Wi-Fi Location Fingerprinting Using an Intelligent Checkpoint Sequence

Guenther Retscher and Hannes Hofer

Department of Geodesy and Geoinformation
TU Wien
Austria

Presented at the FIG Working Week 2016, May 2-6, 2016 in Christchurch, New Zealand
Feature-Based Positioning

Position can be derived by

- Measuring a spatially varying feature (such as RSS)
- Locating the measured value within a database of georeferenced values (reference values)

Typical radio map of one AP
Location Fingerprinting

Current measurements

\[ y = [s^1, s^2, s^3, \ldots] \]

DB

\[ P_1 : \tilde{y}_1 = [\tilde{s}_1^1, \tilde{s}_1^2, \tilde{s}_1^3, \ldots] \]

\[ P_2 : \tilde{y}_2 = [\tilde{s}_2^1, \tilde{s}_2^2, \tilde{s}_2^3, \ldots] \]

\[ P_3 : \tilde{y}_3 = [\tilde{s}_3^1, \tilde{s}_3^2, \tilde{s}_3^3, \ldots] \]

\[ \ldots \]

\[ \ldots \]

\[ ? P(t) \]
Intelligent Checkpoints iCPs

Sparsely distributed reference points in the area of interest

Waypoints along a route have to be passed following a logical sequence

Waypoints are called iCPs

Training measurements on chosen iCPs only

Twofold intelligent because of their intelligent selection and logical sequence along the path

Examples for iCPs in the foyer and staircase of the 3rd floor in a multi-storey office building
Research Questions

• How the iCPs must be chosen so that they are well distinguishable?
• Can it be recognized, when and whether an iCP is passed?
• How the trajectory can be continuously determined between these iCPs?
Evaluation Premises and Definitions

Matching rate \( MR = \frac{\text{number of correctly assigned RSS scans to RPs}}{\text{total number of all RSS scans in positioning phase}} \)

Case 1: -101 dBm for no scan value available in one epoch

\[
\text{Scan}_1 = \left[ S_{1\ AP_1}, S_{1\ AP_2}, \ldots, S_{1\ AP_n} \right] \quad S_{1\ AP_x} \in \left\{ \mathbb{Z}_{<0} \cap \{-100,-99,\ldots,-1\} \right\}
\]

\[
S_{1\ AP_x} = \begin{cases} 
\text{RSS APx in dBm} & \text{if RSS value to APx is obtained} \\
-101 \text{ dBm} & \text{if RSS value to APx not obtained}
\end{cases}
\]

Case 2: NaN (Not a Number) for no scan value available in one epoch

\[
\text{Scan}_2 = \left[ S_{2\ AP_1}, S_{2\ AP_2}, \ldots, S_{2\ AP_n} \right] \quad S_{2\ AP_x} \in \left\{ \mathbb{Z}_{<0} \cap \{-100,-99,\ldots,-1\} \cup \text{NaN} \right\}
\]

\[
S_{2\ AP_x} = \begin{cases} 
\text{RSS APx in dBm} & \text{if RSS value to APx is obtained} \\
\text{NaN} & \text{if RSS value to APx is not obtained}
\end{cases}
\]

where \( n \) is the number of APs given in the vector \( \text{AP}_x = [\text{AP}_1, \text{AP}_2, \ldots, \text{AP}_n] \)
Nearest Neighbour (NN) Algorithm

Euclidean distance $D$ is calculated for each AP in the positioning phase from the DB values:

$$D = \sqrt{(S_{m_{AP1}} - S_{i_{AP1}})^2 + \ldots + (S_{m_{APn}} - S_{i_{APn}})^2}$$

where $[S_{m_{AP1}}, S_{m_{AP2}}, \ldots, S_{m_{APn}}]$ is measured RSS vector for positioning and $[S_{i_{AP1}}, S_{i_{AP2}}, \ldots, S_{i_{APn}}]$ the reference for location $i$ in the used DB.

