Tech Science Press

MagneFi: Multiuser, Multi-Building and Multi-Floor Geomagnetic Field Dataset for Indoor Positioning

Imran Ashraf[1], Muhammad Usman Ali[2], Soojung Hur[1], Gunzung Kim[1] and Yongwan Park[1,*]

[1]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan-si, 38541, Korea
[2]Department of Computer Science, University of Gujrat, Gujrat, 50700, Pakistan
*Corresponding Author: Yongwan Park. Email: ywpark@yu.ac.kr

Abstract: Indoor positioning and localization have emerged as a potential research area during the last few years owing to the wide proliferation of smartphones and the inception of location-attached services for the consumer industry. Due to the importance of precise location information, several positioning technologies are adopted such as Wi-Fi, ultrawideband, infrared, radio frequency identification, Bluetooth beacons, pedestrian dead reckoning, and magnetic field, etc. Although Wi-Fi and magnetic field-based positioning are more attractive concerning the deployment of Wi-Fi access points and ubiquity of magnetic field data, the latter is preferred as it does not require any additional infrastructure as other approaches do. Despite the advantages of magnetic field positioning, comparing the performance of positioning and localization algorithms is very difficult due to the lack of good public datasets that cover various aspects of the magnetic field data. Available datasets do not provide the data to analyze the impact of device heterogeneity, user heights, and time-specific magnetic field mutation. Moreover, multi-floor and multi-building data are available for the evaluation of state-of-the-art approaches. To overcome the above-mentioned issues, this study presents multi-user, multi-device, multi-building magnetic field data which is collected over a longer period. The dataset contains the data from five different smartphones including Samsung Galaxy S8, S9, A8, LG G6, and LG G7 for three geographically separated buildings. Three users including one female and two males collected the data for various path geometry and data collection scenarios. Moreover, the data contains the magnetic field samples collected on stairs to test multi-floor localization. Besides the magnetic field data, the data from inertial measurement unit sensors like the accelerometer, motion sensors, and barometer is provided as well.

Keywords: Magnetic field dataset; magnetic-field based positioning; smartphone sensors; benchmark analysis; indoor positioning and localization

1 Introduction

The wide proliferation of mobile devices, which constitute smartphones mostly, lead to the inception and initiation of several services related to customer's current location. The projected number of smartphone users in 2020 is approximately 3.5 billion which is a 9.3% increase as compared to 2019 [1]. Location-related services require precise location information of the user to fulfill the user's needs both outdoors and indoors. Although both outdoor and indoor environments have their challenges, for indoor no standard positioning technology like global positioning system (GPS) exists. Despite the reduced accuracy in challenging environments like canyons, tall buildings occupied areas and other GPS denied environments, GPS provides the outdoor position within few meters [2]. However, the accuracy of GPS is severely degraded due to physical barriers of the indoor environment where the frequency is blocked by roofs, walls, and other similar interference sources. To overcome such limitations, several positioning technologies have been introduced or adopted that specifically concentrate on indoor positioning.

The availability of a large number of micro-electrical mechanical sensors (MEMS) in modern smartphones opened the doors to utilize smartphones for positioning. Today, many indoor positioning technologies benefit from such sensors as accelerometers, gyroscope, magnetometer, lux meter, and barometer for indoor positioning. Indoor positioning has significant importance from two perspectives: 80% to 90% of our time is spent indoors and it helps to locate persons for emergency response teams for rescue operations [3,4]. Thereby many indoor positioning technologies emerged such as Wi-Fi, radio frequency identification (RFID), infrared, ultra-wideband, Bluetooth, pedestrian dead reckoning (PDR), and magnetic field-based positioning, etc [5–8]. However, positioning technologies that either utilize the available infrastructure like Wi-Fi access points (APs), magnetic field positioning, or smartphone sensors like PDR, are preferred because they do not require additional infrastructure.

Despite the ease of use, simplicity of implementation, and infrastructure independence, the above-mentioned technologies have disadvantages too. PDR requires a starting or reference position to start with and it provides a relevant position over a short time. For that reason, it is traditionally used with other technologies to enhance positioning accuracy. On the other hand, the positioning accuracy of the Wi-Fi is limited due to several factors. First, the intrinsic limitations of radio wave propagation lead to a change in the received signal strength (RSS) which affects Wi-Fi-based fingerprinting solutions in essence [9]. Second, signal absorption, multipath shadowing, and shading are caused in dynamic environments where the signal is absorbed and reflected by different objects and positioning accuracy is affected. Third, environments with human mobility cause human body loss and changes the RSS value and positioning accuracy as a result. Fourth, the diversity in the antenna design and hardware used in heterogeneous devices also affect the RSS value which can cause degradation in positioning [10].

Magnetic field-based indoor positioning has several additional benefits than that of both PDR and Wi-Fi. First of all, like Wi-Fi, it uses the subsisting infrastructure and (commercial-off-the-shelf) applications to estimate the current position of the user. Secondly, it does not require any infrastructure and works with the embedded sensors in the smartphone. Even though Wi-Fi is called infrastructure-less technology, it still needs Wi-Fi APs to perform positioning, magnetic field is the earth's natural phenomenon. Thirdly, the magnetic field data has shown more long-term stability than that of Wi-Fi signals and its mutation over time is very slow. Moreover, it is less affected at all in the environment with human mobility. As a result, it has become an area of wide interest and several frameworks and approaches have been introduced recently. However, the lack of a publicly available dataset for magnetic field positioning makes it very difficult to analyze the proposed approaches on a common

standard. Several important aspects of magnetic field-based positioning are unexplored and under-researched including the smartphone heterogeneity, user diversity, influence of user smartphone's complex attitude, space diversity, etc. To resolve these issues, this study introduces a magnetic field dataset and makes the following contributions

- A comprehensive overview of the existing publicly available magnetic field datasets is presented which points out their limited capacity to evaluate a magnetic data-based positioning approach.
- A large dataset, MagneFi is introduced which contains the data collected over 5 years. Apart from the magnetic field data, the data from accelerometer and motion sensors are also provided to track user activities and walk direction.
- Three persons, one female, and two males are involved in data collection from 5 geographically separated buildings with different indoor settings. For each building, 3 different path trajectories are followed to collect the data. These buildings include both new buildings of concrete, as well as, old brick buildings where the magnetic field anomalies are less unique than those of former buildings.
- To analyze the floor identification capability of the magnetic field data, the data are collected from multiple floors in the same building. The data for stairs are collected to identify 'stairs up' and 'stairs down' events. Moreover, to validate the performance of floor detection, barometer data are also provided.
- Five smartphones are used in the data collection process including Galaxy S8, LG G6, Galaxy A8, LG G7, and Galaxy S9+. Three smartphone orientations are covered to collect the data such as 'navigation', 'call listening', and 'swinging' for each smartphone. For investigating the impact of human mobility on the magnetic field data, the data are collected in a large public exhibition hall considering different levels of human mobility.

