TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Wi-Fi Location Fingerprinting Using an Intelligent Checkpoint Sequence

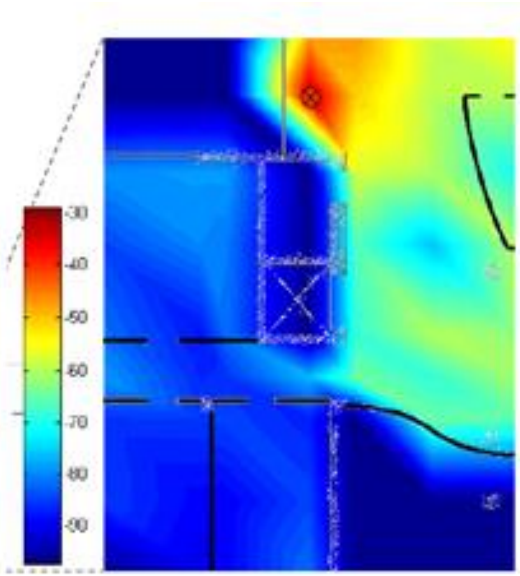
Guenther Retscher and Hannes Hofer

Department of Geodesy and Geoinformation
TU Wien
Austria

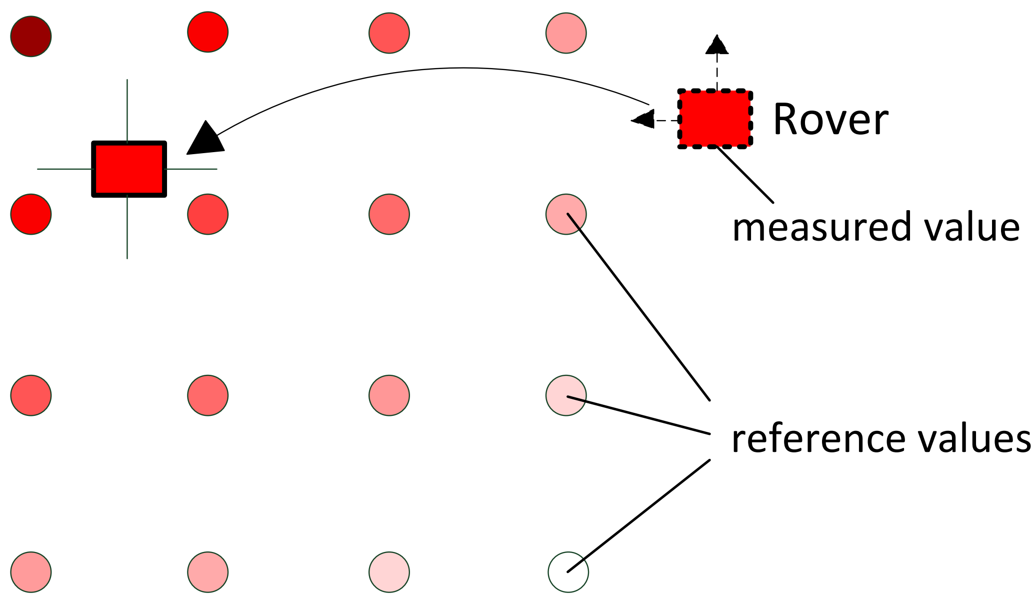
Feature-Based Positioning

Position can be derived by