Allocation of positioning scans to training fingerprinting DB
Calculation Alternatives

Vector Wi-Fi scans:

\[
\begin{align*}
\text{Wi-Fi Scans} &= \left( \begin{array}{c}
P_{\text{Scan No.1}} \\
\vdots \\
P_{\text{Scan No.W}} \\
\end{array} \right) \\
\text{all Scans DB}_{\text{Scan1}} &= \left( \begin{array}{c}
S_{1,AP_1} \ldots S_{1,AP_n} \\
\vdots \\
S_{1,W,AP_1} \ldots S_{1,W,AP_n} \\
\end{array} \right)
\end{align*}
\]

where \([1, \ldots, W]\) is the number \(\text{No.}\) of all Wi-Fi Scans and all Scans \(\text{DB}_{\text{Scan1}}\) is the DB containing all scans \(S_{1,AP_1}\) to \(S_{1,W,AP_n}\)

Calculation of Euclidean distance leads to a distance vector with dimension \(1 \times W\)

The vector is sorted with the respective MATLAB function sort ()

For selection of position \(k\) minimum distances \(D_k = (d_{\text{min}_1}, d_{\text{min}_2}, \ldots, d_{\text{min}_k})\)

are used to find a single position with \(D\) Point \(ID_{S_k} = (ID_{\text{min}_1}, ID_{\text{min}_2}, \ldots, ID_{\text{min}_k})\)
Method 1: Most Frequent Values MFV

MATLAB function `mode()` is applied to vector Point $ID_s$.

$ID$ is selected which exists most frequently in the vector:

$$ID_{selected} = \text{mode} (D \text{ Point } ID_{S_k})$$

If two Point $ID_s$ exist in the vector the first one is selected which has the minimum Euclidean distance $D_k$. 

FIG Working Week Christchurch NZ 05.05.2016
Method 2: Likelihood Algorithm

Probability \( p_j \) for each distance value is calculated

The total probability \( p_{ID} \) for a certain position results from all probabilities \( p_i \) for this position in the form:

\[
p_j = \frac{d_{\text{min}j}^{-1}}{\sum_{i=1}^{k} d_{\text{min}i}^{-1}}
\]

where \( p_{\text{Point } ID} = \sum_{i=1}^{k} p_i \) for all Point \( ID_s \) in the vector\( D \) Point \( ID_{S_k} \) exist

Because the likelihood is higher the smaller the distance value is, these are inverted. \( D_k \) is the sum of the inverted distance values \( k \):

\[
D_k = \sum_{i=1}^{k} \frac{1}{D_{i : \text{PosID}}}
\]

Then the probability for the position Point \( ID \) is calculated as

\[
p_{\text{Point } ID} = \sum_{i=1}^{k} p_i \text{ and for every position in the vector a likelihood calculated}
\]
Outdoor Test Site

RSS scans were measured for the establishment of the fingerprinting DB in the training phase throughout the test site to be able to choose representative iCPs

23 reference points as candidates for iCPs
more than 200 APs
Consideration of all 4 Orientations

<table>
<thead>
<tr>
<th>scenarios</th>
<th>Test DB1 (-101dBm)</th>
<th>Test DB2 (NaN)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean DB</td>
<td>median DB</td>
</tr>
<tr>
<td>-------------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>SM1 DB</td>
<td>94,1%</td>
<td>67,5%</td>
</tr>
<tr>
<td>SM2 SB</td>
<td>95,7%</td>
<td>86,2%</td>
</tr>
<tr>
<td>joint DB</td>
<td>95,8%</td>
<td>59,7%</td>
</tr>
<tr>
<td>mean MR</td>
<td>95,2%</td>
<td>71,1%</td>
</tr>
<tr>
<td>SM1 DB</td>
<td>91,6%</td>
<td>92,0%</td>
</tr>
<tr>
<td>SM2 SB</td>
<td>92,9%</td>
<td>96,1%</td>
</tr>
<tr>
<td>joint DB</td>
<td>91,4%</td>
<td>88,2%</td>
</tr>
<tr>
<td>mean MR</td>
<td>92,0%</td>
<td>92,1%</td>
</tr>
</tbody>
</table>

Best matching rates are achieved if DB mean (-101 dBm) and test DB1 are combined.

Weighting favours more stable APs with less temporal variations which is usually the case in public spaces.
Matching rates are in average 4.9% higher over all scenarios and calculation variants
Main advantage is reduction of the number of RSS scans to be tested by a factor of 4
### Comparison MFV - Likelihood Alg.