The rest of the paper is structured as follows. Section 2 provides the description and critical analysis of the publicly available magnetic field datasets. The details of the data collection process, smartphone used for the collection, and scenarios are discussed in Section 3. Section 4 explains the structure of the dataset, its attributes, and the process of using it for positioning. Discussion is given in Section 5 while Section 6 concludes the study. Owing to the use of several acronyms in this study, Tab. 1 shows a list of the acronyms.

Table 1: A summary of acronyms used in the current study

Acronym.	Abbreviation
AP	Access Point
BE	Business and Economic Engineering
COEX	Convention Exhibition Center
CS	Computer Science
EE	Electrical Engineering
GPS	Global Positioning System
GPU	Graphical Processing Unit
hPa	Hecto Pascal
IMU	Inertial Measurement Unit
IPIN	Indoor Positioning and Indoor Navigation
IT	Information Technology

(Continued)

Table 1: Continued

Acronym.	Abbreviation
LCD	Liquid Crystal Display
MEMS	Micro Electro Mechanical Systems
PDR	Pedestrian Dead Reckoning
RFID	Radio Frequency Identification
RIC	Regional Innovation Center
RSS	Received Signal Strength

2 Related Work

Several datasets have been presented before that provide the magnetic field data to test indoor positioning approaches. These datasets can be categorized as publicly available and access restricted or available on request. The latter datasets are collected and used in several research works and lack various aspects of the magnetic field data. Moreover, they are specifically collected for testing the proposed approaches in the associated research papers. Therefore, these datasets are not discussed here. The former category, on the other hand, collects the data for public access and covers various data collection scenarios, path geometry, indoor environment, users, and heterogeneous devices. The datasets which are included in the discussion here are only publicly available datasets. So, all publicly available magnetic field datasets are discussed in this section.

A magnetic field dataset called UJIIndoorLoc-Mag is presented in [11] at the indoor positioning and indoor navigation (IPIN) conference 2015. The data are collected using two smartphones of different brands including a Google Nexus 4 and LG G3, both phones have Android 5.0. A computer laboratory separated by desktop tables is used to collect the data along the corridors and intersections. The dimension of the laboratory is 260 m² and 8 corridors are used for the data. Data collection scenarios include the user walking between designated starting and ending points which are manually marked. For each segment of the data, the starting, the ending, and the coordinates where the user makes turns are recorded along with the data. Additionally, the data from the accelerometer and motion sensors are recorded to analyze the walking speed of the user and perform resampling if necessary.

The research [12] presents a dataset that comprises the magnetic field and Wi-Fi data collected from a smartphone and smartwatch. For this purpose, Sony Xperia M2 and LG W110G Watch R are used by a user for the data collection. The collection of Wi-Fi data and the magnetic for the same ground truth positions makes it possible to perform hybrid solutions for enhanced positioning accuracy. Besides, accelerometer and orientation sensors data help to determine user walking patterns and walking direction. The data are collected along a long corridor with two offices at the start and end of the corridor and a hall in the middle. The path geometry is complex with frequent turns across the corridor and hall.

A multi-source dataset is introduced in [13] that contains the data both from the magnetic field, as well as, the Wi-Fi. In addition, the data from IMU sensors such as gyroscope and accelerometer is provided to track the user's device orientation. The indoor setting contains corridors and intersections containing turns in various directions. Besides, the data are collected in two directions. Single smartphone, i.e., Google Nexus 4 operating on Android 5.0.0 is used for the data collection, and 8 corridors and 19 intersections are scanned for the data. The data are collected during a period of six months from April to September 2014 to analyze the time-related mutation of the magnetic field data.

The authors provide a dataset to perform hybrid indoor positioning in [14]. The dataset contains the magnetic field, Wi-Fi, radio frequency identification (RFID), and Bluetooth data collected using Samsung Galaxy Yong GT-S5360 operating on Android 4.4.4. The developed module can provide the data for all the technologies mentioned above, however, the unavailability of an RFID scanner in the smartphone leads to empty records for RFID. For the Wi-Fi and Bluetooth, 30 access points (APs) and 9 Bluetooth devices are installed in the data collection environment. Ground truth positions are calculated manually whereas the collected data is sent to the server for processing and storage. The dataset also contains the data from multi-levels making it possible to perform multi-floor positioning and localization.

MagPIE dataset is presented in [15] at IPIN 2017 which contains the magnetic field and IMU sensors data. Two different buildings are used for data collection with a different indo1or setting and path geometry. For data collection, two platforms are used including a wheeled robot and a human surveyor. Same smartphones are used with the robot and human surveyor, i.e., Motorola Moto Z Play and Lenovo Phab 2 Pro. The smartphones are mounted on the robot in a vertical position facing the screen parallel to the body of the robot while the human surveyor holds them in the navigation mode. Besides the data collection along an empty path, several ferromagnetic materials like an industrial fan, furniture, and personal computers, etc. are placed alongside the path to analyze temporal changes in the magnetic field data.

The above-cited research works present various datasets with the primary aim of providing the magnetic field dataset to evaluate the positioning approaches and compare their performance on a benchmark. However, these datasets lack in several ways and cannot cover all the aspects of magnetic field positioning. First, the magnetic field data changes from one smartphone to another due to the sensitivity, noise, and error tolerance of the embedded magnetic sensor, and proper evaluation requires the data from multiple smartphones. Predominantly, the majority of the datasets contain data from a single smartphone only. Secondly, the impact of the user's height cannot be analyzed if a single user collects the data which is the case with many of the datasets described above. Thirdly, the path trajectory is simple and does not fit the real-life scenarios of shopping malls, train stations, and airports, etc. to depict the positioning performance of the approaches for the real world. Fourthly, the long-term behavior of the magnetic field data is not recorded which shows mutation over time. Last but most important, the complex attitude of the user during walking or standing is not recorded where the user moves his phone into different orientations. All the datasets are collected with a single orientation only. A summary of the advantages and limitations of the discussed research works is provided in Tab. 2.

Table 2: A summary of advantages and disadvantages of magnetic field benchmark datasets

Ref.	Advantages	Disadvantages
[11]	-Two smartphones are used -Corridors and intersections are covered -High accuracy for data collection coordinates -IMU data are provided	-Only one orientation is used -Data collection in a single building -No multi-level data -Single user data collection -Long term mutation is not recorded

(Continued)

Table 2: Continued

Ref.	Advantages	Disadvantages
[12]	-Wi-Fi and magnetic field data are collected -Data from a smartphone and smartwatch -Data for corridors, offices, and one hall -Complex path trajectory is used -IMU data are available	-Fixed orientation of the smartphone -No multi-level and multi-building data -Single user data collection -Long term mutation is not recorded
[13]	-Magnetic field and Wi-Fi data -Data collection in two opposing directions -Path includes corridors and intersections -Records the data during 6 months -IMU data are available	-Fixed orientation of the smartphone -Single user and single device data collection -Single building data collection -Long term mutation is not recorded
[14]	-Wi-Fi, Bluetooth, and magnetic field -Deployment of APs and Bluetooth beacons -Data for multi-level corridors and intersections for large area -Lightweight client-side module -IMU data are available	-Fixed orientation of the smartphone -Data collection in a single building -Single user and device data collection -Long term mutation is not recorded
[15]	-Magnetic field from multiple buildings -Data using a robot and human surveyor -Two smartphones are used -Records temporal changes of magnetic field -IMU data are available	-Fixed orientation of the smartphone -Single user data collection -Path trajectory for data collection is not available -Long term mutation is not recorded

3 Data Collection Process

The data collection process involves four smartphones, three buildings, three different phone holding orientations, and path trajectories. The details of each of these characteristics are provided separately in the following sections.

