26th Conference Radioelektronika 2016, April 19-20, Košice, Slovak Republic

Miskolc IIS Hybrid IPS: Dataset for Hybrid Indoor Positioning

Zsolt Tóth, Judit Tamás

Institute of Information Science, University of Miskolc, Hungary
tothzs@iit.uni-miskolc.hu, tamas13@iit.uni-miskolc.hu

ABSTRACT

A dataset for hybrid indoor positioning systems is presented in this paper. The existing indoor positioning datasets are focused on a single technology such as WiFi RSSI or Magnetometer. But the presented dataset has data from W-LAN and Bluetooth interfaces, and Magnetometer. The measurements were performed in a three–story building and covers about 50% of the entire building. Moreover, both symbolic and absolute positions are defined for each measurement. The measurements were recorded by the ILONA system and stored in a MySQL database. The dataset is available in both SQL and CSV format.

Index Terms— Dataset, Indoor Positioning, Fingerprinting, WiFi, Bluetooth, Magnetometer

1. INTRODUCTION

Indoor Positioning and Localization is a hot topic in mobile computing these days. It is challenging due to the unique features of the indoor environment. Moreover, the comparison of the existing solutions is difficult due to the different test environments and the variety of the applied technologies. Although, many indoor positioning methods can be found in the literature [1, 2], these solutions are usually evaluated empirically. These empirical results can be used for testing the performance in a given test environment, but they cannot be used to compare the different solutions. Hence, a common test environment is required to compare the different indoor positioning systems properly.

Indoor Positioning can be considered a machine learning problem. UCI Machine Learning Repository [3] contains two indoor positioning datasets [4, 5]. One of these datasets contains WiFi fingerprints and the other dataset has measurements about the variations of the Earth's magnetic field. Thus, these datasets are limited to a single technology.

This paper presents a dataset for hybrid indoor positioning systems. WiFi and Bluetooth interfaces and Magnetometer are used to create the measurement in a given location. Both absolute and symbolic position is given for each measurements. The measurements were done in the Institute of Information Science at the University of Miskolc, that is a three story–building made of reinforced concrete. Hence the GPS signals are usually unavailable in the building. Moreover, even the GSM service is unavailable in the center of the building. The presented measurements are stored in a MySQL database that data model was presented in [6]. The dataset is also available in CSV format which is easier to process and widely used among dataset providers.

The described work was carried out as part of the TÉT_12_SK_2013_0019 project.

2. RELATED WORKS

Indoor Positioning Systems are usually classified based on the applied technology. The early systems required specific client devices for positioning. Active Badge [7] system used infrared to locate people. The badges emit unique signals which were received by the installed readers. Active Bats [8] was based on ultrasound technology and it used specific devices that were installed into the ceiling. WALRUS system [9] also uses ultrasound for room–level positioning.

Modern systems usually use the smart phones as clients due to their popularity, availability and low cost. Smart phones have various built–in sensors which affects on the performance of the system. Movement sensors [10], Magnetometer [11, 12] were used to track the movement of the device in the building. Bluetooth [13] and WLAN interfaces are both used for positioning purposes. Although RFID based solutions [14, 15, 16, 17] have promising results, RFID reader has not been integrated into smart phones yet.

WLAN based indoor positioning systems are popular due to their low installation cost and wide availability. These systems are generally based on client–server architecture and the positioning is performed by the server. Fingerprinting method was presented in the Radar [18] system first. The Horus [19, 20] was also based on fingerprinting and it showed that the performance of the system can be improved by client side filtering techniques. Off–line and on–line phases are usually distinguished by these systems. Site survey is performed in the off–line phase which is tedious, time–consuming and costly. Localization service is available in the on–line phase of the system. Efforts are made to merge the off–line and on–line phases [21]. Some of the popular WLAN based indoor positioning system are compared in [22].