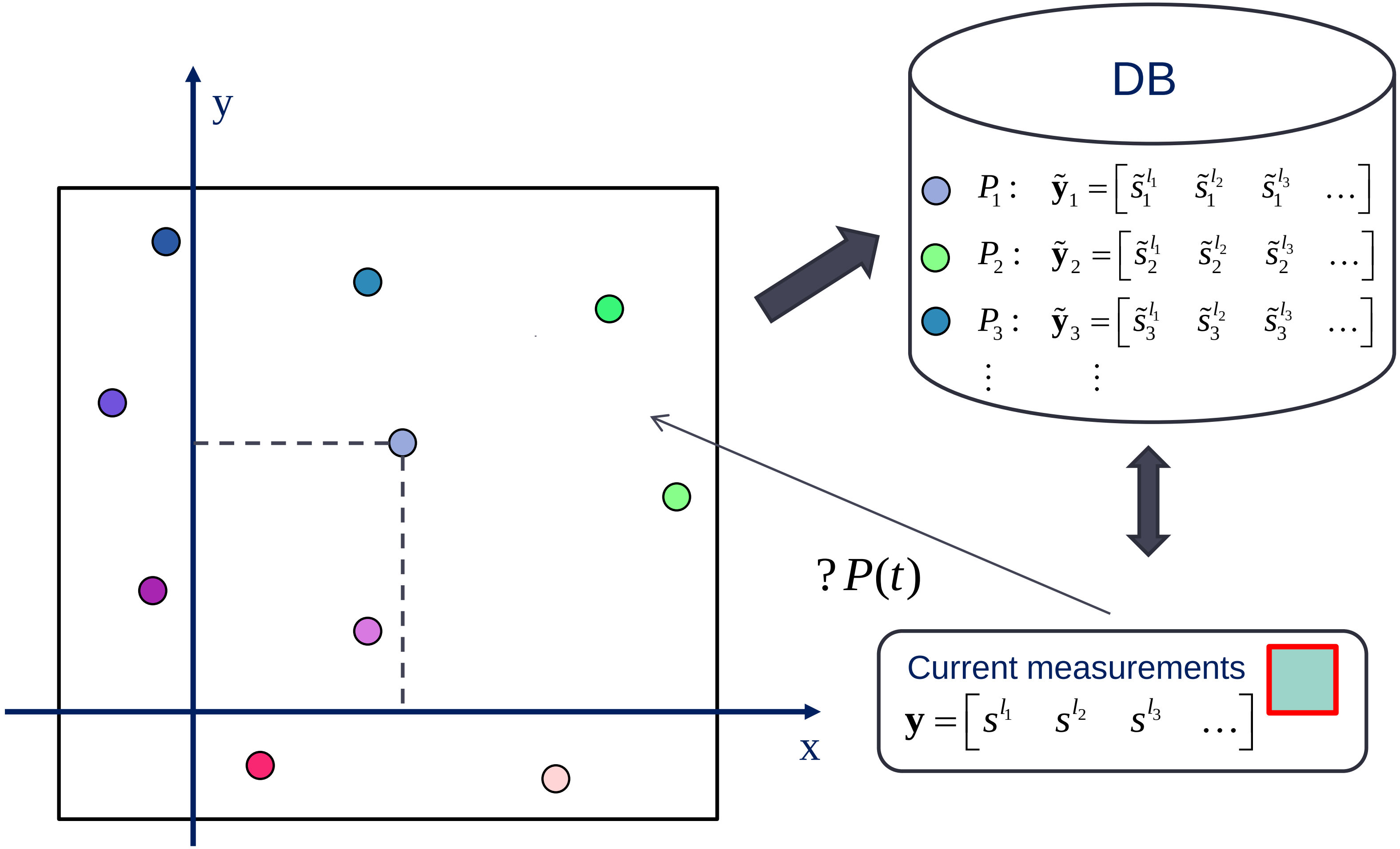
- Measuring a spatially varying feature (such as RSS)
- Locating the measured value within a database of georeferenced values (reference values)



Typical radio map of one AP



Location Fingerprinting



Intelligent Checkpoints iCPs

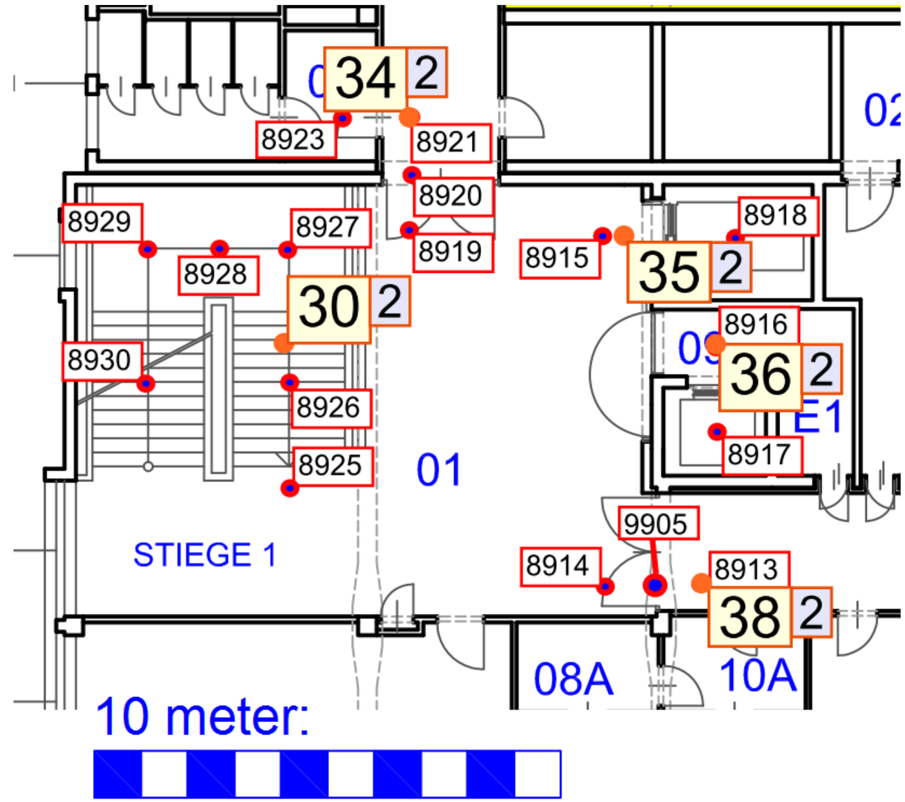
Sparsely distributed reference points in the area of interest