3.1 Smartphones Used for Data Collection

For analyzing the impact of using heterogeneous smartphones on the positioning accuracy with the magnetic field data, four devices are used to collect the data. Tab. 3 shows the names and descriptions of the sensors used for data collection. Six smartphones from two different companies are used to gather the data including Samsung and LG. Besides different models of the same brands are used as well to analyze the impact of model change on the magnetic field data. For example, Galaxy S8 and Galaxy S9+ are used by Samsung, and G6 and G7 are used by LG companies. Since the vendors and the model of the used magnetometer are different, it is important to give the specifications like the sensitivity, temperature, etc. It will be helpful to analyze the positioning accuracy with the brands of smartphones that use different vendors and different models of the same vendors for the embedded magnetic sensor.

Table 3: Names and descriptions of smartphone sensors used to collect the magnetic field data

Sensor	Vendor	Description
Samsung Galaxy A8		
SM-A530 N Galaxy A8	Samsung	Octa-core, Adreno 540 GPU, Android 9.0. (Pie), 4 GB RAM
Magnetometer (AK09918C)	Asahi Kasei	3-axis, 16-bit, sensitivity 0.15 mT/LSB, temperature −30 to +85°C, 6.0 mA [16]
Accelerometer (LSM6DSL)	Asahi Kasei	3-axis, 16-bit, sensitivity 0.061 mg/LSB, temperature −40 to +85°C, 0.25 mA [17]
Gyroscope (LSM6DSL)	STMicroelectronics	3-axis, 16-bit, sensitivity 125 mdps/LSB, temperature −40 to +85°C, 6.1 mA
Samsung Galaxy S8		
SM-G950 N Galaxy S8	Samsung	Octa-core, Adreno 550 GPU, Android 8.0 (Oreo), 4 GB RAM
Magnetometer (AK09916C)	Asahi Kasei	3-axis, 16-bit, sensitivity 0.15 mT/LSB, temperature −30 to +85°C, 6.0 mA [18]
Accelerometer (LSM6DSL)	STMicroelectronics	3-axis, 16-bit, sensitivity 0.061 mg/LSB, temperature −40 to +85°C, 0.25 mA [17]
Gyroscope (LSM6DSL)	STMicroelectronics	3-axis, 16-bit, sensitivity 125 mdps/LSB, temperature −40 to +85°C, 6.1 mA
Samsung Galaxy S9+		
SM-G965U Galaxy S9+	Samsung	Octa-core, Adreno 630 GPU, Android 10, 6 GB RAM
Magnetometer (AK09916C)	Asahi Kasei	3-axis, 16-bit, sensitivity 0.15 mT/LSB, temperature −30 to +85°C, 6.0 mA [18]
Accelerometer (LSM6DSL)	STMicroelectronics	3-axis, 16-bit, sensitivity 0.061 mg/LSB, Temperature −40 to +85°C, 0.25 mA [17]
Gyroscope (LSM6DSL)	STMicroelectronics	3-axis, 16-bit, sensitivity 125 mdps/LSB, Temperature −40 to +85°C, 6.1 mA
LG G6		
LGM-G600L LG G6	LG	Quad-core, Adreno 530 GPU, Android 7.0 (Nougat), 4 GB RAM
Magnetometer (AK09915C)	Asahi Kasei	3-axis, 16-bit, sensitivity 0.15 mT/LSB, temperature −30 to +85°C, 6.0 mA [19]
Accelerometer (BMI-160)	Bosch	3-axis, 16-bit, Temperature −40 to +85°C, 0.18 mA [20]
Gyroscoope (BMI-160)	Bosch	3-axis, 16-bit, Temperature −40 to +85°C, 0.9 mA [20]

(Continued)

Table 3: Continued

Sensor	Vendor	Description
LG G7		
LM-G710 N LG G7 ThinQ	LG	Octa-core, Adreno 630 GPU, Android 10.0, 4 GB RAM
Magnetometer (AK09918C)	Asahi Kasei	3-axis, 16-bit, sensitivity 0.15 mT/LSB, temperature −30 to +85°C, 1.1 mA [16]
Accelerometer (IAM-20680)	TDK-Invesense	3-axis, 16-bit, Temperature −40 to +85°C, 0.24 mA [21]
Gyroscoope (IAM-20680)	TDK-Invesense	3-axis, 16-bit, Temperature −40 to +85°C, 0125 mA [21]

3.2 Data Collection Application

Due to the unavailability of a smartphone app for data collection in a custom environment, a data collection app is required. In this regard, a custom data collection app is designed using the Android Studio 3.5, released on 9 Aug 2019. Fig. 1 shows the snapshots of various windows of the custom-built app. Images for all data collection scenarios, as well as, the coordinates for ground truth locations are fed into the app. App uses different application programming interfaces of the Android operating system to collect the data from IMU sensors of the smartphone.
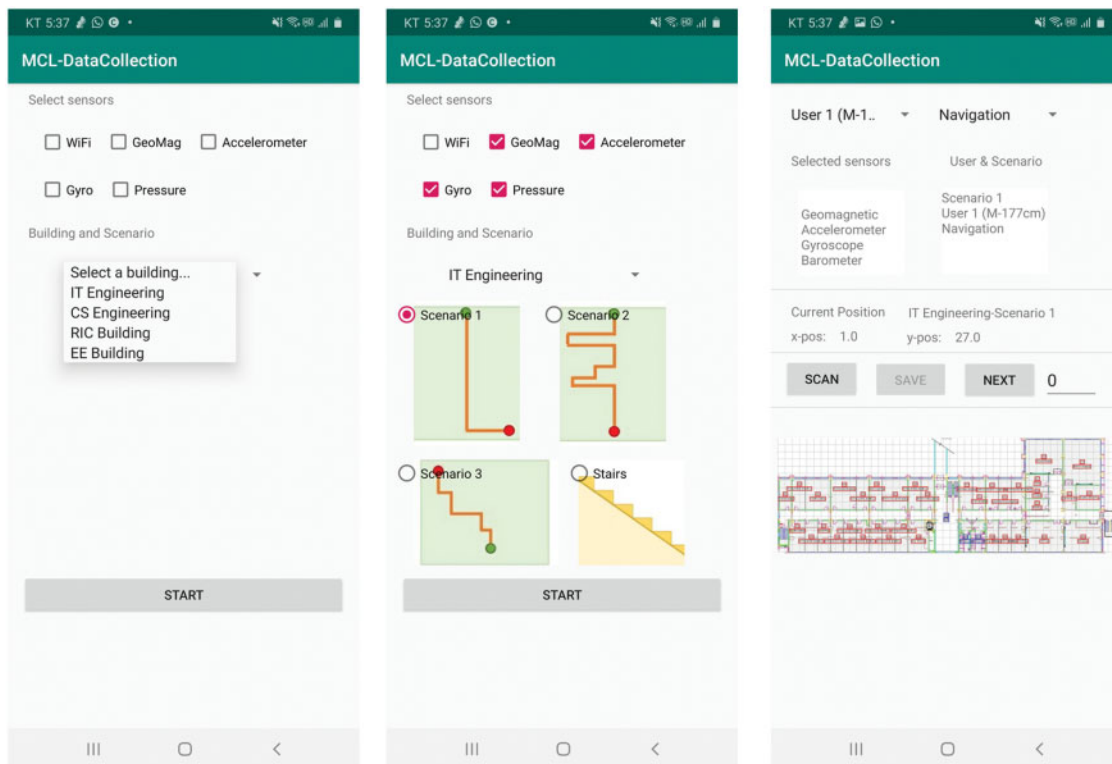
Figure 1: Snapshots of the Android app used for the data collection

