UJIINDOORLOC - A NEW MULTI-BUILDING AND MULTI-FLOOR DATABASE FOR WLAN FINGERPRINT-BASED INDOOR LOCALIZATION PROBLEMS

- Carried out by Joaquín Torres-Sospedra et al. at 2 Spanish universities.
- Largest & 1st publicly available database in literature.
- Objective: To provide an objective database for comparing positioning systems & WLAN-fingerprinting algorithms.

UJIINDOORLOC WLAN INTRODUCTION

- GPS (Global Positioning System): A navigational system that uses satellites in order to provide accurate positional data.
- WLAN (Wireless Local Area Network): Wireless network that allows 2 or more devices to connect & communicate via their high-frequency radio waves. Usually includes an access point (WAP) to the Internet.
- GPS is lost when indoor and can only solve OUTDOOR localization problems.
- Many studies use WLAN and mobile phones to solve indoor localization problems but their databases are not as detailed and publicly available as UJIINOORLOC.

UJIINDOORLOC WLAN INTRODUCTION

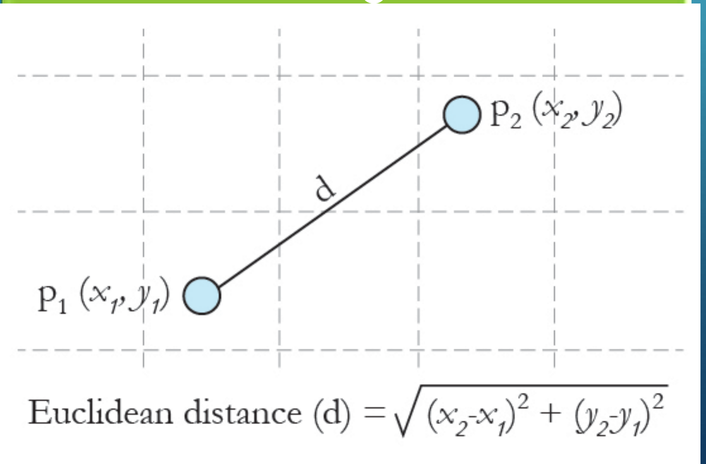
- RSSI (Received Signal Strength Indicator): This is a measurement of the power present in a received radio signal.
- WLAN fingerprint-based indoor localization methods are based on RSSI values and comprise of 2 phases:

1) Calibration – Construction of a radio map showing where the users should be detected.
2) Operation – User acquires signal strengths from the visible access points of the WLAN that can be detected from the user's position. This becomes the test sample which is sent to the server for comparison of training samples.

UJIINDOORLOC WLAN BASELINE

- Indoor Localization System – Used k-Nearest Neighbour technique (kNN-1NN) along with Euclidean Distance:

Calculation of the Euclidean Distance of the current fingerprint with respect to the fingerprints in the training set.



If there is <u>1</u> candidate with the shortest distance.

The location of the current fingerprint corresponds to the location of the Euclidean's closest training fingerprint.

If there is <u>more than 1</u> candidate with the shortest distance.

The location corresponds to the average of the Euclidean's training fingerprints on the "winning" building and floor. Tie = Error.

UJIINDOORLOC WLAN MAIN CHARACTERISTICS

- Covers a surface of 108,703m^2 – 3 buildings with 4 or 5 floors.
- 933 different places (reference points) appeared in the database.
- 19,938 sample points were obtained for training/learning and 1,111 were obtained for validation/testing – 21,049 in total.
- Testing samples were taken 4 months after the training samples to ensure dataset independence.
- 520 different WAPs (Wireless Access Points) appeared in the database.
- Data collected by more than 20 users with 25 different mobile device models.

UJIINDOORLOC WLAN TRAINING SET GENERATION

- Android apps were used to create the database. They provided geographic information of the interior of the buildings i.e. the training reference points localization. – CaptureLoc for Training Data & ValidationLoc for Testing Data.

CaptureLoc captures records.

→

Information from CaptureLoc is sent to centralized server.

→

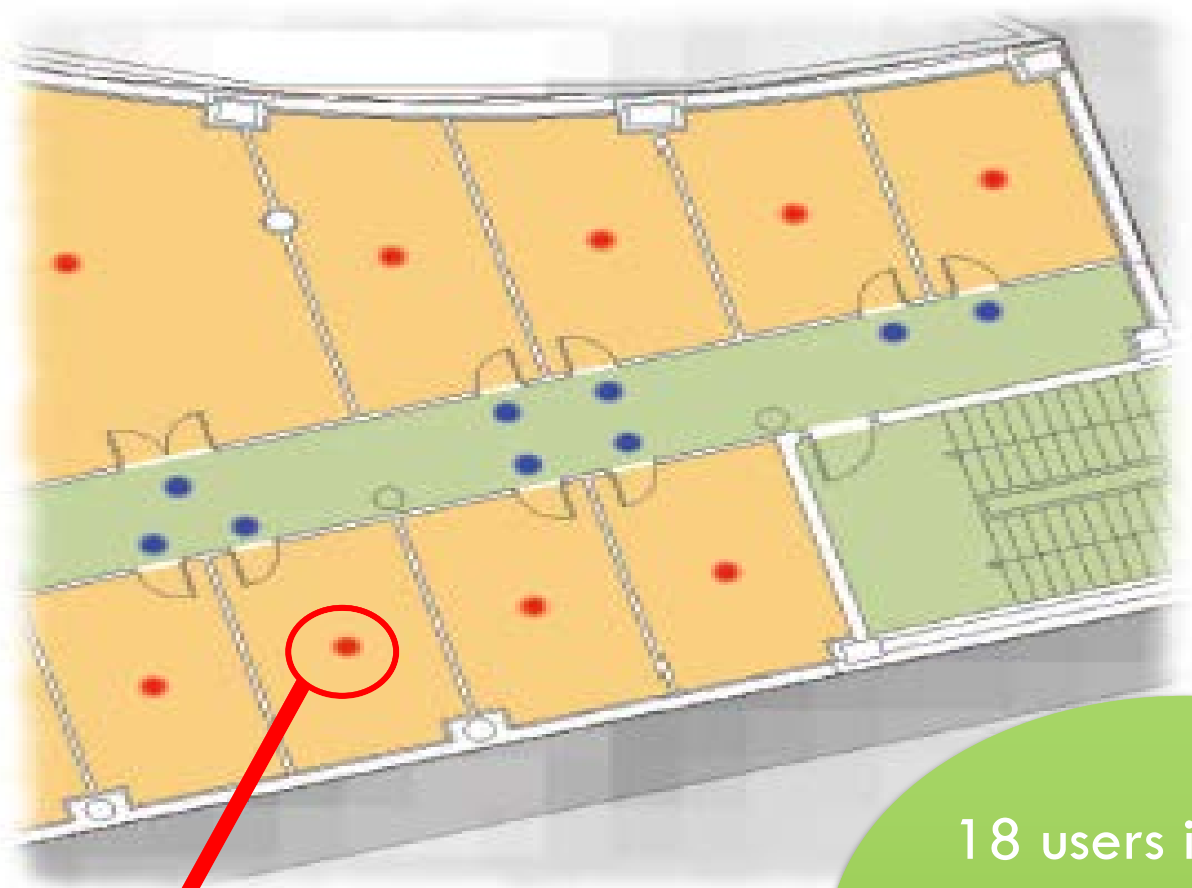
Centralized server permanently stores the information.