Datasets are widely used for evaluation and comparison

978-1-5090-1674-7/16/$31.00

of various machine learning algorithms. The UCI Machine Learning Repository [3] contains more than 300 datasets for various tasks such as classification, clustering and regression. This repository contains two datasets [4, 5] related to indoor positioning and the measurements were performed in a multi–building and multi–floor environment. One of these datasets [4] contains approximately 20.000 instances of WiFi fingerprints. The other dataset [5] has about 40.000 instances and it shows the variation of the Earth's magnetic field. However these datasets allow the comparison of various indoor positioning methods, but they are limited to one technology.

3. MEASUREMENTS

3.1. Environment

The measurements were performed in the Institute of Information Science at the University of Miskolc. Figure 1 shows the structure of the building. Three part of the floors can be distinguished based on their purpose. The offices and the laboratories are placed at both side of the building and their access is restricted. The hall and the corridors have public access. Finally, storages and rest rooms are in the center of the floor. The walls are made of reinforced concrete so GPS signal are usually unavailable within the building. Furthermore, the installation of numerous WiFi access points was necessary to provide WLAN access in the building.

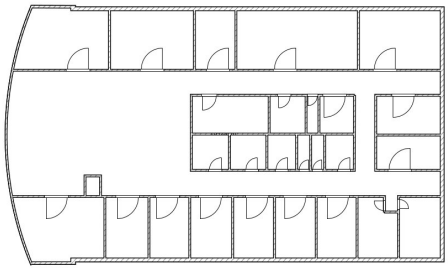
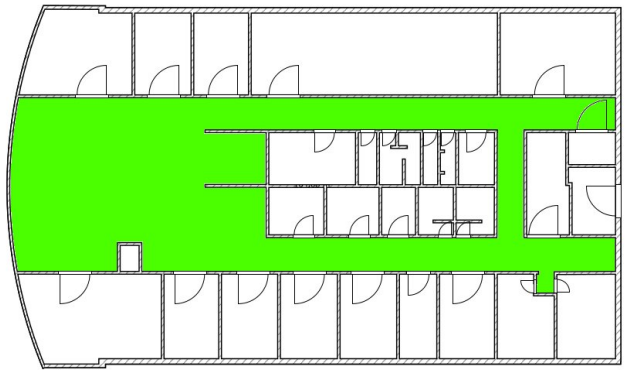


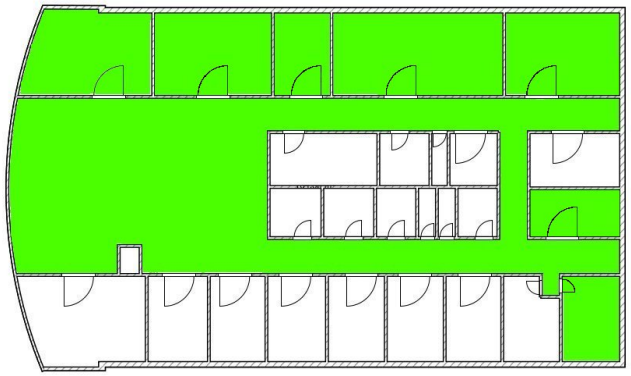
Fig. 1. Structure of the Institute

The dataset do not cover the entire building because the access to the offices and storage rooms is restricted. Thus the dataset cover about 50% of the building which is shown in Figure 2. The number of reference point is proportional to the size of the rooms. The colored places show where the measurements were performed that are the corridors on each floor and the laboratories and lecture hall XXVI on the 1st floor.

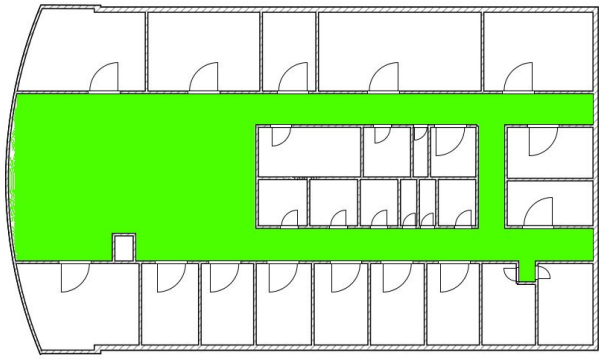
Table I sums up the total and the covered area and their ratio for each floor. Each floor is $1425m^2$. The corridors and the hall are approximately $465.75m^2$ and cover about 33% of the building. Laboratories and a lecture hall are placed on one side of the first floor and they were included into the measurement so the dataset covers about 82% of the first floor. During our measurement, approximately 50% of the institute was recorded in the dataset and the other part of the building was unavailable due to security issues.