3.3 Data Collection with Different Orientations

Predominantly, the fixed orientation of the smartphone is used for evaluating the magnetic field-based positioning approaches. The magnetic field data changes substantially with the change in smartphone orientation, so a navigation mode is adopted where the smartphone is kept in front of the user's body with the user moving in various directions (left, right, back) but the orientation is not changed. But in the real-life, the attitude of the user is complex where the user moves the smartphone in different orientations like call listening, phone in the pocket and swinging in hand, etc. To model these orientations of the user, this study adopts the three most common smartphone orientations, for the data collection.

Navigation mode is the most commonly used orientation for magnetic field-based indoor positioning approaches [22–24] where the phone is hand-held in front of the body. The user can move in any direction but the orientation of the phone is kept constant. Fig. 2a shows the position and axes of the magnetic field data for this orientation. The phone is held in front of the body approximately at the navel height of the user with the LCD facing upwards.
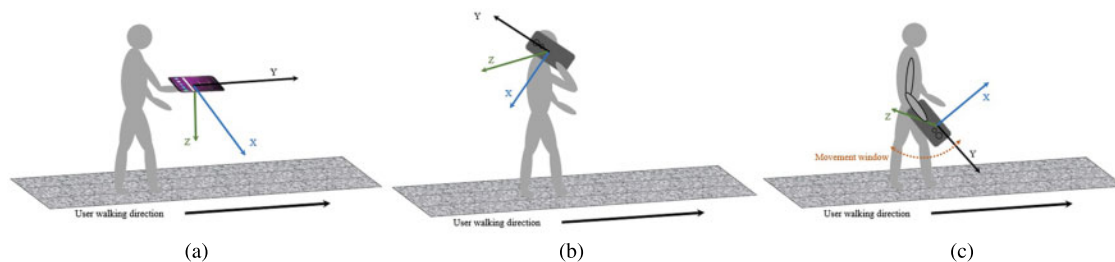

(a)   (b)   (c)

Figure 2: Smartphone orientations used for the data collection, (a) Navigation orientation, (b) Call listening orientation, and (c) Phone swinging orientation. The x, y, and z-axis show the magnetic field components

Call listening mode is seldom used for magnetic field-based positioning, however, for PDR approaches it has been a common mode of investigation [25,26]. In this mode, the phone is held beside the ear in an upside-down position where the phone is tilted clockwise. The LCD of the smartphone is facing inside towards the user's face. Fig. 2b provides the axes details for magnetic field components in this orientation.

Phone swinging orientation is also not a preferred orientation to test the magnetic field-based positioning approaches [27,28]. This orientation is complex and the nature of the magnetic field data is difficult to handle while walking. The continuous swing of the arm to forward and backward directions generates the data that changes due to the arm swing and complicates the positioning process. The LCD of the smartphone is facing inside towards the user's leg in this orientation. Fig. 2c shows the position of the phone and magnetic field components for this model. The 'movement window' in Fig. 1c indicates the space window where the phone moves forward and backward when the user walks with the phone.

3.4 Spatial Diversity and Path Trajectory for the Magnetic Field Data Collection

Path trajectory is an important element to evaluate the performance of the magnetic field-based indoor positioning approaches as the complexity of the path has a direct influence on the accuracy. Predominantly, simple paths with few turns are followed in many positioning approaches [29,30].

Several initial magnetic fields-based positioning works followed simple path trajectories for feasibility studies for both handheld devices and robot navigation [23,31,32].

Complex path trajectories should be added for the data collection, which includes turns in various directions, as well as, the data from various indoor set up such as large spaces, rooms, and corridors, etc. For the current dataset, three different buildings in the university are used. For each building, different scenarios are designed ranging from simple to complex paths. Fig. 3 shows the scenarios and path geometry followed for the information technology (IT) building in the Yeungnam University. The black and green circles on the map indicate the starting and ending positions of the data collection, respectively.

(a)

(b)

Figure 3: (Continued)

(c)
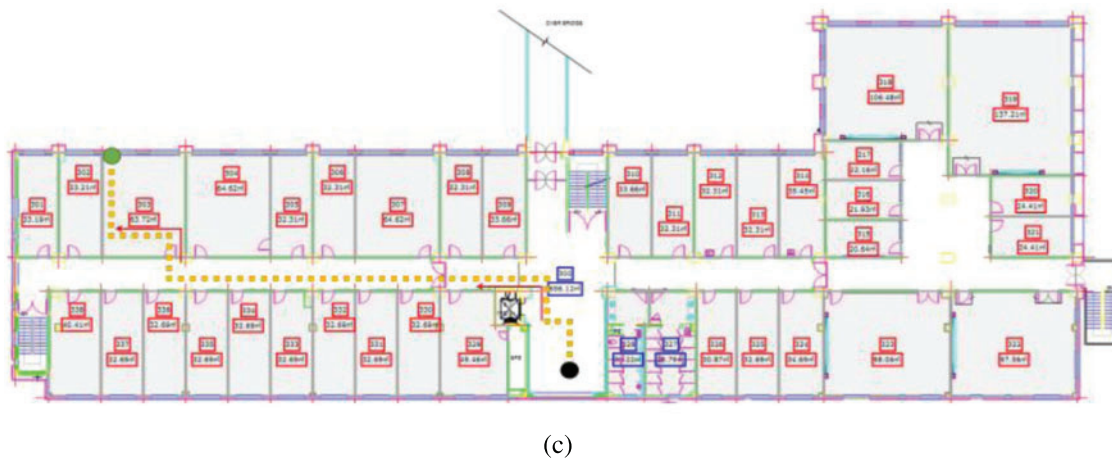
Figure 3: Path geometry followed for the magnetic field data collection in the IT building

In addition to the three scenarios followed for the data collection, the data are collected for the multi-floor environment for the IT building as well. It involves the data collection on the stairs for a three-floor building. The dimensions of the stairs are shown in Fig. 4. Each step has a dimension of 33×15 cm^2 and floors are connected with a 'V' shape stairs model with each leg having 10 steps. The roof is made of concrete while a ceiling is added at a height of 241 cm from the floor level. For the magnetic field data, the data are collected at each 2nd step of the stairs. Moreover, the data are collected in both downward and upward directions for 'stairs down' and 'stairs up' events. Similarly, four other buildings are involved in the data collection process including the computer science (CS) building, electrical engineering (EE), regional innovation center (RIC), and business and economics (BE) department building.