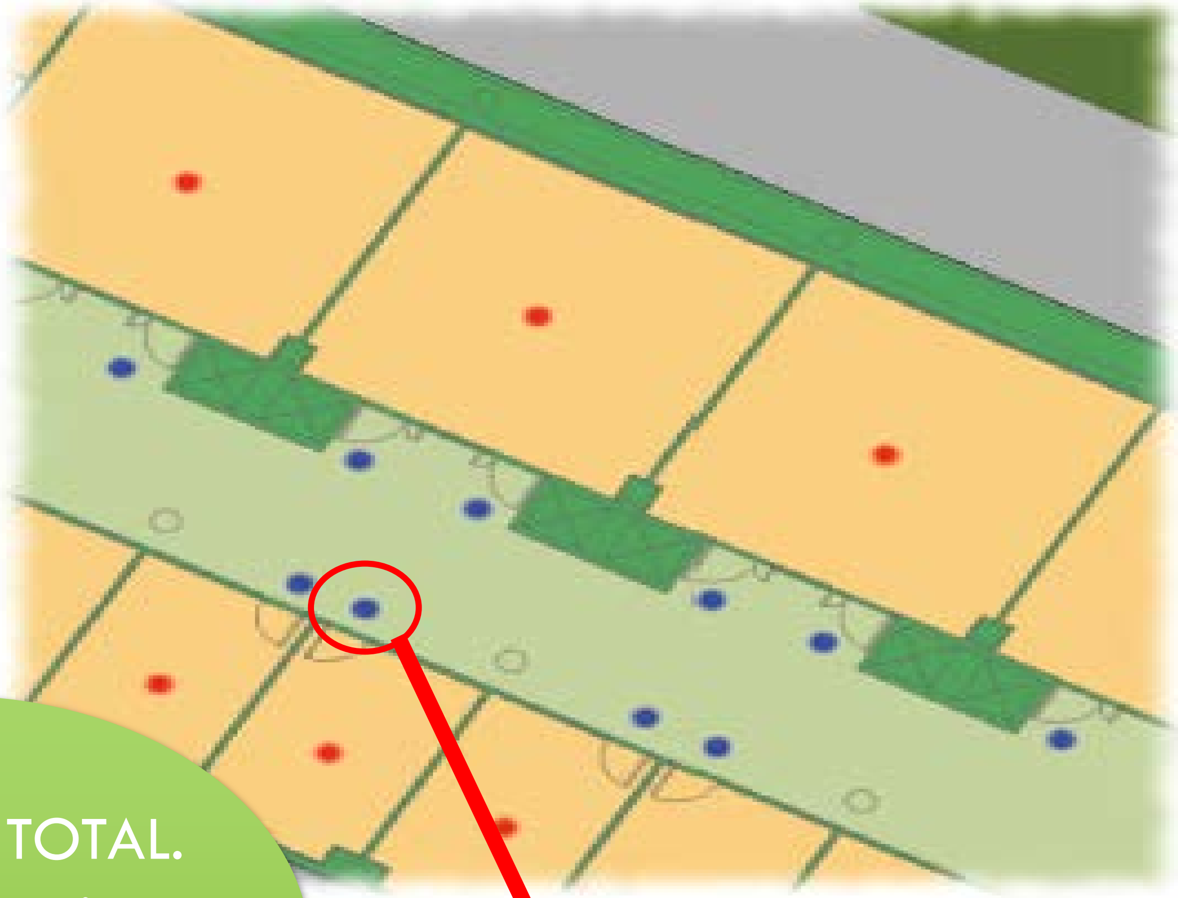
→

Process is repeated 10 times for each captured location.

UJIINDOORLOC WLAN TRAINING SET GENERATION

UJIINDOORLOC WLAN TRAINING SET GENERATION





The centre of the room is the inside reference point of a closed space.

18 users in TOTAL. Reference points were covered by least 2 users!

The front of each door is the outside reference point for a closed space.

UJIINDOORLOC WLAN VALIDATION SET GENERATION

ValidateLoc captures more points for validation purposes.

ONLY WAPs and RSSI levels from ValidateLoc are sent to centralized server.

ValidateLoc gets a point in the building (longitude, latitude & floor) from the centralized server.

ValidateLoc ensures the location is accurate via asking the user.

UJIINDOORLOC WLAN VALIDATION SET GENERATION

ValidateLoc ensures the location is accurate via asking the user.

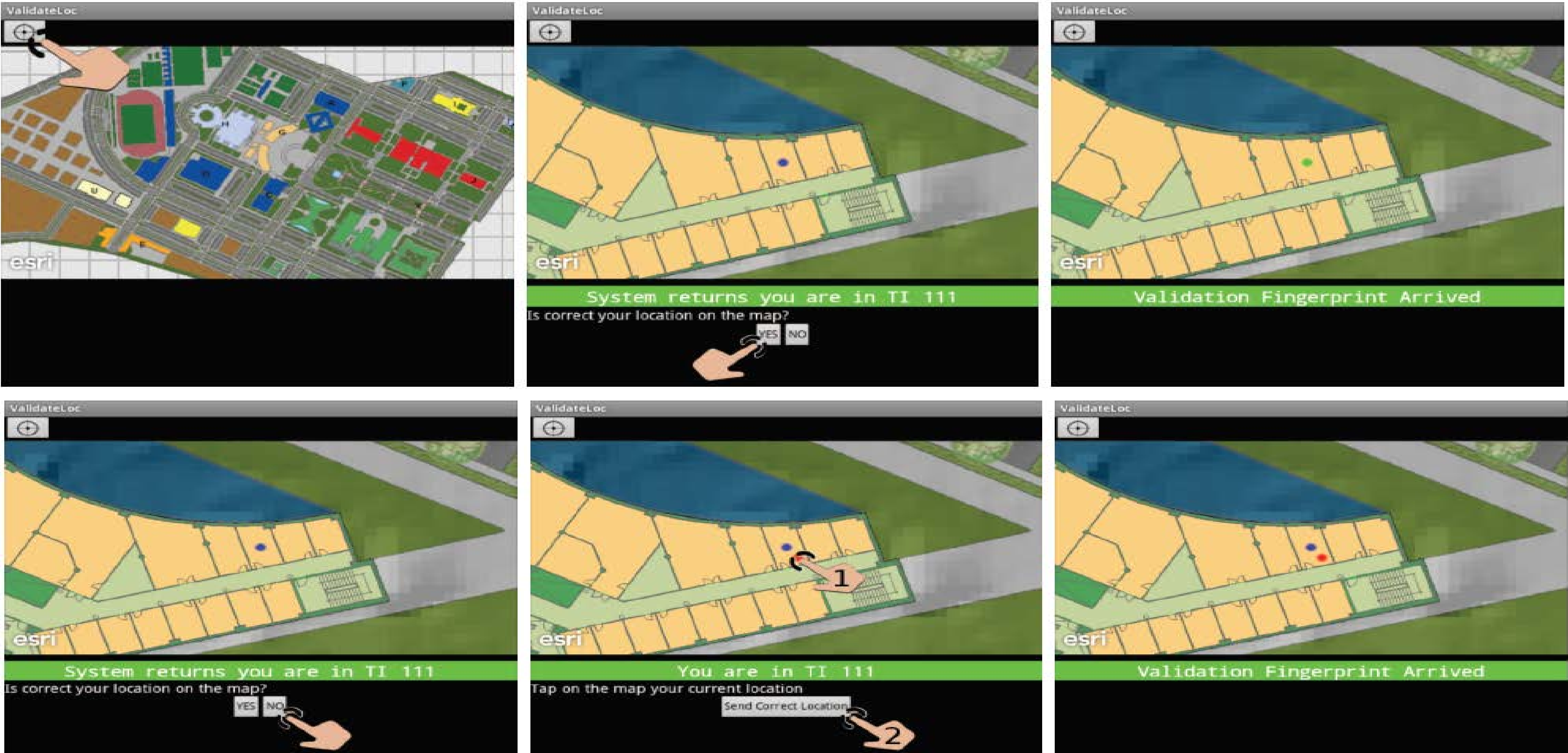
If accurate, the WiFi fingerprint & the successfully predicted localization are sent to the centralized server.