(a) Ground Floor



(b) 1st Floor



(c) 2nd Floor

Fig. 2. Covered Area

3.2. Infrastructure

ILONA system was used to record the measurements which is a centralized indoor positioning system. The system is based

TABLE I. Coverage of the building

ZONE	AREA (m^2)	COVERED (m^2)	RATIO
GROUND FLOOR	1425	465.75	32.68 %
FIRST FLOOR	1425	1174.66	82.43%
SECOND FLOOR	1425	465.75	32.68%
INSTITUTE	4275	2106.16	49.27%

on client–server model and built from loosely coupled components. The positioning methods, storage and most of the business logic are implemented at the server side. The clients were designed to run with low resources and to be easy to implement due to the huge variety of the smart phones. Thus, the clients are used to read sensor data and send the measurement to the server for further processing.

The web application was implemented in Java and Spring. It consists of loosely coupled components which have specific tasks such as measurements, positioning, navigation and tracking. Only the measurement component were used by the system to create the dataset. The measurements are stored in a MySQL database [6].

Samsung Galaxy Young GT-S5360 smart phones were used as clients to collect the measurements. Android 4.4.4 run on the client devices with CynagenMod updates. The application performed a measurement in tree steps. First, the corresponding sensor data was read. Then, the read data was converted into the suitable format and were wrapped. Finally, the wrapper object was serialized in JSON and sent to the server via HTTP.

4. DATASET DESCRIPTION

The measurements were recorded in a MySQL database by the ILONA system [6]. The dataset is available as an SQL script so it can be used to create the same environment in other systems. The database was exported in a standard CSV format which is preferred by many data mining tools.

4.1. MySQL Database

Figure 3 shows the schema of the database whose main components are the `Measurement` and the `Position` tables. The measurements, positions and zones have unique identifiers. A zone refers to a disjunct part of the building such as "Laboratory 101", "Office 107B" or "West Corridor". The zones are considered as symbolic positions. In the dataset, each position record contains both absolute and symbolic position. The (x, y, z) absolute coordinates are defined as the distance from a given point of the building.

The measurements contain the position, a time stamp and data from various sensors. The position was set by the users. The following sensors are supported by the model. The magnetometer's measurement is stored as the three component and the rotation. The result of RFID reader and Bluetooth scan are stored as a set of sensed devices. Finally, the WiFi RSSI values are stored as a key–value pair. However the model support the RFID values, the client device had no RFID reader.

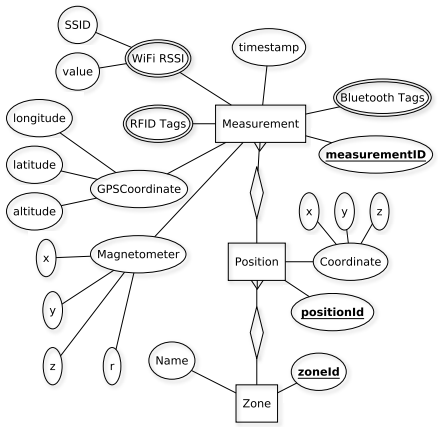


Fig. 3. Schema of the Database

The database store the measurements in a normalized schema and the usage of an RDBMS is required. Because these data is tend to be analyzed, the database was exported and converted into a CSV format which is easily distributable and supported by almost any data analyst tool.

4.2. CSV Format

Each measurement is converted into a record in the CSV format. A measurement record consists of the following three parts which is shown in Figure 4. Measurement description stands in the first part of the record. The second part contains the position information. Measured values are placed at the rest of the record.