(a)   (b)   (c)
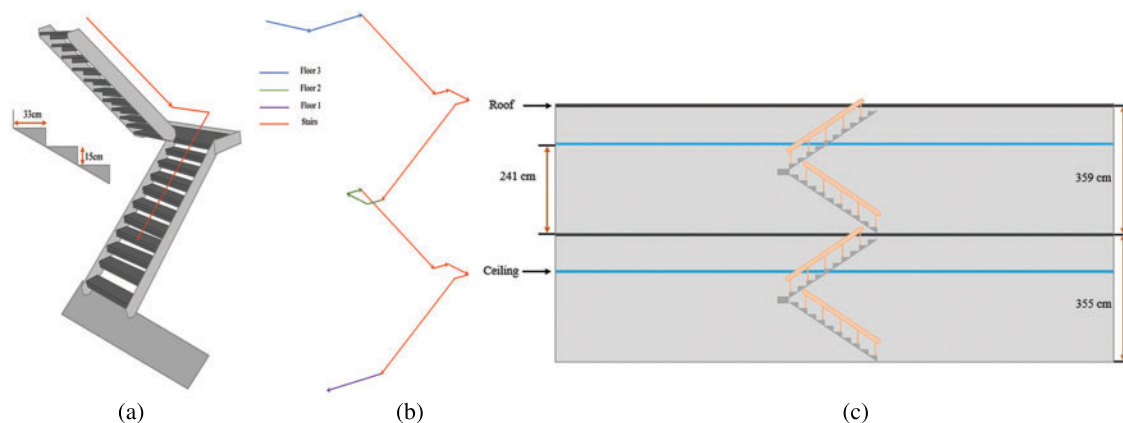
Figure 4: Path geometry for multi-floor stairs between floor 1 and floor 3 of the IT building, (a) Number and size of stairs steps, (b) Direction followed during the data collection involving floor surface and stairs, and (c) The height of each floor and the measurements of roof and ceilings

Fig. 5 shows the paths followed to collect the IMU sensors and magnetic field data in CS, EE, and RIC buildings, respectively. The data are collected in a grid where each point of the grid is separated by a distance of 1 m. The data collected in different buildings are helpful to test the performance

of various state-of-the-art approaches in different indoor settings. However, the performance of a positioning approach in a room like a laboratory or an office is also important. Predominantly, previous approaches and datasets cover only the former aspect while the latter is ignored. This study covers this aspect as well and collects the data in a computer laboratory of $9 \times 7\ m^2$ for this purpose. The data are collected at marked location points at a distance of 1 m as shown in Fig. 6.
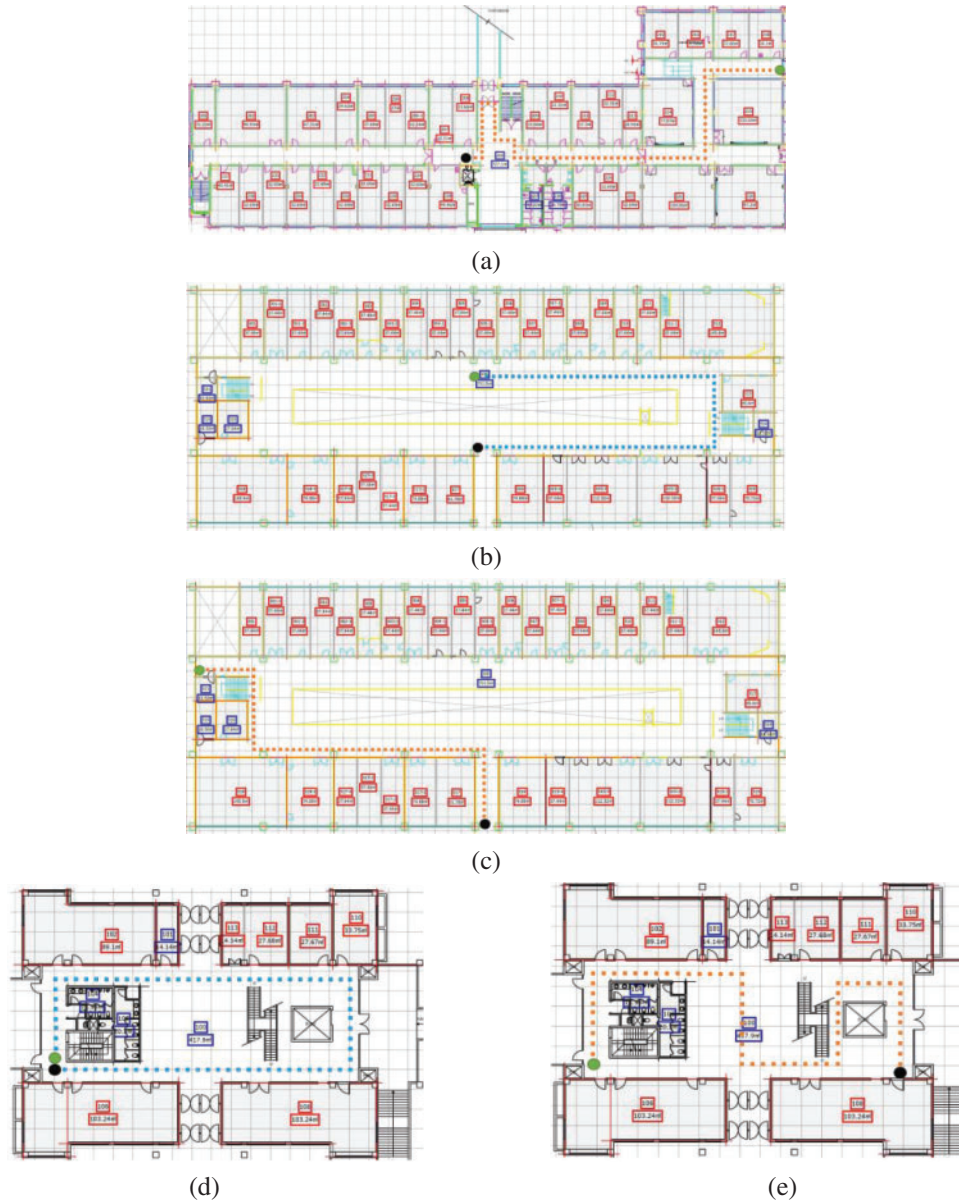

(a)

(b)

(c)

(d)   (e)