The WiFi fingerprint & the successfully predicted localization are permanently stored in the server.

If inaccurate, user is asked to select the real localization and this is sent to the centralized server.

The real localization selected by the user is permanently stored in the server.

UJIINDOORLOC WLAN VALIDATION SET GENERATION



FACTORS THAT MAY AFFECT SCANNED WAPS

- Main factors that they found affected the number of scanned WAPs:
  1. Location (WiFi coverage)
  2. Mobile phone model (Android version & hardware)
  3. How the device is held

UJIINDOORLOC WLAN DATABASE DESCRIPTION

TABLE II. EXAMPLE OF ONE DATABASE ENTRY (7754-TH RECORD). IT WAS CAPTURED ON JUNE, 4TH 2013 (12:02:22PM GMT+02) BY USER11 WITH A *HTC Wildfire S A510e* (ANDROID VERSION 2.3.5). THE DEVICE DETECTED 14 WAP (NEGATIVE RSSI VALUES) ON THE REFERENCE POINT LOCATED OUTSIDE OFFICE 111 ON THE THIRD FLOOR OF THE *TI* BUILDING.

[1] WAP_{001}	...	[520] WAP_{520}	[521] Longitude	[522] Latitude	[523] Floor	[524] BuildingID	[525] SpaceID	[526] Rel.Pos.	[527] UserID	[528] PhoneID	[529] Time
-97	...	+100	-7594.7...	4864983.9...	3	0	111	2	11	13	1370340142

Real-world coordinates

Space identifiers

Timestamp set by centralized server

TABLE III. EXTRACT OF THE VECTOR THAT REPRESENTS THE RSSI VALUES. MAC ADDRESSES HAVE BEEN ANONYMIZED DUE TO PRIVACY REASONS.

WAP_{001}	...	WAP_{031}	WAP_{032}	WAP_{033}	WAP_{034}	WAP_{035}	WAP_{036}	...	WAP_{520}
-97	...	+100	-97	+100	+100	-65	-65	...	+100

Android provided integer RSSI values.

WAP identifiers (MAC addresses) are linked to vector positions.

UJIINDOORLOC WLAN DATABASE DESCRIPTION

TABLE IV. EXTRACT OF THE WAPS LIST WITH 14 RSSI VALUES PROVIDED BY *getScanResults()*.

pos. in list	WAP Identifier	RSSI level
1^{st}	WAP_{032}	$-97dBm$
2^{nd}	WAP_{001}	$-97dBm$
3^{rd}	WAP_{268}	$-97dBm$
4^{th}	WAP_{150}	$-94dBm$
	...	
11^{th}	WAP_{036}	$-65dBm$
12^{th}	WAP_{035}	$-65dBm$
13^{th}	WAP_{142}	$-48dBm$
14^{th}	WAP_{143}	$-46dBm$

Method in Android class to obtain list of detected WAPs.

RSSI Levels in dBm represents the detected WAP's signal intensity

-100dBm = very weak signal
0dBm = extremely good signal
+100dBm = undetected WAP

Ordering actually depends on mobile device's model & Android version.

This data represents ONLY 1 Wifi Scan!

UJIINDOORLOC WLAN RESULTS

TABLE XI. UJIINDOORLOC RESULTS WITH 1NN IN CONJUNCTION WITH *Euclidean Distance*

Error in positioning	$7.9m$
Success rate	89.92%
Time	$495.26 \pm 0.54ms$

Average error in meters of validation fingerprints correctly located inside corresponding building & floor.

% of validation fingerprints correctly located inside corresponding building & floor.

Average time in milliseconds taken to find the precise location (longitude, latitude, floor) per fingerprint.

More than 10% - Incorrect localization
0.27% - Wrong building
9.81% - Wrong floor

UJIINDOORLOC WLAN PROBLEMS

- Reducing the redundant access points and keeping a complete coverage.
- Although some WAPs are visible, majority of them are hidden to human eye. WLAN antennas are normally located in restricted areas or in the ceiling.
- Detection of low-coverage places. Adding new antennas can improve the localization algorithm.
- Detection of WLAN collision places where some WAPs are emitting in the same channel – Degradation of WAP connectivity.
- The validation fingerprints were taken 3 months later than the training ones, and some WAPs disappeared and new ones were introduced.

UJIINDOORLOC WLAN CONCLUSIONS

- This paper introduces a NEW database for indoor localization based on a WLAN fingerprinting environment.
- The database description consists of the features, procedures as well as the applications used to generate the database.
- Unlike other databases:
  1. The samples taken can be considered as realistic as human users were used.
  2. A variety of mobile devices as well as a large number of users were used.
  3. A large area was covered. More buildings with more than 1 floor were used and their internal structures differed.
  4. Validation samples have been provided.
- This database can therefore be presented as a common, public database that can be used for comparisons.

THANK YOU FOR LISTENING!
NEXT IS CHONGFENG'S PRESENTATION!

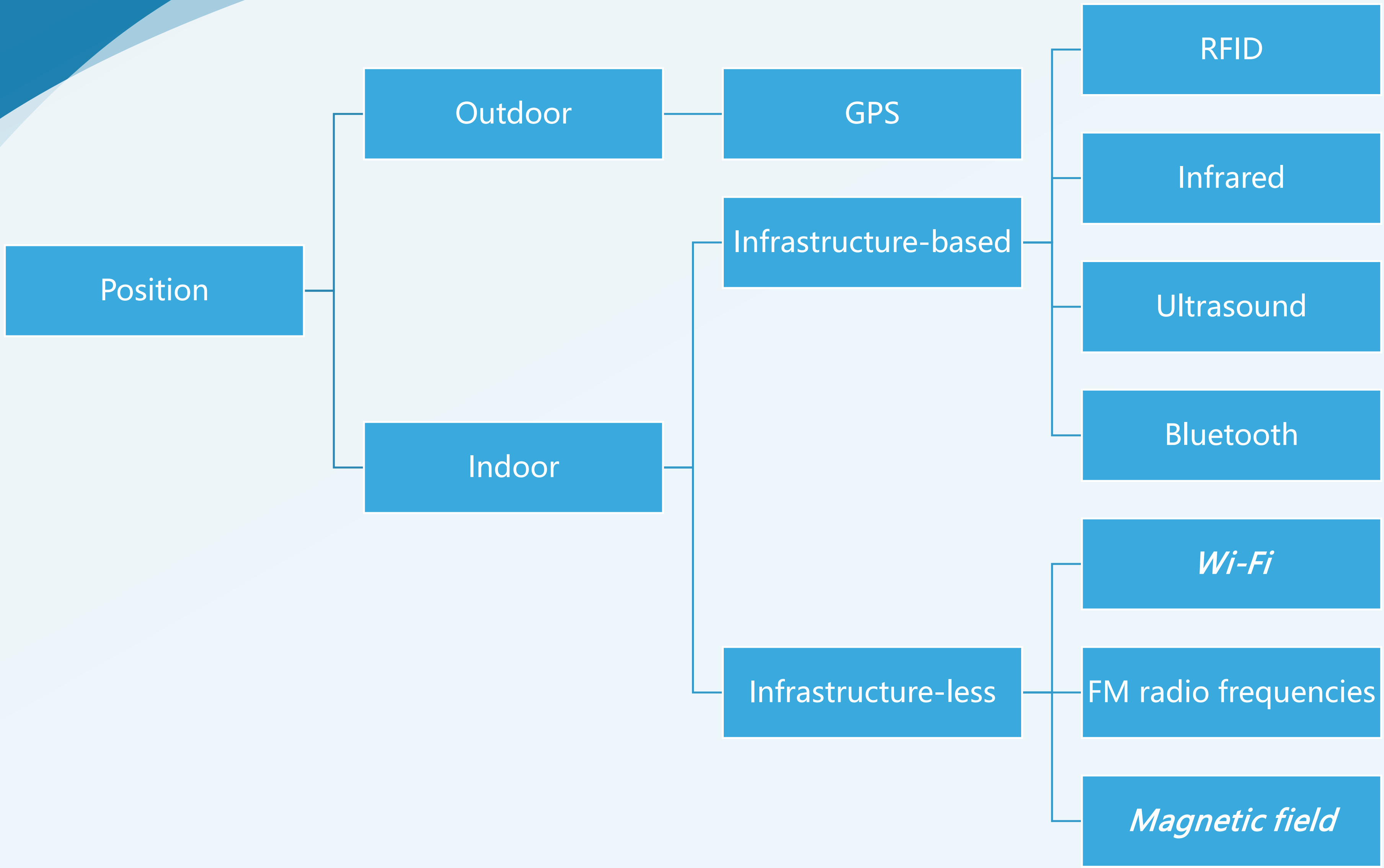
UJIIndoorLoc-Mag: A New Database for Magnetic Field-Based Localization Problems