The Measurement information part contains an identifier and a time stamp. Client device generated the identifiers by Java `UUID` class thus these identifiers are unique. The timestamp is in the format of `YYYY-MM-DD hh:mm:ss`.

Measurements contain both absolute and symbolic positions. The absolute position is given by three coordinates which are positive values because the origin and the axes was set properly. Symbolic position are given by the identifier and the name of the zone.

The measurements part contains the acquired data from the magnetometer, WLAN card and Bluetooth interface. The magnetometer gives back a 3 dimensional real vector. The WiFi RSSI values are placed between the 11–42 position where each position has a corresponding Access Points. The RSSI value is an integer in $[-255, 0]$ or `null` if the access point was not sensed in the given location. Results of Bluetooth scan are placed in the last part of the record. Each

Measurement Information		Position Information		Measurements		
Id	Timestamp	Absolute	Symbolic	Magnetometer	WiFi RSSI	Bluetooth
1	2	3–5	6– 7	8–10	11 – 42	43 – 65

Fig. 4. Structure of CSV Dataset

position between 43 and 65 has a corresponding device which is identified by a string which contains its name and MAC address. In these positions the record contains `1` if the device was within range and `0` otherwise.

5. EVALUATION OF THE DATASET

The construction of the dataset was made at the end of February of 2016. The measurements were performed in weekend in order to reduce the noise. The dataset contains 1571 records with 65 attributes. These measurements cover about $2000m^2$ which is approximately the half of the building. Each measurement was performed in a unique absolute position. Symbolic positions were defined by 22 zones. During the measurements 31 WiFi access point and 22 Bluetooth enabled device were sensed.

The WiFi Access Point were already installed and the WLAN was used for communication too. The 31 access point seems to be far more than necessary to cover the building. But a few of these access points belong to the nearby student hostel or other buildings. The others are owned by one of the three departments which are placed in the building. Six of these access point were sensed less than 10 times and the most frequently available access point was sensed more than 1200 times. The range of the RSSI values is $[-96, -34]$.

Nine of the Bluetooth devices were installed for the measurement and the others were used by visitors. Each measurement sensed 0 to 10 bluetooth enabled device. On average, about 4 bluetooth enabled device was sensed by a measurement. This phenomena fits the the setup of the measurements and can be explained with the relatively short range of Bluetooth.

The CSV format can be easily processed with data mining tools such as Rapid Miner.

6. CONCLUSION

A one–building, multi–level dataset for hybrid indoor localization was presented in this paper which is available in SQL and CSV formats. The measurements contain data from WLAN and Bluetooth interfaces and Magnetometer sensor which makes it unique because the already existing indoor positioning datasets are focused on only one technology. Both symbolic and absolute positions are defined for each measurement. The same type of devices was used to perform the measurements thus the differences do not caused by the client's sensor set. The dataset covers approximately 50% of a building of $4275m^2$. The dataset allows the comparison and evaluation of hybrid indoor positioning systems.

7. REFERENCES

[1] Hui Liu, Houshang Darabi, Pat Banerjee, and Jing Liu, "Survey of wireless indoor positioning techniques and systems," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 37, no. 6, pp. 1067–1080, 2007.

[2] Gabriel Deak, Kevin Curran, and Joan Condell, "A survey of active and passive indoor localisation systems," *Computer Communications*, vol. 35, no. 16, pp. 1939–1954, 2012.

[3] M. Lichman, "UCI machine learning repository," 2013.

[4] Joaquın Torres-Sospedra, Raúl Montoliu, Adolfo Martınez-Usó, Joan P Avariento, Tomás J Arnau, Mauri Benedito-Bordonau, and Joaquın Huerta, "Ujiindoorloc: A new multi-building and multi-floor database for wlan fingerprint-based indoor localization problems," in *Proceedings of the fifth conference on indoor positioning and indoor navigation*, 2014.

[5] Joaquin Torres-Sospedra, David Rambla, Raul Montoliu, Oscar Belmonte, and Joaquin Huerta, "Ujiindoorloc-mag: A new database for magnetic field-based localization problems," in *Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on*. IEEE, 2015, pp. 1–10.