Figure 5: Path geometry used CS, EE, and RIC buildings, (a) Scenario for CS building, (b) Scenario 1 for the EE building, (c) Scenario 2 for data collection in EE building, (d) Scenario 1 for IACT, and (e) Scenario 2 for IACT building
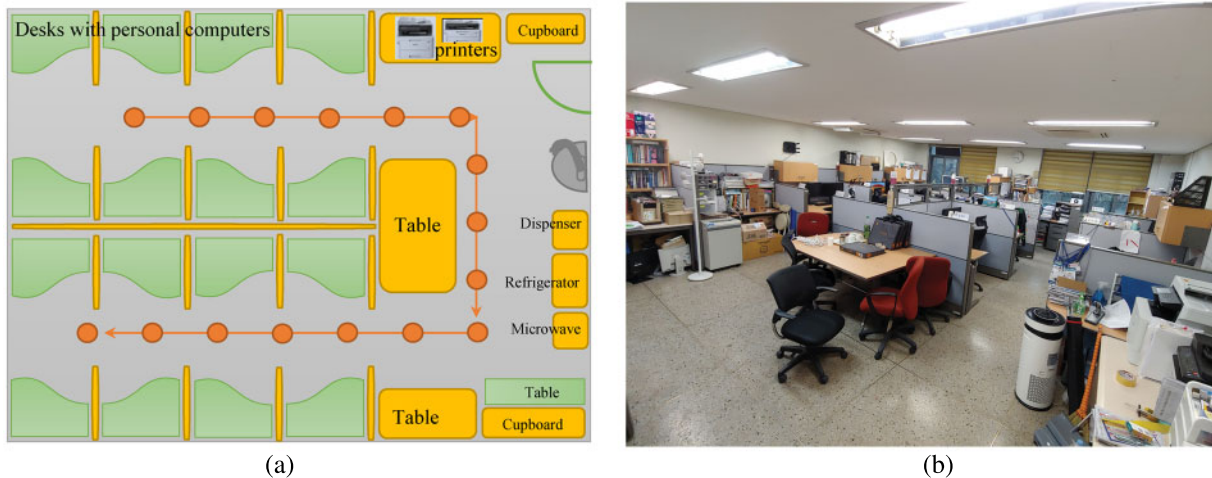
Figure 6: The path followed for data collection in a laboratory, (a) Map of the data collection points separated by 1 m, and (b) Snapshot of the room showing indoor environment

Two additional scenarios considered for the data collection are a public exhibition hall with different levels of human mobility and buildings without concrete. For the former, a public exhibition hall called Starfield COnvention EXhibition (COEX) center is selected. Located in Seoul city of Korea, the COEX center comprises many seminar and conference rooms and exhibition halls. The hall which is selected for the data collection has a large of $108 \times 106 \, m^2$. The hall area is covered, however, it does not has any concrete walls. During the data collection, an exhibition was in place and the hall area was divided into small units by the stalls and tables placed by different companies, as shown in Fig. 7. For data collection, two human mobility scenarios are covered.

- Small to medium human mobility, where the number of people in the hall varies between 50 to 350 at the time of the data collection.
- High human mobility, when more than 350 people are present in the hall at the time of data collection.

The BE building is selected for the data collection due to its construction structure which involves using red bricks and lacking concrete. Experiments show that the buildings containing concrete exhibit high magnetic field variation than those that do not contain concrete. It is due to the presence of iron in the concrete which is a ferromagnetic material and interferes with the earth's natural magnetic field. For evaluating the performance of the magnetic field-based positioning, the data from such buildings are not available. So this study specifically considers this building for the data collection.

The BE building has an uneven structure of the floor which includes a few steps of stairs on the floor. Consequently, the floor number is not changed, but the up and down level of the floor affects the positioning performance. Using the data from such environments is also helpful to determine users' activities of going up and down. Fig. 8 shows the path followed to collect the data for uneven floor structure from BE building.

Fig. 8a shows the path used for data collection in BE building while Fig. 8b shows that the path is subdivided into two paths. For better presentation path 1 and path 2 are shown as magenta and blue circles, respectively, in Fig. 8.
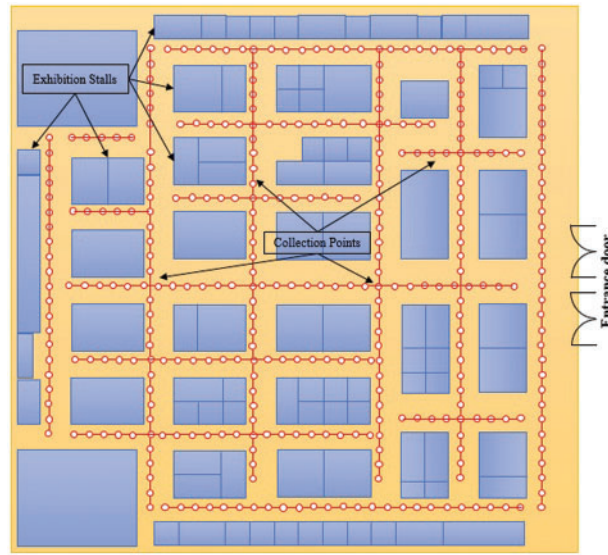
Figure 7: Data collection points at COEX. The area is separated by stalls. Red circles show the ground truth points used for the data collection

| (a) | (b) |

Figure 8: Data collection path in BE building (a) Path followed for the data collections, and (b) Details for the path and stairs. For better illustration, the path is split into path 1 and path 2

Data collection starts from Path 1 which has two descending stairs containing 8 steps each and the height of each step is 13.30 cm. Path 2 has three ascending stairs with 8, 5, and 8 steps, respectively. Path 1 and path 2 are used to explain the path geometry; for the data collection, both paths are connected. Colors for different path segments are used to indicate the height of the path segments. For example, the end of path 1 and the start of path 2 shown in red color, has the same height. Similarly, green segments in path 1 and path 2 have the same height while the blue and grey segments have different heights from other path segments.

3.5 Data Collection Strategy

The magnetic field data collection involves two strategies for all the buildings and path geometries including static data and continuous data. Where the former strategy involves collecting the data on designated location points, called 'ground truths'. The data are collected by standing stationary at the given points and 125 to 150 data samples are collected at a sampling rate of 10 *Hz/s*. In the latter strategy, on the other hand, the data are collected while walking on the designated path at a consistent speed.

The data collection strategy is adopted to serve multiple purposes. First, the data from the static collection strategy can be used to build the fingerprint databases which is the most frequently used positioning approach for magnetic field-based indoor positioning. It requires multiple samples for the same location which are normalized to a single feature vector to build the fingerprint database. Static data from other users can be used for testing the positioning approach. Second, continuous data is traditionally used for locating the user in real-time. It contains the motion sensors data including gyroscope and accelerometer which helps to enhance the positioning accuracy by estimating the traveled distance and direction of the user. Third, although the data are collected at a consistent speed, the walking speed of the same user varies while turning to different directions. Similarly, since the data collection involves various users, the walking speed varies from one user to another. The data collected at various speeds helps in analyzing the impact of walking speed for magnetic field-based positioning approaches. Additionally, the data from the barometer helps to determine the floor as well as can be used for different events like going up or down and stairs detection. Last but the most important is the analysis of the walking pattern also called gait analysis using the magnetic field data and IMU sensors data from different users.