By: Joaquín Torres-Sospedra, David Rambla, Raul Montoliu,
Oscar Belmonte, and Joaquín Huerta

Keywords—Indoor Localization; Magnetic field; Database; Comparison of methods

Contect

01 Introduction



01 Introduction

the *UJIIndoorLoc-Mag* database: First publicly available database that could be used to make comparisons among different methods in this field.

- Consists of 281 continuous samples (270 for training and 11 for testing).
- Taken in our 260m2 (15x20m approx.) laboratory.
- Each sample comprises a set of discrete captures taken along the 8 corridors (including intersections) of the laboratory.
- Data is selected per 0.1 seconds.
- Almost 40,000 discrete captures obtained from the **magnetometer, accelerometer and orientation sensor** of a mobile phone.

01 Introduction

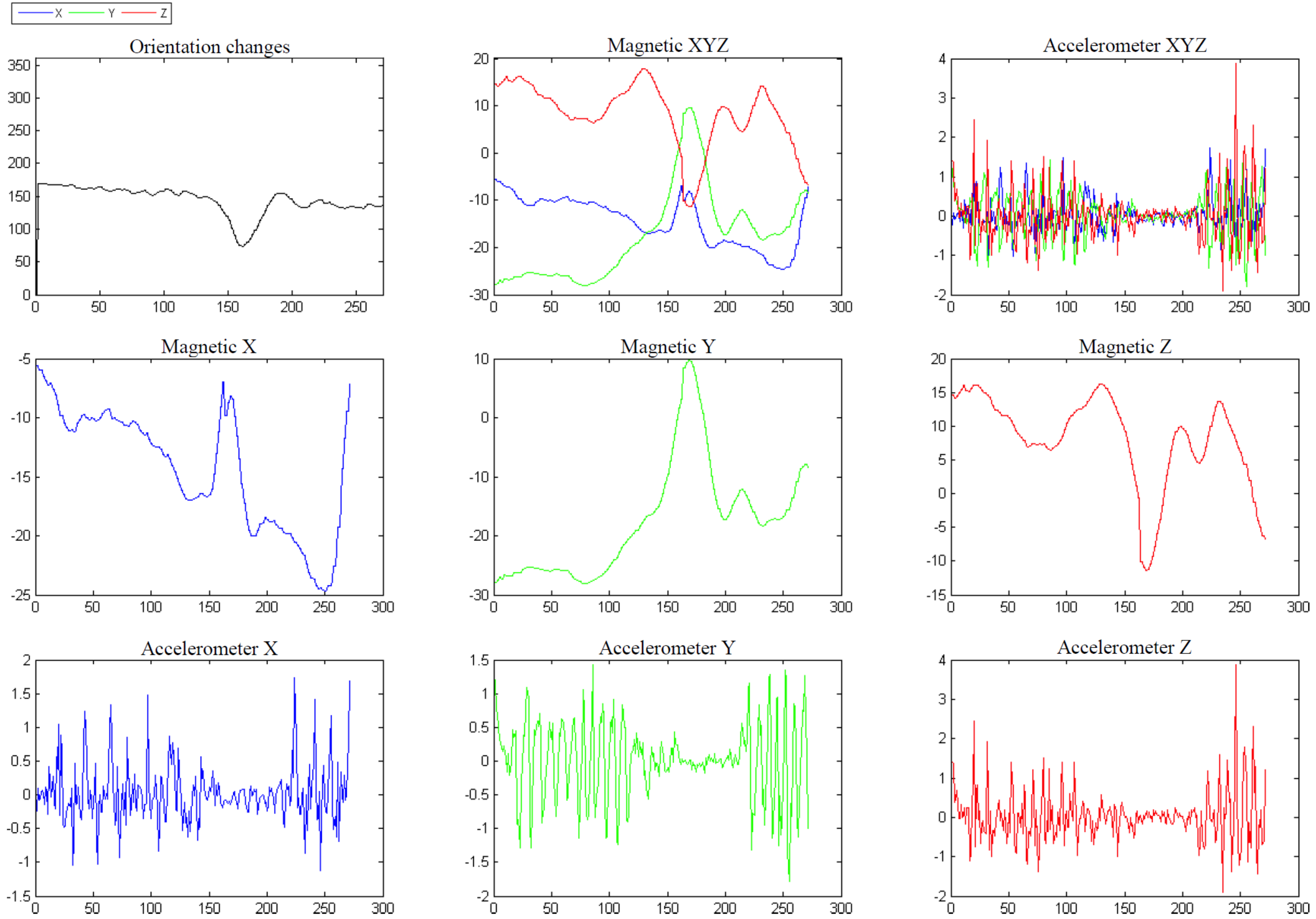


Fig. 6. Data values collected for a trajectory. Simplified orientation is shown for visualization purposed. Moreover the values of magnetic and accelerometer are also shown in separate plots.