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

Waypoints are called iCPs

Training measurements on chosen iCPs only

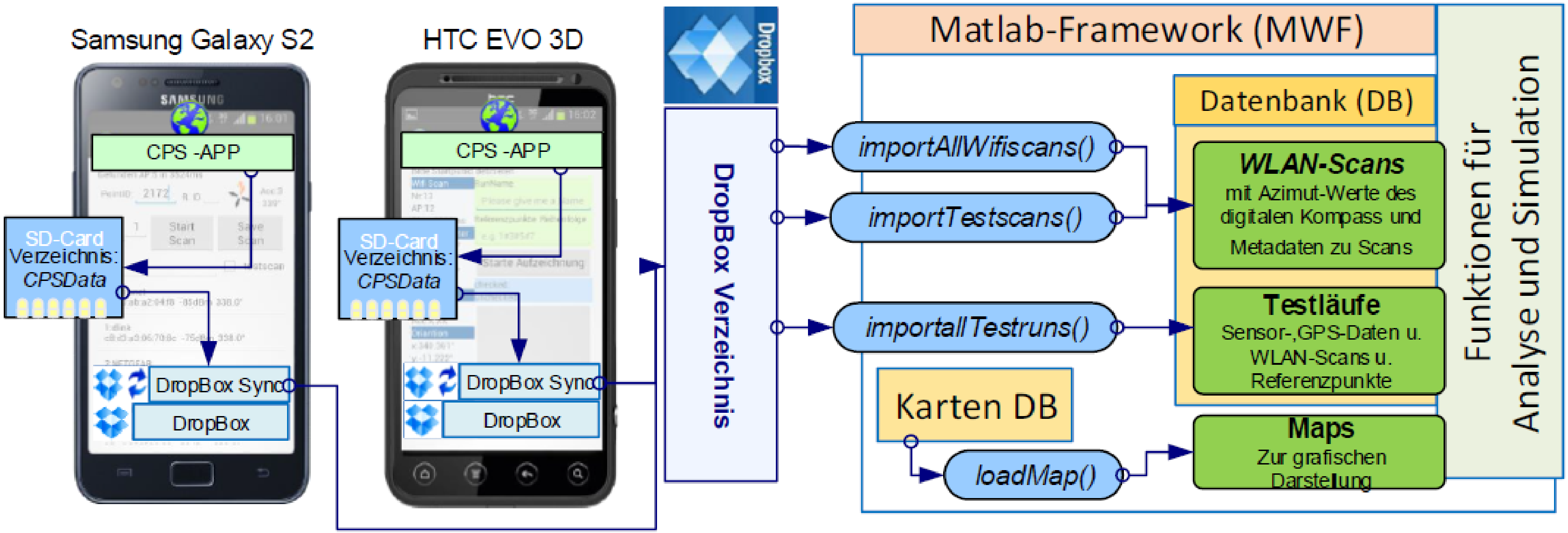
Twofold intelligent because of their intelligent selection and logical sequence along the path



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 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

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

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

$$\text{Scan1} = \left[S1_{AP1}, S1_{AP2}, \ldots, S1_{APn}\right] \quad S1_{APx} \in \left\{\mathbb{Z}_{<0} \cap \{-100, -99, \ldots, -1\}\right\}$$

$$S1_{APx} = \begin{cases} \text{RSS APx in dBm} & \text{if RSS value to APx is obtained} \\ \text{-101 dBm} & \text{if RSS value to APx not obtained} \end{cases}$$

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

$$\text{Scan2} = \left[S2_{AP1}, S2_{AP2}, \ldots, S2_{APn}\right] \quad S2_{APx} \in \left\{\mathbb{Z}_{<0} \cap \{-100, -99, \ldots, -1\} \cup \text{NaN}\right\}$$

$$S2_{APx} = \begin{cases} \text{RSS APx in dBm} & \text{if RSS value to APx is obtained} \\ \text{NaN} & \text{if RSS value to APx is not obtained} \end{cases}$$