4 Structure of Magnetic Field Benchmark Dataset

The dataset is recorded as Excel files and stored in folders and subfolder structures for easy and efficient access. Fig. 9 shows the directory structure of the MagneFi dataset. As described previously, the data are collected using static and continuous strategies, they are also stored in separate folders. The main folder contains 'static data' and 'continuous data' sub-folders for the data collected using two different strategies. Both strategies are adopted for each building selected for the data collection including CS, EE, IT, and IACT buildings. 'Navigation', 'call listening', and 'swinging' orientations are used for data collection where the definitions for these orientations are provided in Section 3.4. The data collected on the stairs are put in a separate folder, called, 'stairs'. For data collection involving stairs, only the 'navigation' style is used by multiple users.

The data for each orientation is stored in separate folders, where each folder contains the data for different scenarios designed for data collection, as shown in Fig. 9b. Each scenario involves the data collected from five different smartphones including Galaxy A8, LG G6, Galaxy S8, LG G7, and Galaxy S9+ smartphones. Three users collect the data for each smartphone where the details for users' heights and gender are given as follows

- User 1: male with a height of 177 *cm*
- User 2: male with a height of 174 *cm*
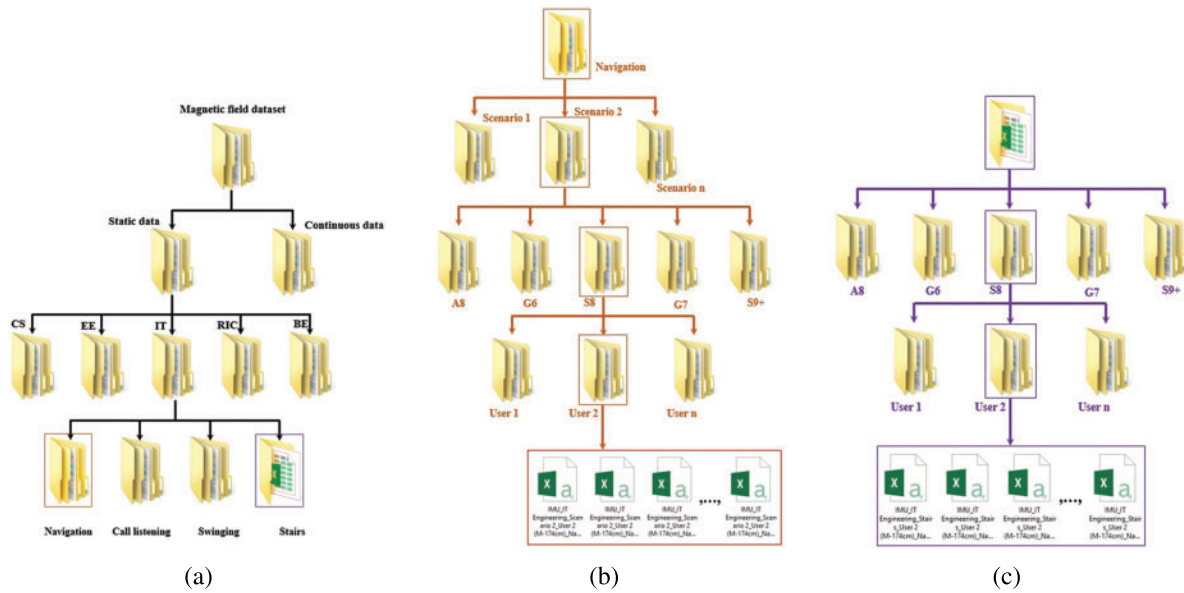- User 3: female with a height of 168 *cm*

(a)   (b)   (c)

Figure 9: Structure of the dataset, (a) Main directory, (b) Directory structure for navigation orientation data collection, and (c) Directory structure to access the data for stairs in IT building

The structure for storing the data for various buildings and orientation styles is identical to that of what has been described. Under the folder named, 'User', the data files are stored as Excel sheets. Each sheet contains data in 16 columns as shown in Fig. 10, where the detailed description of each column is given.

Figure 10: Snapshot of the Excel worksheet containing the magnetic field and IMU sensors data

1. Indicates the time stamp which is taken using the current time from the smartphone used for the data collection. Time column has a format 'yyyy.mm.dd hh:mm:ss'.
2. Provides the x coordinate of the location point at which the data are collected. It indicates the local coordinate according to the map and dimension information of the building.
3. Provides the y coordinate of the location and indicates the local coordinate. The x and y coordinates are available for the static data while for the continuous data these coordinates are not recorded. For continuous data, the users walk between the starting and ending points at a consistent speed and the data are recorded continuously.

4, 5, and 6 give the magnetic field data for its x, y, and z components, respectively. Magnetic x, y, and z components represent the North, West, and vertical component of the magnetic field, respectively. Total magnetic intensity F is not provided, however, it can be calculated using

$$F = \sqrt{Mag_x{}^2 + Mag_y{}^2 + Mag_z{}^2} \tag{1}$$

7, 8, and 9 indicate the acceleration recorded from the accelerometer of the smartphone and represent the x, y, and z acceleration. The values for x, y, and z acceleration are given in m/s². Total acceleration can be measured using

$$A = \sqrt{Acc_x{}^2 + Acc_y{}^2 + Acc_z{}^2} \tag{2}$$

10, 11, and 12 show the values for the gyroscope x, y, and z-axis, respectively. The values show the angular velocity and are given in *radians/s*.

13, 14, and 15 provide device orientation with its x, y, and z-axis. Orientation values are used both in the static data and continuous data to track the orientation of the user as these values are different for various events like call listening, phone swinging, and navigation, etc. Similarly, when changing directions in the walking mode, the orientation vector can be used to determine the direction of the user like right or left. The difference between the 'Gyro' and 'Orn' columns is that the former shows the raw values from the gyroscope while the latter is calculated using rotation matrix, accelerometer reading, and magnetometer reading.

16. Indicates the atmospheric pressure measure from the barometer of the used smartphone. The pressure is given in hecto Pascal (*hPa*) which is equal to the millibar pressure unit.

5 Preliminary Results Using Dataset

To show the use of the presented dataset, approach [33] is used for obtaining the positioning results. Results are obtained using the magnetic field and IMU sensors data for this purpose. As shown in Fig. 11, results indicate that using the presented dataset, accurate indoor positioning can be performed. Results in Fig. 11 are for EE building using Galaxy S9+, S8, and LG G6 smartphones.
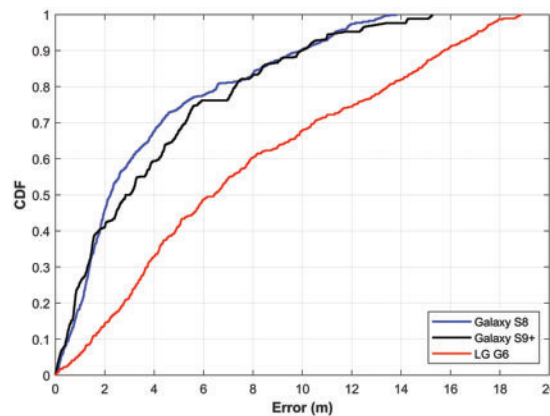
Figure 11: CDF for positioning error using the presented dataset

6 Discussions

Magnetic field-based indoor positioning gained wide attention from academia and industry during the last few years and a large number of positioning frameworks and approaches have been presented. However, such approaches have not been tested on a standard benchmark dataset and individually collected datasets have been utilized. The problem with this approach is two-fold. First, the results of the proposed frameworks and methods cannot be reproduced by other researchers. Second, comparing the performance is not possible due to the unavailability of the dataset used for such approaches. As a result, providing a common benchmark dataset for magnetic field-based indoor positioning is of great significance. Despite previously available datasets, the shortcomings in the collected data do not ensure the analysis of all important aspects of the magnetic field data.