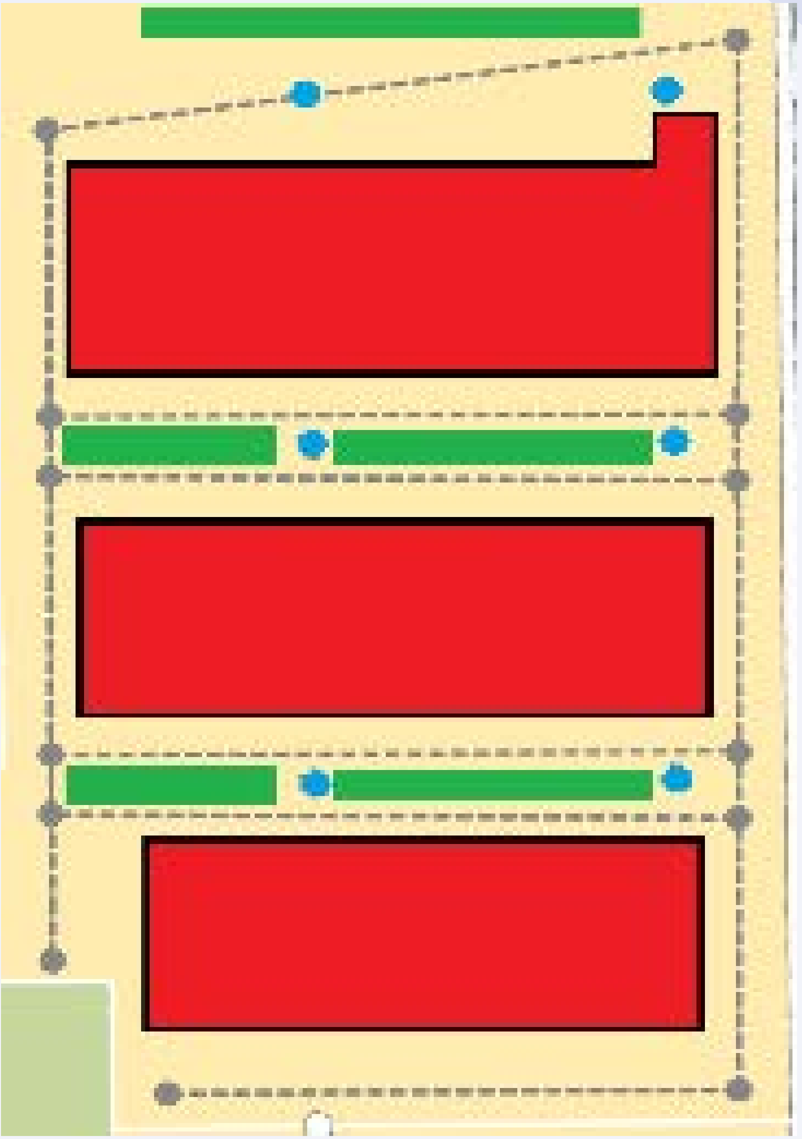
01 Introduction

There are three types of database.

1. **Continuous samples taken in a lineal environment (like a corridor).**
2. Discrete samples taken in a lineal environment.
3. Discrete samples taken in a two dimensional space.

a single continuous sample corresponds to a sequence of some consecutive discrete samples taken in a lineal environment.

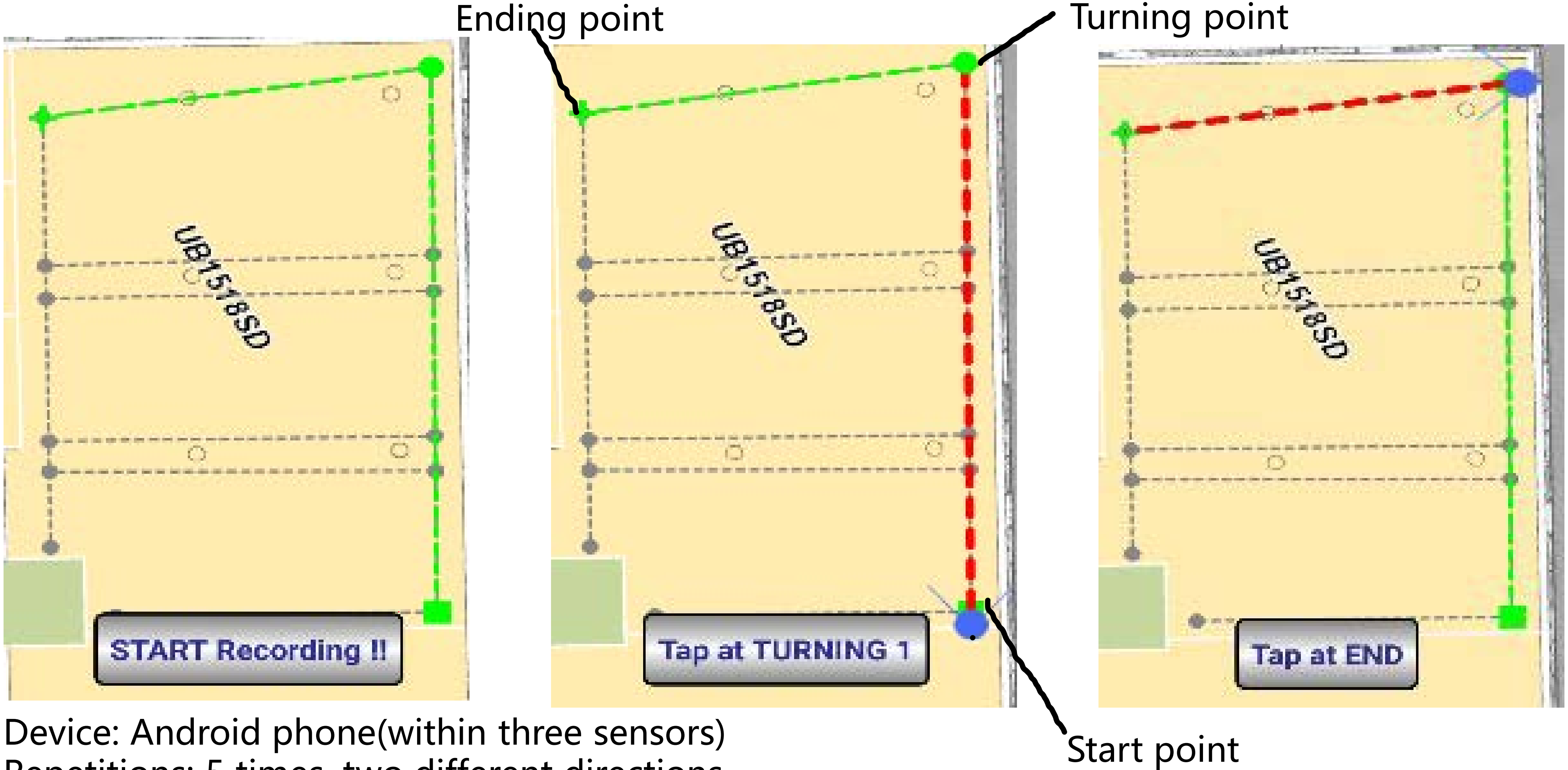
02 Experiment



The lab’s map where desktop tables are highlighted in red, bookcases in green, and Columns in blue.

8 corridors
19 intersections

02 Experiment



Device: Android phone(within three sensors)
Repetitions: 5 times, two different directions
Period: every 0.1 seconds velocity: normal(1.39m/s approx.)

03 Description of database files

270 samples
80 are single corridor
190 are two corrdors

Samples

Training purpose — One or two corridors

Testing purpose — Multiple corridors

11 samples

03 Description of database files

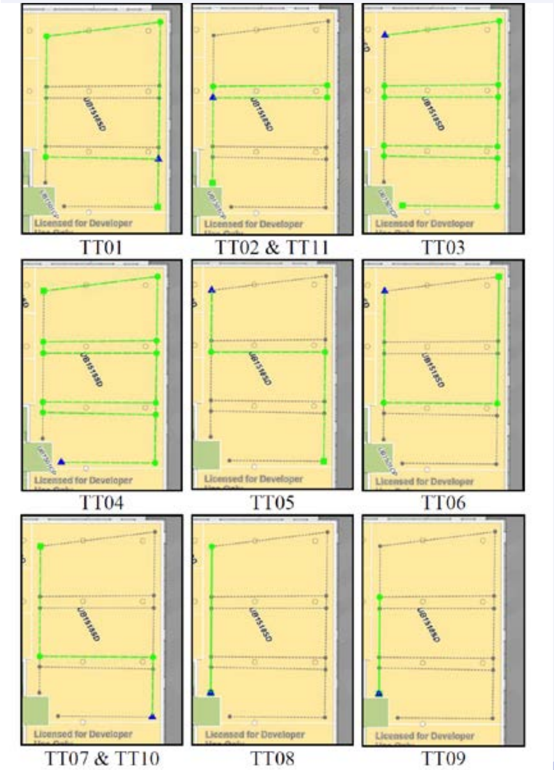
270 samples
80 are single corridor
190 are two corrdors