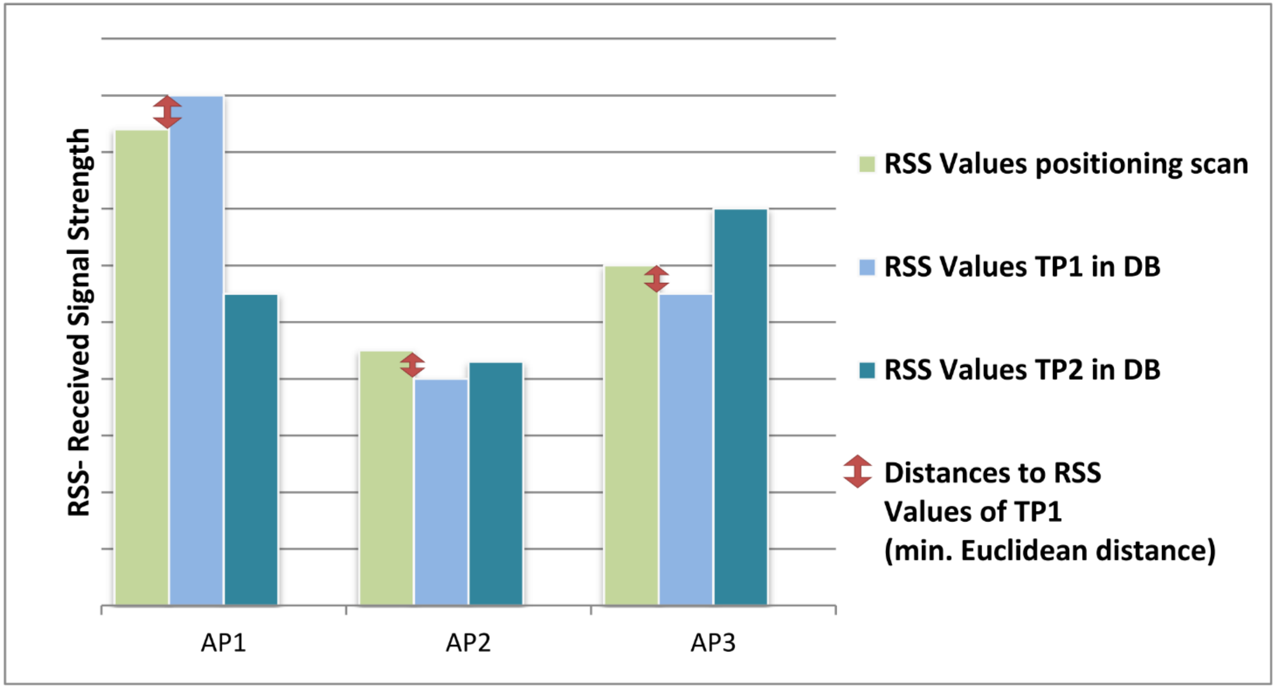
where *n* is the number of APs given in the vector $AP_x = [AP_1, AP_2, \ldots, AP_n]$

Nearest Neighbour (NN) Algorithm

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

$$D = \sqrt{\left(Sm_{AP1} - Si_{AP1}\right)^2 +,\ldots,+\left(Sm_{APn} - Si_{APn}\right)^2}$$

where $\left[Sm_{AP1}, Sm_{AP2}, \ldots, Sm_{APn}\right]$ is measured RSS vector for positioning and $\left[Si_{AP1}, Si_{AP2}, \ldots, Si_{APn}\right]$ the reference for location i in the used DB



Allocation of positioning scans to training fingerprinting DB

Calculation Alternatives

Vector Wi-Fi scans:

$$\text{Wi-Fi Scans} = \begin{pmatrix} P_{Scan\ No.1} \\ \vdots \\ P_{Scan\ No.W} \end{pmatrix} \qquad \text{all Scans DB}_{Scan1} = \begin{pmatrix} S1_{1,AP1} \cdots S1_{1,APn} \\ \vdots \quad \ddots \quad \vdots \\ S1_{W,AP1} \cdots S1_{W,APn} \end{pmatrix}$$

where [1, ... , W] is the number *No*. of all Wi-Fi Scans and all Scans DB_{Scan1} is the DB containing all scans $S1_{1,AP1}$ to $S1_{W,Apn}$

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 $D_k = \left(d_{\min_1}, d_{\min_2}, \ldots, d_{\min_k}\right)$

are used to find a single position with $D\ \text{Point}\ ID_{S_k} = \left(ID_{\min_1}, ID_{\min_2}, \ldots, ID_{\min_k}\right)$

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

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:

$$p_j = \frac{d_{\min j}^{-1}}{\sum_{i=1}^{k} d_{\min i}^{-1}}$$

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:

$$D_k = \sum_{i=1}^{k} \frac{1}{D_{i:PosID}}$$

Then the probability for the position Point ID is calculated as

$p_{\text{Point}\,ID} = \sum_{i=1}^{k} p_i$ and for every position in the vector a likelihood calculated

Outdoor Test Site

23 reference points as candidates for iCPs

more than 200 APs

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

Consideration of all 4 Orientations

scenarios	Test DB1 (-101dBm)			
	mean DB		median DB	
	-101dbm	*NaN*	*-101dbm*	*NaN*
SM1 DB	94,1%	67,5%	90,2%	66,8%
SM2 SB	95,7%	86,2%	92,6%	85,5%
joint DB	95,8%	59,7%	94,2%	55,8%
mean MR	95,2%	71,1%	92,3%	69,4%

scenarios	Test DB2 (NaN)			
	mean DB		median DB	
	-101dbm	*NaN*	*-101dbm*	*NaN*
SM1 DB	91,6%	92,0%	90,2%	92,3%
SM2 SB	92,9%	96,1%	90,4%	96,1%
joint DB	91,4%	88,2%	90,7%	88,6%
mean MR	92,0%	92,1%	90,4%	92,3%

Best matching rates are achived 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

Consideration of Heading

scenarios	Test DB1 (-101dBm)			
	mean DB		median DB	
	-101dbm	*NaN*	*-101dbm*	*NaN*
SM1 DB	92,7%	86,7%	90,9%	90,9%
SM2 SB	96,1%	92,6%	95,4%	95,4%
joint DB	95,2%	79,6%	94,5%	94,5%
mean MR	**94,7%**	**86,3%**	**93,6%**	**93,6%**

scenarios	Test DB2 (NaN)			
	mean DB		median DB	
	-101dbm	*NaN*	*-101dbm*	*NaN*
SM1 DB	88,1%	93,4%	84,6%	93,7%
SM2 SB	92,9%	94,3%	92,9%	94,3%
joint DB	91,2%	90,5%	89,4%	90,0%
mean MR	**90,7%**	**92,7%**	**89,0%**	**92,7%**