<table>
<thead>
<tr>
<th>scenarios</th>
<th>MFV algorithm</th>
<th></th>
<th>Likelihood algorithm</th>
<th></th>
<th>mean MR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DB1(-101dBm)</td>
<td>DB2(NaN)</td>
<td>DB1(-101dBm)</td>
<td>DB2(NaN)</td>
<td></td>
</tr>
<tr>
<td>SM1 DB</td>
<td>93,0%</td>
<td>91,3%</td>
<td>92,7%</td>
<td>93,0%</td>
<td>92,50%</td>
</tr>
<tr>
<td>SM2 SB</td>
<td>95,4%</td>
<td>95,4%</td>
<td>95,7%</td>
<td>95,4%</td>
<td>95,48%</td>
</tr>
<tr>
<td>joint DB</td>
<td>95,4%</td>
<td>92,6%</td>
<td>95,6%</td>
<td>93,1%</td>
<td>94,18%</td>
</tr>
<tr>
<td>mean MR</td>
<td>94,6%</td>
<td>93,1%</td>
<td>94,7%</td>
<td>93,8%</td>
<td>94,05%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>scenarios</th>
<th>MFV algorithm</th>
<th></th>
<th>Likelihood algorithm</th>
<th></th>
<th>mean MR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DB1(-101dBm)</td>
<td>DB2(NaN)</td>
<td>DB1(-101dBm)</td>
<td>DB2(NaN)</td>
<td></td>
</tr>
<tr>
<td>SM1 DB</td>
<td>91,6%</td>
<td>90,6%</td>
<td>92,0%</td>
<td>90,9%</td>
<td>91,3%</td>
</tr>
<tr>
<td>SM2 SB</td>
<td>96,5%</td>
<td>94,0%</td>
<td>96,5%</td>
<td>94,3%</td>
<td>95,3%</td>
</tr>
<tr>
<td>joint DB</td>
<td>94,7%</td>
<td>91,7%</td>
<td>94,7%</td>
<td>91,9%</td>
<td>93,3%</td>
</tr>
<tr>
<td>mean MR</td>
<td>94,3%</td>
<td>92,1%</td>
<td>94,4%</td>
<td>92,4%</td>
<td>93,3%</td>
</tr>
</tbody>
</table>

Matching rates differ at most by around 1.1% compared to previous results.
Method which uses all RSS scans in the vectors leads to slightly lower MRs where combined DB falls short of only 0.2%.
No significant differences if heading is not or is considered.
Main advantage, however, is that number of operations to be carried out are reduced by around three quarter.
Highest MRs are achieved if DB1 is combined with mean DB (-101 dBm) 
Combination with DB1 containing the minimum values with fingerprinting DB leads to good results 
DB2 can be combined with fingerprinting DB without that MRs get substantially worse 
Best averaged MRs are 89.7% with DB2 and 76.4% with DB1
2 Different DBs with Known Orient.

Highest MRs have not changed, however, averaged MR is increased.
Combination of Test DB1 and DB with NaN values could achieve best MRs over 90%.
Reason is that the visibility of particular APs fluctuates less while scanning in a certain orientation.
Situation arises much less often if a RSS value is stored in fingerprinting DB but not measured in test scan in positioning phase.
Discussion of Main Results

- Between the calculation variants of different DBs no significant difference
- Arithmetic mean led in all tests to slightly better MRs
- Better results are obtained if all measured RSS scans are used for fingerprinting
- Better results if a single DB for a certain mobile is employed
- High MRs if the locations of the iCPs are selected in an intelligent manner, e.g. outdoors not to close to each other around building corners
- With use of magnetometer similar results but significant reduction of number and duration of calculation by three quarter
- Application of logical sequence between the iCPs is a simple attempt to reduce the number of possible user locations
- MR of 93 % if a site specific weighting vector is applied
- For more complex environments an advanced vector graph allocation can be applied and implemented
GPS Trajectory

mean deviation 10.1 m
iCP and INS Trajectory

mean deviation 3.0 m
Integration of Smartphone INS

Eucledian distances of a certain iCP calculated from continuous RSS scans while walking along a trajectory