Samples

Training purpose — One or two corridors

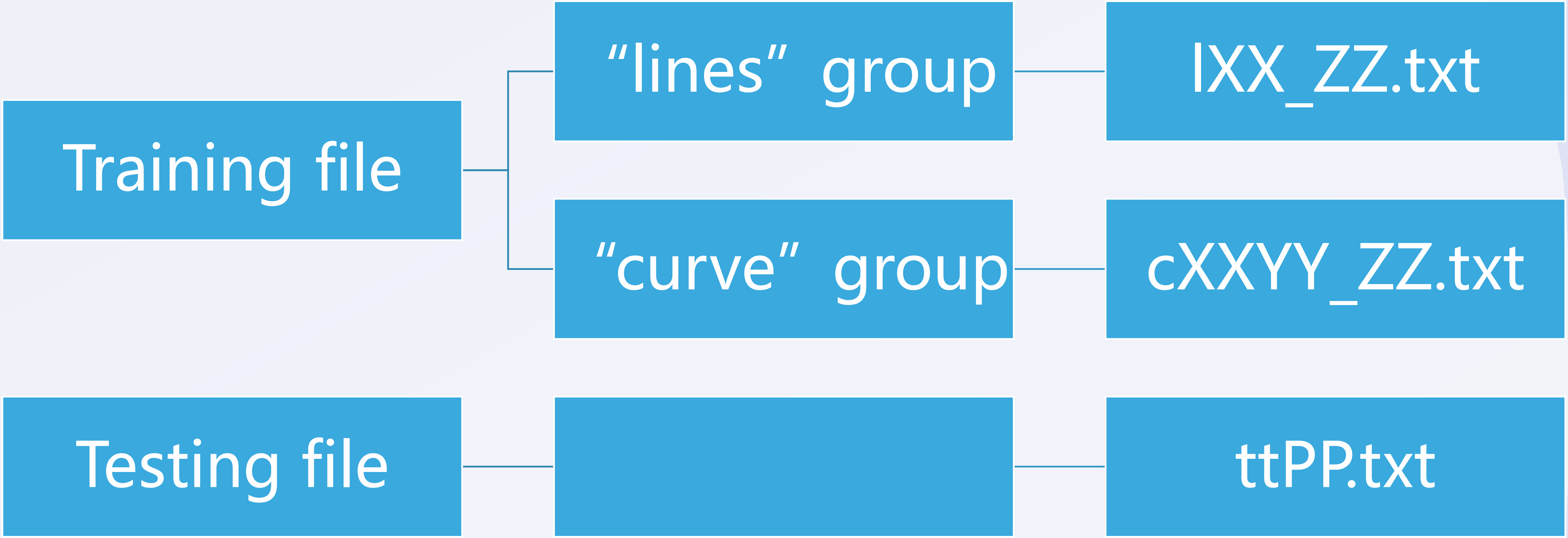
Testing purpose — Multiple corridors

11 samples



03 Description of database files

281 continuous samples are stored as independent text files. The training ones are grouped into two main categories “lines” and “curves” .



03 Description of database files

281 continuous samples are stored as independent text files. The training ones are grouped into two main categories “lines” and “curves” .

File	Description
lXX_ZZ.txt	XX :the number of corridor & orientation (n or r) ZZ :repetition. Example: l3r_03.txt
cXXYY_ZZ.txt	XX and YY :the number of corridor & orientation for the first and second corridors in the two corridors trajectory ZZ : repetition. Example: c5n1r_05.txt
ttPP.txt	PP : the complex testing trajectory number

03 Description of database files

Data has been stored as a simple text file as follows:

ts_1	mx_1	my_1	mz_1	ax_1	ay_1	az_1	ox_1	oy_1	oz_1
…									
ts_n	mx_n	my_n	mz_n	ax_n	ay_n	az_n	ox_n	oy_n	oz_n
<m>									
lat_1	lon_1	lat_2	lon_2	FS_1	LS_1				
…									
lat_m	lon_m	lat_{m+1}	lon_{m+1}	FS_m	LS_m				

n : the number of samples collected in the trajectory at a 0.1 seconds frequency

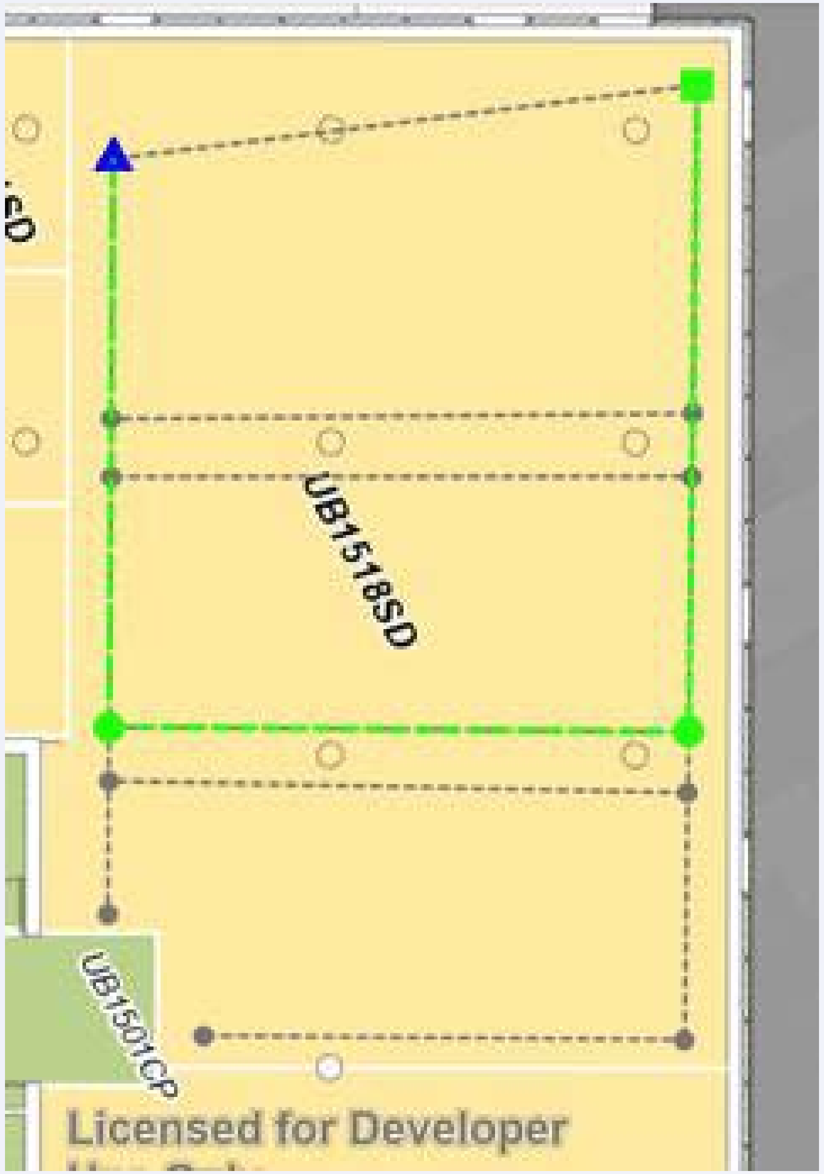
m : the number of segments (corridors) in the trajectory

Each sample contains the timestamp *ts* and the values from magnetometer, accelerometer and orientation sensors in the three axes, which are denoted with *mx*, *my*, *mz*, *ax*, *ay*, *az*, *ox*, *oy* and *oz*.