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.

scenarios	MFV algorithm		Likelihood algorithm		mean MR
	DB1(-101dBm)	DB2(NaN)	DB1(-101dBm)	DB2(NaN)	
SM1 DB	***93,0%***	91,3%	92,7%	***93,0%***	92,50%
SM2 SB	95,4%	95,4%	***95,7%***	95,4%	95,48%
joint DB	95,4%	92,6%	***95,6%***	93,1%	94,18%
mean MR	94,6%	93,1%	***94,7%***	93,8%	94,05%

scenarios	MFV algorithm		Likelihood algorithm		mean MR
	DB1(-101dBm)	DB2(NaN)	DB1(-101dBm)	DB2(NaN)	
SM1 DB	91,6%	90,6%	***92,0%***	90,9%	91,3%
SM2 SB	***96,5%***	94,0%	***96,5%***	94,3%	95,3%
joint DB	***94,7%***	91,7%	***94,7%***	91,9%	93,3%
mean MR	94,3%	92,1%	***94,4%***	92,4%	93,3%

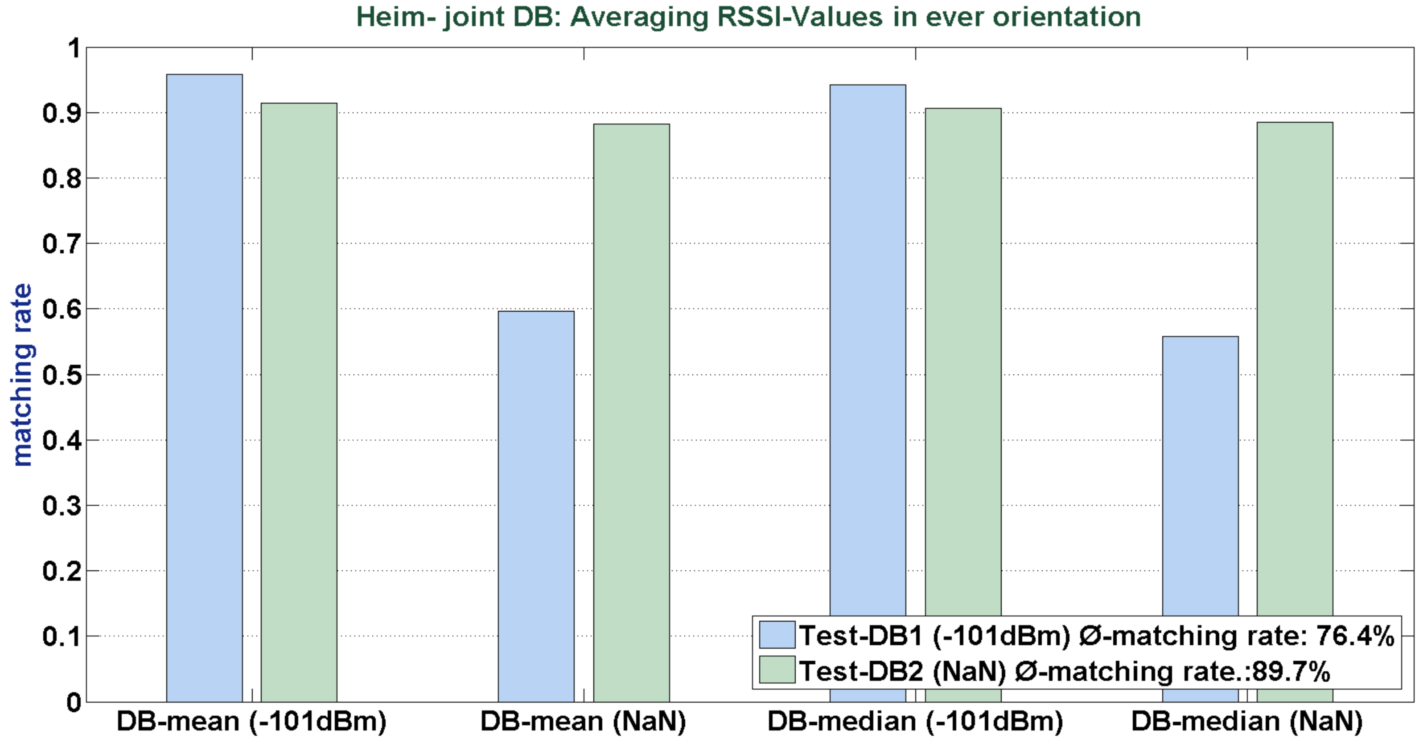
Matching rates differ at most by around 1.1% compared to previous results