[6] Richárd Németh Judit Tamás Zsolt Tóth, Péter Magnucz, "Data model for hybrid indoor positioning systems," *Production Systems and Information Engineering*, vol. 7, no. 1, pp. 67–80, 12 2015, HU ISSN 1785-1270.

[7] Roy Want and Andy Hopper, "Active badges and personal interactive computing objects," *Consumer Electronics, IEEE Transactions on*, vol. 38, no. 1, pp. 10–20, 1992.

[8] Andy Ward, Alan Jones, and Andy Hopper, "A new location technique for the active office," *Personal Communications, IEEE*, vol. 4, no. 5, pp. 42–47, 1997.

[9] Gaetano Borriello, Alan Liu, Tony Offer, Christopher Palistrant, and Richard Sharp, "Walrus: wireless

acoustic location with room-level resolution using ultrasound," in *Proceedings of the 3rd international conference on Mobile systems, applications, and services*. ACM, 2005, pp. 191–203.

[10] Tibor Tajti, "Indoor localization with mobile phone," In *Proceedings of the 1st International Conference on Future RFID Technologies* [23].

[11] Binghao Li, Thomas Gallagher, Andrew G Dempster, and Chris Rizos, "How feasible is the use of magnetic field alone for indoor positioning?," in *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*. IEEE, 2012, pp. 1–9.

[12] Zoltan Horvath Tianhang Wu Cui Xuan Andrea Baranyi Sandor Jene Ferenc Brachmann, Ildiko Jenak, "The effects of hardware and software-based signal distortion in multi-platform indoor positioning systems," in *Proceedings of the 11th International Symposium on Location-Based Services*, Georg Gartner and Haosheng Huang, Eds., Nov. 2014.

[13] Silke Feldmann, Kyandoghere Kyamakya, Ana Zapater, and Zighuo Lue, "An indoor bluetooth-based positioning system: Concept, implementation and experimental evaluation.," in *International Conference on Wireless Networks*, 2003, pp. 109–113.

[14] Kyuwon Han and Sung Ho Cho, "Advanced landmarc with adaptive k-nearest algorithm for rfid location system," in *Proceedings of the 2nd IEEE International Conference on Network Infrastructure and Digital Content (ICNIDC'10), Sep 24–26, 2010, Beijing, China*. IEEE Piscataway, NJ, USA, 2010, pp. 595–598.

[15] Jeffrey Hightower, Roy Want, and Gaetano Borriello, "Spoton: An indoor 3d location sensing technology based on rf signal strength," *UW CSE 00-02-02, University of Washington, Department of Computer Science and Engineering, Seattle, WA*, vol. 1, 2000.

[16] Jeffrey Hightower, Chris Vakili, Gaetano Borriello, and Roy Want, "Design and calibration of the spoton ad-hoc location sensing system," *unpublished, August*, 2001.

[17] Lionel M Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek P Patil, "Landmarc: indoor location sensing using active rfid," *Wireless networks*, vol. 10, no. 6, pp. 701–710, 2004.

[18] Paramvir Bahl and Venkata N Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*. Ieee, 2000, vol. 2, pp. 775–784.

[19] Moustafa Youssef and Ashok Agrawala, "Handling samples correlation in the horus system," in *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*. IEEE, 2004, vol. 2, pp. 1023–1031.

[20] Moustafa Youssef and Ashok Agrawala, "The horus wlan location determination system," in *Proceedings of the 3rd international conference on Mobile systems, applications, and services*. ACM, 2005, pp. 205–218.

[21] Chenshu Wu, Zheng Yang, Yunhao Liu, and Wei Xi, "Will: Wireless indoor localization without site survey," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 24, no. 4, pp. 839–848, 2013.

[22] Zsolt Tóth Szabolcs Karsai, "Comparison of wifi-based indoor positioning techniques," In *Proceedings of the 1st International Conference on Future RFID Technologies* [23].

[23] *Proceedings of the 1st International Conference on Future RFID Technologies*, Nov. 2014.