Lat_i and lon_i corresponds to the coordinates (latitude & longitude in decimal degrees) of the initial, intermediate (intersections) and final points.

FSi and *LSi* state for the i-th trajectory' s first and last sample respectively in the full sequence of samples collected during the trajectory mapping.

03 Description of database files



TT06

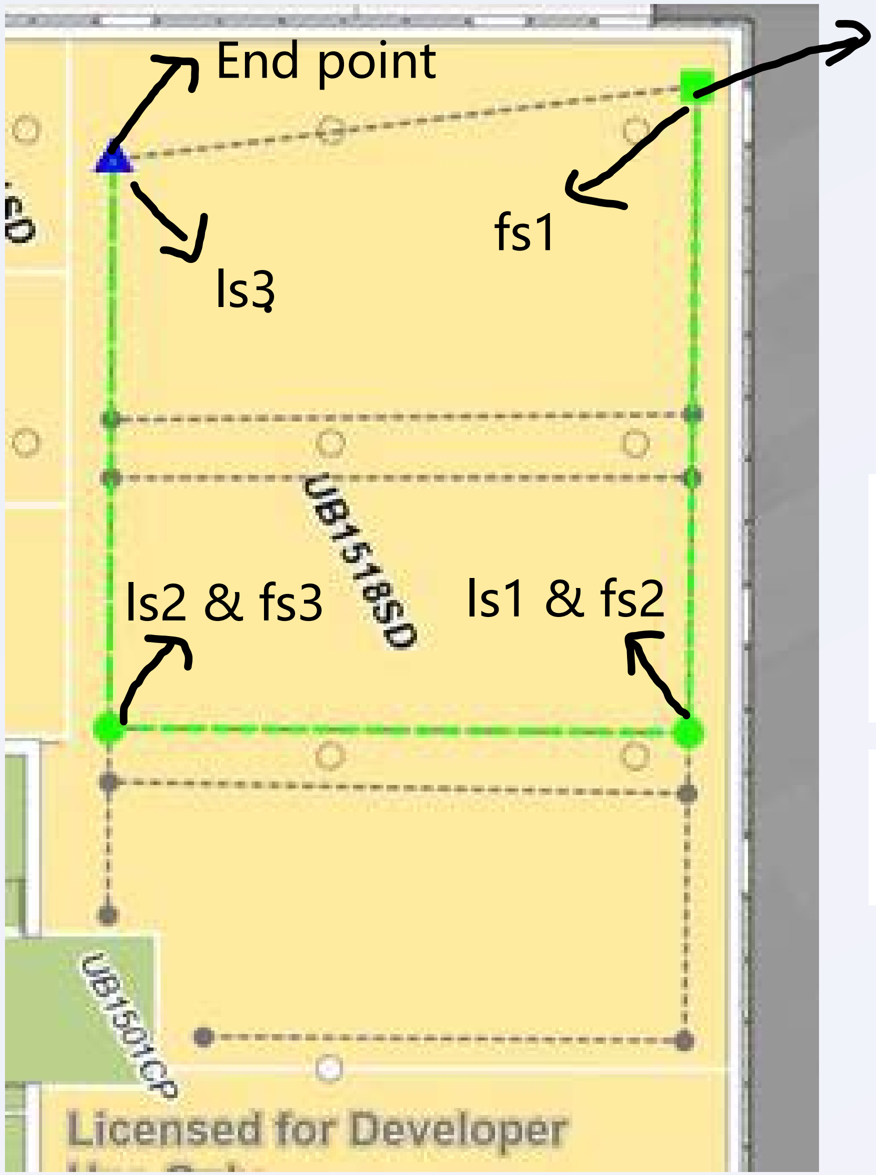
The text files composed by two well-differentiated parts separated by the row indicating the number of segments in the trajectory: 1) the sequence of discrete samples taken during the trajectory mapping, and 2) the configuration data.

Part 1: contains the timestamp, the vector data from magnetometer, accelerometer and orientation.

ts_{24}	1417178330528	ts_{25}	1417178330629
mx_{24}	24.899292	mx_{25}	24.719238
my_{24}	-10.319519	my_{25}	-11.219788
mz_{24}	-49.55902	mz_{25}	-49.319458
ax_{24}	-0.12917818	ax_{25}	-0.15856716
ay_{24}	0.52311563	ay_{25}	0.68318987
az_{24}	-0.19135952	az_{25}	-0.15023136
ox_{24}	-64.537674	ox_{25}	-62.273254
oy_{24}	-21.03711	oy_{25}	-21.420563
oz_{24}	0.15363675	oz_{25}	0.5122262

Two consecutive samples (vertically represented here) from 6-th testing trajectory

03 Description of database files



TT06

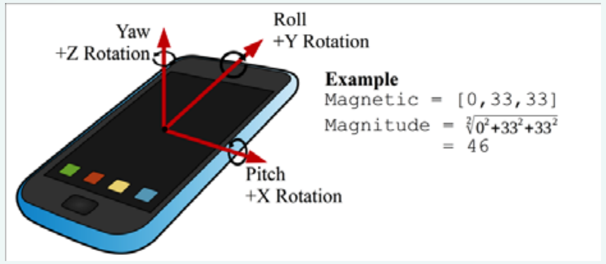
Part 2: contains the information about location of initial, intermediate and ending points. These can be associated to corridor segments and turnings.

lat_1	lon_1	lat_2	lon_2	FS_1	LS_1
…					
lat_m	lon_m	lat_{m+1}	lon_{m+1}	FS_m	LS_m

39.99389	-0.07375	39.99393	-0.07384	0	71
39.99393	-0.07384	39.99386	-0.07389	72	159
39.99386	-0.07389	39.99388	-0.07394	160	223

the configuration part for the 6-th testing trajectory

04 features



- Magnetic field: Provides a vector that corresponds to the strength and direction of the magnetic field. But it is not trivial to detect a user's orientation change (turn).
- Orientation : Record raw data from the orientation sensor. Provides the direction vector and the values are measured in degrees.
- Acceleration: Can detect the user's steps and therefore estimate the velocity. (by resampling or different speed configuration)
- Timestamp: With the start & end point and normal velocity, user's position can be analysis.
- Start/end/turning points: Improve the accurate.

05 Baseline

- Two simple baseline methods are developed and tested to provide a starting point that any more sophisticated indoor localization algorithm should be able to overcome.
- The first one uses a *discrete method* and the second one uses *continuous method*. Both algorithms ONLY use the training samples taken on the 8 corridors and from the magnetometer.
- The k-NN algorithm [11] with k = 1 has been used to estimate the location of each test sample, so the test current location would correspond to the most similar train sample. The location of the most similar sample in the training set is the one assigned to the test sample.

05 Baseline
--discrete method

- Discrete captures extracted from the continuous training samples. Same procedure performed for testing capture.
- The capture consists 5 features: lat, lon, mX, mY, mZ.
- 8943 samples for training and 4380 for testing totally.