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

2 Different DBs in 4 Orientations



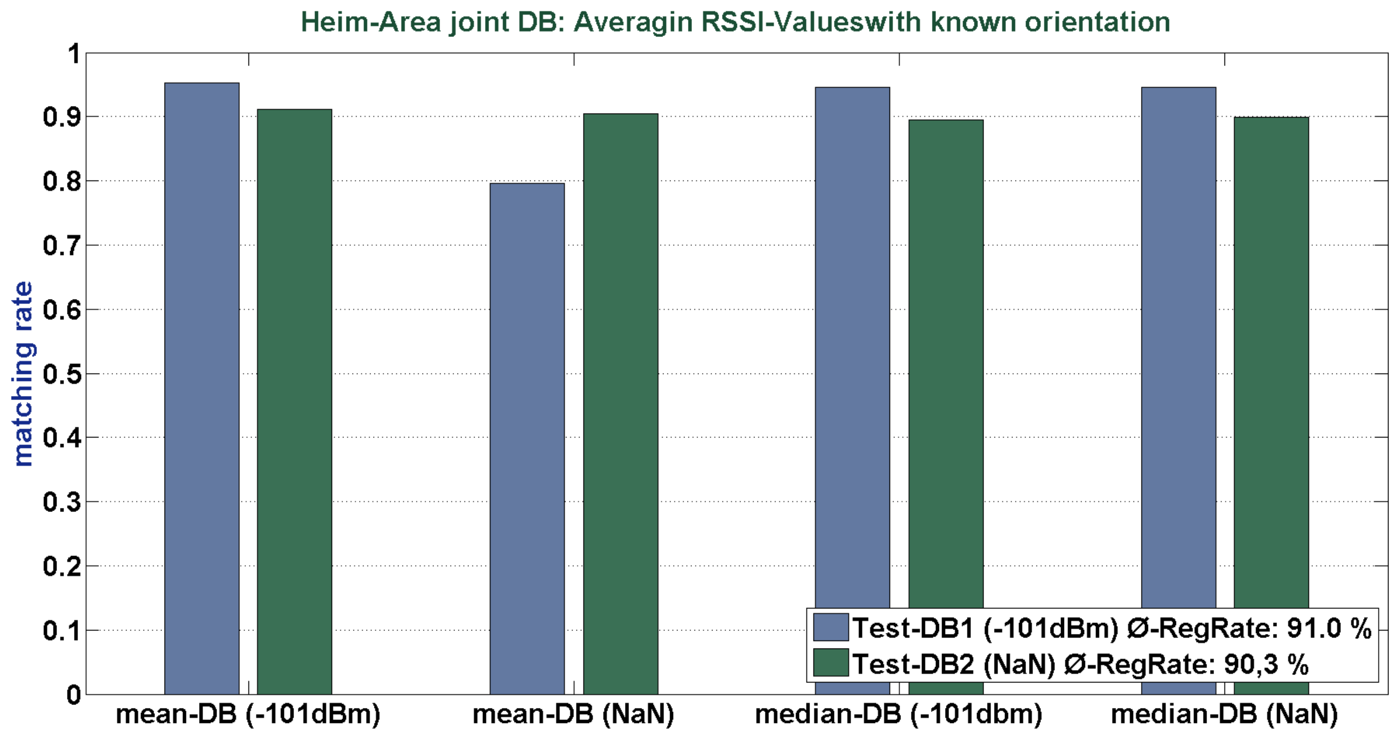
Highest MRs are achieved if DB1 is combined with mean DB (-101 dBm)

