R. P. Martin, "Bootstrapping a location service through geocoded postal addresses," in *Proc. of the Third International Symposium on Location- and Context-Awareness*, 2007.

[8] P. Nurmi, S. Bhattacharya, and J. Kukkonen, "A grid-based algorithm for on-device gsm positioning," in *Proc. of the 12th Int. Conf. on Ubiquitous Computing*. ACM, 2010.

[9] Ekahu, "Ekahau positioning engine," ekahau.com, 2011, [Online; accessed 25-July-2012].

[10] O. Woodman and R. Harle, "Rf-based initialisation for inertial pedestrian tracking," *Pervasive Computing*, 2009.

[11] J. Ledlie, J. Park *et al.*, "Molé: a scalable, user-generated wifi positioning engine," in *Proc. of the second International Conference on Indoor Positioning and Indoor Navigation*, 2011.

[12] P. Bolliger, "Redpin-adaptive, zero-configuration indoor localization through user collaboration," in *Proc. of the 1st int. workshop on Mobile entity localization and tracking in GPS-less environments*. ACM, 2008.

[13] I. Borg and P. Groenen, *Modern Multidimensional Scaling: Theory and applications*. Springer Verlag, 2005.

[14] Y. Shang, W. Ruml, Y. Zhang, and M. Fromherz, "Localization from mere connectivity," in *Proc. of the 4th ACM international symposium on Mobile ad hoc networking & computing*. ACM, 2003.

[15] Y. Shang and W. Ruml, "Improved mds-based localization," in *23rd Annual Joint Conference of the IEEE Computer and Communications Societies*. IEEE, 2004.

[16] B. Wei, W. Chen, and X. Ding, "Advanced mds based localization algorithm for location based services in wireless sensor network," in *Ubiquitous Positioning Indoor Navigation and Location Based Service*, oct. 2010.

[17] J. Koo and H. Cha, "Autonomous construction of a wifi access point map using multidimensional scaling," *Pervasive Computing*, 2011.

[18] T. Pulkkinen, T. Roos, and P. Myllymäki, "Semi-supervised learning for wlan positioning," *Artificial Neural Networks and Machine Learning*, 2011.

[19] P. Jaccard, "The distribution of the flora in the alpine zone," *New Phytologist*, vol. 11, no. 2, 1912.

[20] D. Kim, Y. Kim, D. Estrin, and M. Srivastava, "Sensloc: Sensing everyday places and paths using less energy," in *Proc. of the 8th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2010.

[21] J. Montavont and T. Noel, "Ieee 802.11 handovers assisted by gps information," in *Int. Conf. on Wireless and Mobile Computing, Networking and Communications*, 2006.

[22] V. Otsason, A. Varshavsky, A. LaMarca, and E. de Lara, "Accurate gsm indoor localization," *Journal of Pervasive and Mobile Computing*, vol. 3, no. 6, pp. 698–720, December 2007.

[23] T. Pulkkinen, S. Bhattacharya, and P. Nurmi, "Abstracting positions for indoor location based services," in *IUI Workshop on Location Awareness for Mixed and Dual Reality*, 2011.

[24] T. Yang, J. Liu *et al.*, "A fast approximation to multidimensional scaling," in *Proc. of Workshop on Computation Intensive Methods for Computer Vision*, 2006.