Euclidean's distance between two samples, $m_1 = [m_{X,1}, m_{Y,1}, m_{Z,1}]$ $m_2 = [m_{X,2}, m_{Y,2}, m_{Z,2}]$ is

$$d(m_1, m_2) = \sqrt{(m_{X,1} - m_{X,2})^2 + (m_{Y,1} - m_{Y,2})^2 + (m_{Z,1} - m_{Z,2})^2}$$

05 Baseline

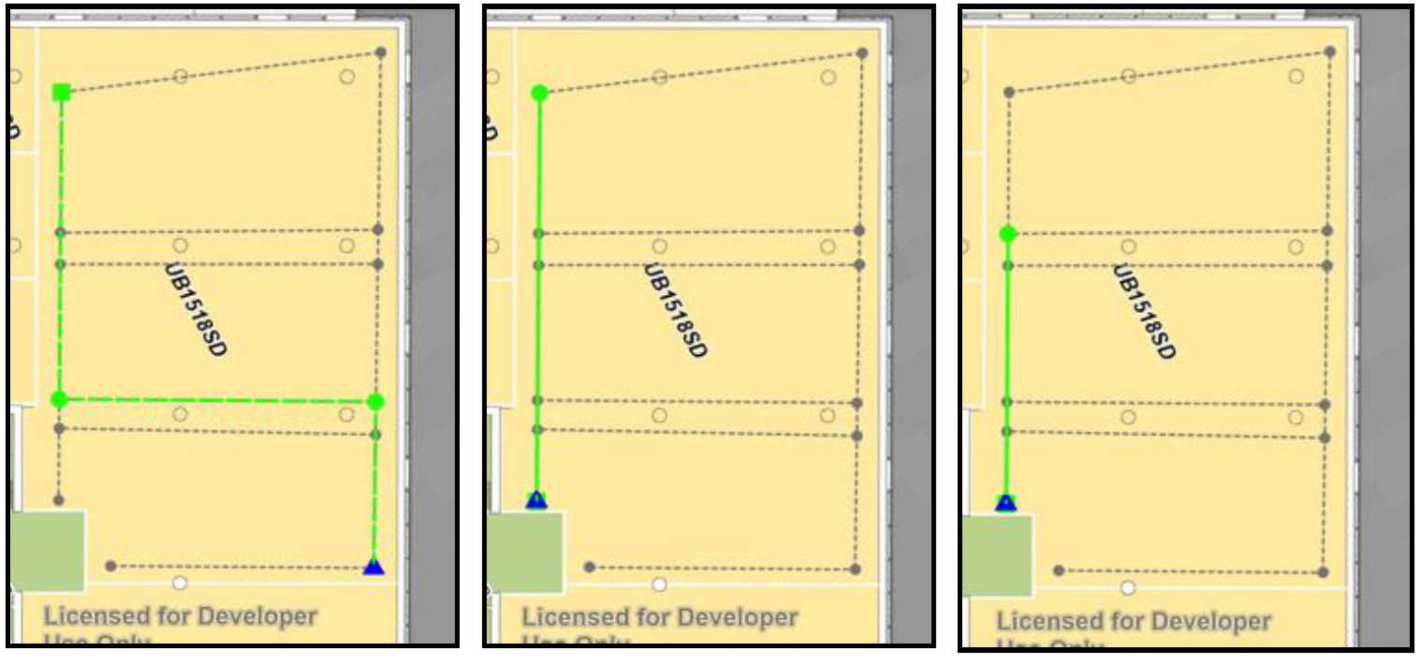
--*Continuous method*

- each continuous training sample is divided in several subsamples of 5 seconds each one. For instance, if a sample is 10 seconds long and has 100 discrete samples, then it is divided in 6 continuous subsamples, [1-50], [11-60], ..., [51-100].
- The capture consists 5 features: lat, lon, mX, mY, mZ.
- 540 samples for training and 231 for testing totally.

Euclidean' s distance between two samples, $vm_1 = [vm_{X,1}, vm_{Y,1}, vm_{Z,1}]$ and $vm_2 = [vm_{X,2}, vm_{Y,2}, vm_{Z,2}]$ is

$$d_{vec}(vm_1, vm_2) = \frac{1}{N}\sum_{i=1}^{N} d(vm_1[i], vm_2[i])$$

05 Baseline --error



TT07 & TT10 TT08 TT09

7	*211*	*7.72±0.29*	*8*	*6.24±1.51*
8	*246*	*9.26±0.22*	*16*	*6.58±0.88*
9	*196*	*3.33±0.21*	*11*	*1.25±0.34*
10	*223*	*7.46±0.3*	*10*	*5.33±1.36*
11	*287*	*6.65±0.18*	*13*	*4.7±0.81*
mean		*7.23±0.38*		*6.05±0.43*

Errors of two simple baseline are high, while the continuous method’ s error is small than discrete method’ s.

is:

points corresponds to the latitude (lat) ongitude (lon) coordinates in decimal rees are not expressed in lineal meters. *haversine formula can be used instead.* formula only use the training samples taken on the 8 corridors and from the magnetometer. Better results can be achieved if training samples taken at intersections were considered in the algorithm.

06 Conclusions

- This paper introduces a new database for indoor localization, *UJIIndoorLoc-Mag*, on the basis of variations on the magnetic field.
- The database description including the features used in the database. The procedure and the applications used to generate the database have also been described.
- Two simple baseline methods were introduced to show the viability of the usage of the magnetic database
- Further work will be focused on increasing the amount of samples of the database and a more robust indoor positioning method.

SIMILARITIES

- Both present a **detailed and public** database for WiFi Fingerprinting & Magnetic Field based indoor localization methods.
- Database for the training set is created via an Android app with several sensors.
- Additional hardware is unnecessary.
- Both use the kNN algorithm in baseline.

DIFFERENCES

WLAN FINGERPRINT-BASED	MAGNETIC FIELD-BASED
2 sensors are used to determine indoor localization	3 sensors are used to determine indoor localization
Data is divided into 2 sets: training & testing. Training is more complex	Data is divided into 2 sets: training & testing. Testing is more complex
3 buildings with 4 or 5 floors with various rooms are used to obtain the training & testing data.	Only 1 room in 1 building and on 1 floor is used to obtain the training & testing data.
Approximately 495ms was taken to collect each WiFi fingerprint.	Data was collected every 0.1s.
Baseline uses only 1 method.	Baseline uses 2 methods: 1st is discrete. 2nd is continuous.
Data is based on positioning – user must remain stationary.	Data is created as continuous samples.

DIFFERENCES

WLAN FINGERPRINT-BASED DATABASE

TABLE II. EXAMPLE OF ONE DATABASE ENTRY (7754-TH RECORD). IT WAS CAPTURED ON JUNE, 4TH 2013 (12:02:22PM GMT+02) BY USER11 WITH A *HTC Wildfire S A510e* (ANDROID VERSION 2.3.5). THE DEVICE DETECTED 14 WAP (NEGATIVE RSSI VALUES) ON THE REFERENCE POINT LOCATED OUTSIDE OFFICE 111 ON THE THIRD FLOOR OF THE *TI* BUILDING.

[1] WAP_{001}	...	[520] WAP_{520}	[521] Longitude	[522] Latitude	[523] Floor	[524] BuildingID	[525] SpaceID	[526] Rel.Pos.	[527] UserID	[528] PhoneID	[529] Time
-97	...	+100	-7594.7...	4864983.9...	3	0	111	2	11	13	1370340142

~~MAGNETIC FIELD-BASED DATABASE~~

ts_1 mx_1 my_1 mz_1 ax_1 ay_1 az_1 ox_1 oy_1 oz_1

...

ts_n mx_n my_n mz_n ax_n ay_n az_n ox_n oy_n oz_n

<m>

lat_1 lon_1 lat_2 lon_2 FS_1 LS_1

...

lat_m lon_m lat_{m+1} lon_{m+1} FS_m LS_m

a piece of discrete sample

a continuous sample