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%

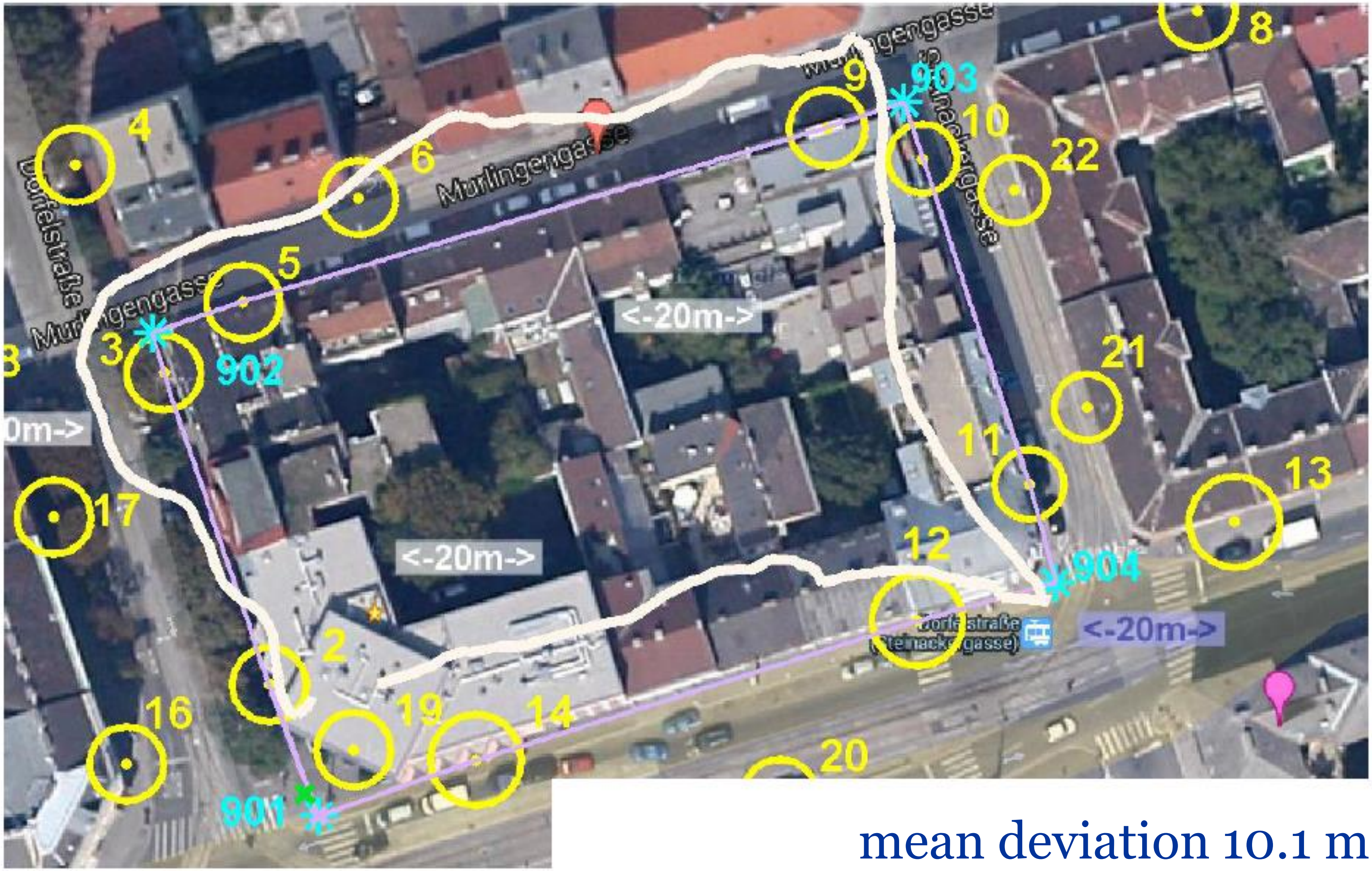
Reason is that the visibility of particular APs fluctuates less while scanning in a certain orientation

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

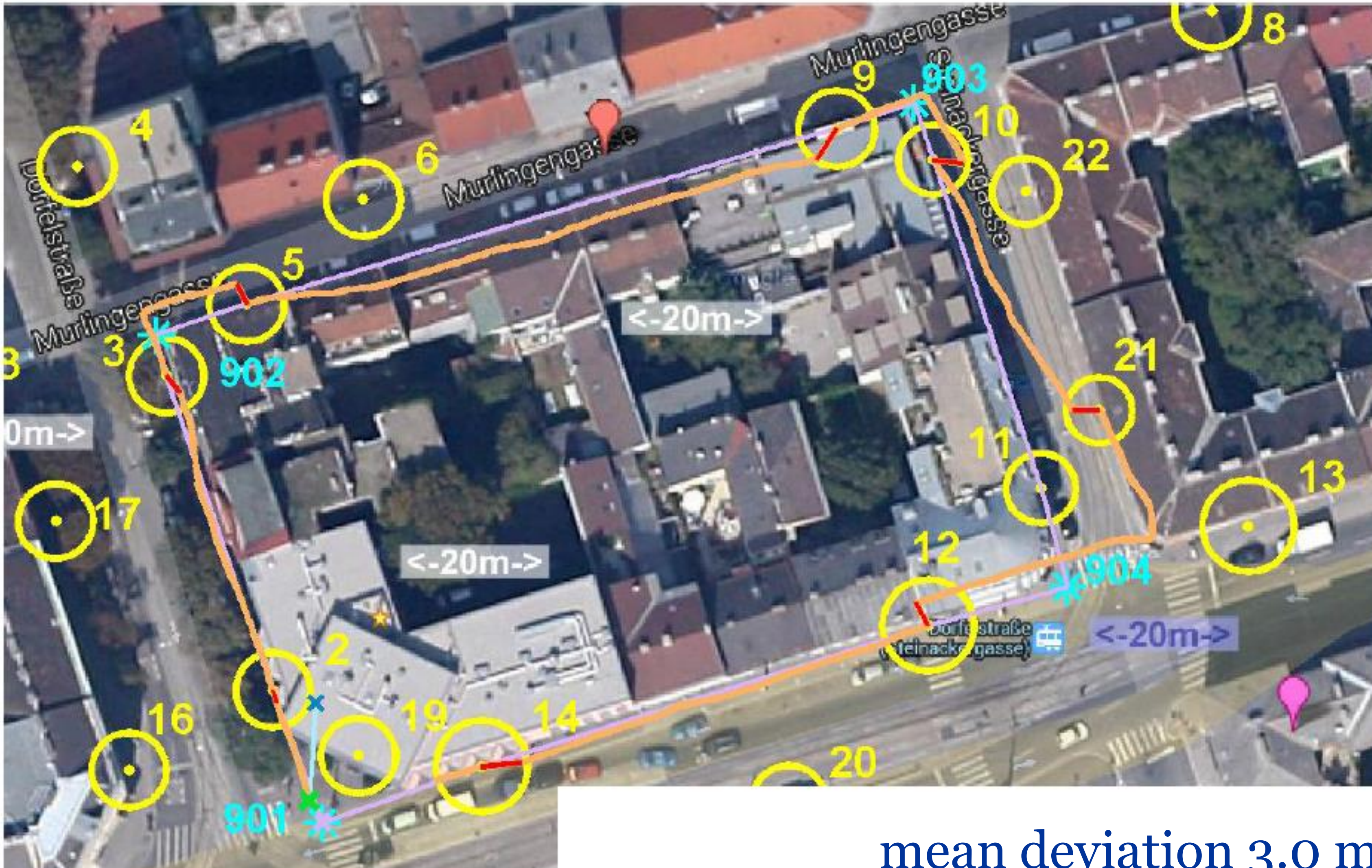
- Between the claculation variants of different DBs no significant difference
- Arithmetic mean led in all tests to slightly better MRs
- Better results are obtained if all measured RSS scans are used for fingerprinting
- Better results if a single DB for a certain mobile is employed
- High MRs if the 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



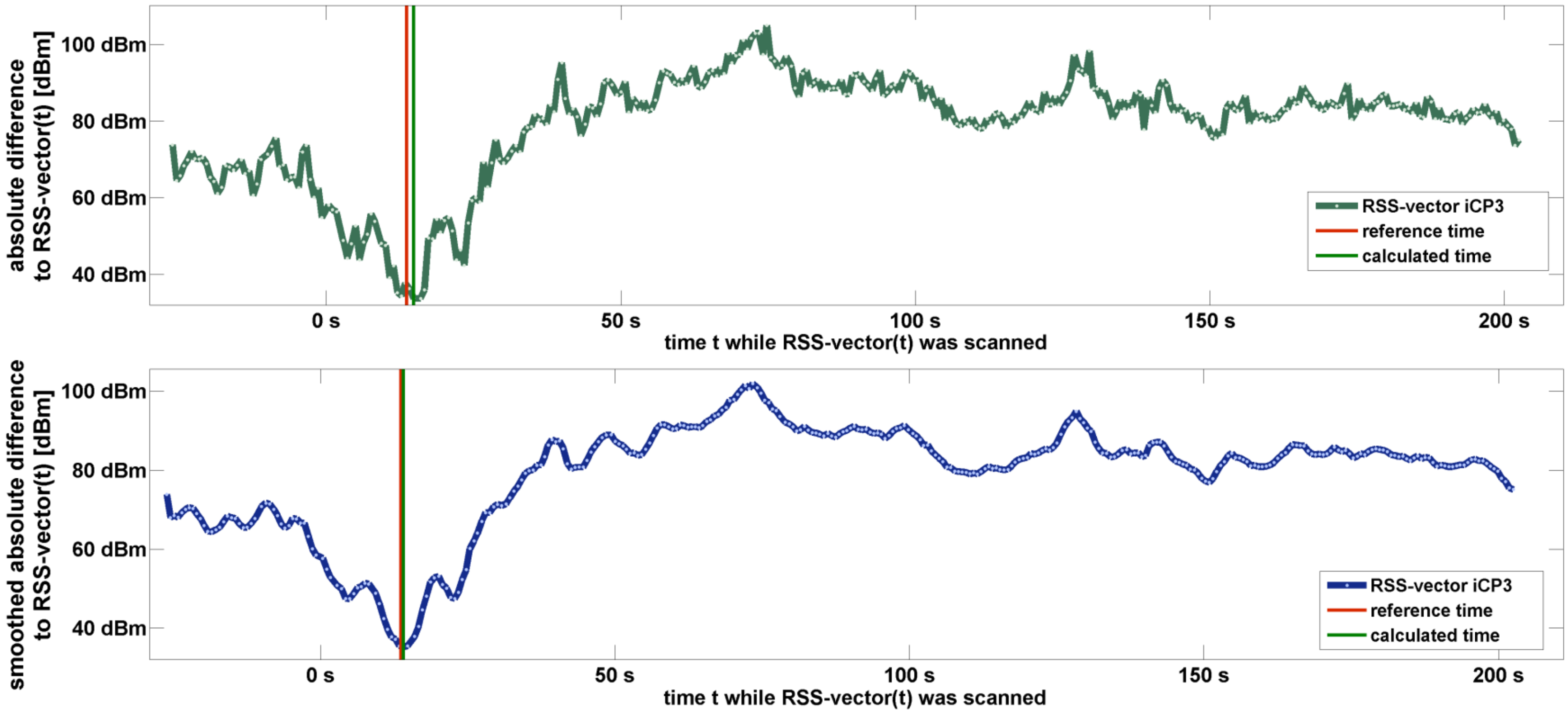
mean deviation 10.1 m

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