To overcome the limitations of the publicly available datasets, this study focuses on those aspects of the magnetic field data which have been missed. This study first highlights the important aspects of the magnetic field data concerning its use for indoor positioning and localization and then defines various strategies to incorporate those aspects into the data collection process. The first element involves using multiple smartphones for the data collection contrary to the previous studies where predominantly a single smartphone is used for the same purpose. Providing the magnetic field data from multiple smartphones enables the testing of positioning approaches and analyzing the impact of using heterogeneous smartphones. Current experiments and previous studies indicate that the positioning accuracy is different when used with different smartphones.

By and large, the available benchmark datasets contain the magnetic field data from the single orientation of the smartphone, and various events of the user such as call listening, and phone swinging are not recorded. The magnetic field data changes substantially when the orientation of the smartphone is changed and the positioning accuracy dramatically changes with it. In the real-life, the user attitude is complex and involves multiple orientations of the smartphone. So, using a fixed orientation for magnetic field positioning is not very practical when its practical implementation is considered. This study provides the data for three of the most commonly used orientations of the smartphone including navigation, call listening, and phone swinging. Furthermore, different users have collected the data as the change in the magnetic field data has been reported for the users of different heights.

The positioning accuracy of the magnetic field data is reported differently in different buildings given that the same positioning approach is utilized. It indicates that the distribution of the magnetic field is different in different buildings and the uniqueness of the magnetic field fingerprints varies from one place to another. Indoor settings, the geographical position of the building, and the presence of the different amounts of ferromagnetic materials cause the different distribution of the magnetic field data. Subsequently, the magnetic field data from various buildings that are geographically separated and have different indoor environments is important to the researcher community. For this purpose, the current study collects the data from four different buildings, as well as, various path trajectories are followed in each building.

Magnetic field data can be leveraged to identify specific floors as reported in [25] and for further experimentation, the data from the multi-floor environment are provided in the current study. In addition, the data are collected on the stairs involving three floors to investigate the 'stairs up' and 'stairs down' events. The data from the barometer helps to verify the results of the magnetic field data for floor change detection. Similarly, the data from IMU sensors including accelerometer and motion sensors help to track user short-term position and can be fused with the magnetic field data to enhance the positioning accuracy.

The data are collected using static and continuous collection strategies where the former involves collecting several samples at specified location points while the latter strategy is used to collect the data when the user is walking along the specified path between the starting and ending points. Since various users are involved in the data collection, the influence of the walking speed can be analyzed using the provided dataset. The provided dataset holds several advantages over the previous publicly available dataset and a comparison of the characteristics of the dataset is given in Tab. 4.

Table 4: Comparison of publicly available magnetic field datasets

Characteristics	MagneFi	[11]	[12]	[13]	[14]	[15]
Users	Multiple	Multiple	Single	Single	Single	Single
Smartphone	Multiple(5)	Multiple(2)	Single	Single	Single	Multiple(2)
Buildings	Multiple(5)	Single	Single	Single	Single	Multiple(2)
Timespan	Long	Short	Short	Short	Short	Short
Orientations	Multiple(3)	Single	Single	Single	Single	Single
Collection strategy	Static+cont.	Cont.	Static	Cont.	Cont.	Static
Building type	Old & new	Single	Single	Single	Single	Single
Uneven floor	Yes	No	No	No	No	No
Stairs	Yes	No	No	No	No	No
Multi-floor	Yes	No	No	No	No	No
Office/lab	yes	No	No	No	No	No

7 Conclusion

With an increase in processing capabilities of the embedded sensors, smartphones can be utilized for several additional tasks than mere calling, play games, and watching videos. As a result, the smartphone industry led to the inception of several applications for the consumer market today that focus on this market like online marketing, on-the-go services, and telemedicine, etc. Being called

location-based services, they aim at providing services oriented to consumer location and require precise location information of the user to provide a high quality of service. Consequently, among other positioning technologies such as Wi-Fi, BLE, and pedestrian dead reckoning, magnetic field-based positioning has emerged as an attractive candidate for indoor positioning and localization. Despite being easy to implement, and simple to adapt, magnetic field-based positioning requires wide research efforts to analyze its various characteristics that influence the positioning accuracy. A large number of positioning approaches have been proposed, however, the lack of publicly available datasets makes it very difficult to evaluate the performance of such approaches on common benchmark datasets. Existing datasets lack several important aspects of the magnetic field data such as data from heterogeneous smartphones involving multiple orientations and different spatial considerations. This study overcomes such limitations of the magnetic field dataset by introducing the MagneFi which is a multi-user, multi-device, multi-building, and multi-floor dataset collected over almost 5 years. It involves 5 smartphones, 3 users including a female and two males of different heights, 5 different buildings with a multi-floor environment, and 3 orientations of the smartphone. The data for the stairs, as well as the laboratory, are collected to study the stairs up and down events and positioning performance of the magnetic field-based approaches in both small and large indoor areas. We believe that the collected dataset will serve the purposes of the research community to investigate and test the magnetic field-based approaches on common benchmark datasets.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Oberlo, "How many people have smartphones in 2020?," https://www.oberlo.com/statistics/how-many-people-have-smartphones#:~:text=In%202020%2C%20the%20number%20of,rate%20is%20at%2045.4%20percent, 2020.

[2] H. A. Karimi, *Advanced Location-Based Technologies and Services*, CRC Press, Tayler & Francis, FL, USA, 2013.

[3] S. Blog, "What are consumers doing on their smartphones anyway," https://www.salesforce.com/blog/2018/02/consumer-smartphone-use.html, [Online; accessed 19-May-2020].

[4] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan *et al.,* "Diversity in smartphone usage," in *Proc. Int. Conference on Mobile Systems, Applications and Services*, San Francisco, USA, pp. 179–194, 2010.

[5] P. Bahl and V. N. Padmanabhan, "Radar: An in-building rfbased user location and tracking system," in *Proc. INFOCOM, Tel Aviv*, Israel, vol. 2, pp. 775–784, 2000.

[6] M. Youssef and A. Agrawala, "The Horus WLAN location determination system," in *Proc. International Conference on Mobile Systems, Applications, and Services*, Seattle Washington, USA, pp. 205–218, 2005.

[7] L. M. Ni, Y. Liu, Y. C. Lau and A. P. Patil, "LANDMARC: Indoor location sensing using active RFID," in *Proc. IEEE PermCom*, Texas, USA, pp. 407–415, 2003.

[8] I. Ashraf, S. Hur and Y. Park, "Smartphone sensor based indoor positioning: Current status, opportunities, and future challenges," *Electronics*, vol. 9, no. 6, pp. 891, 2020.

[9]   H. Zou, X. Lu, H. Jiang and L. Xie, "A fast and precise indoor localization algorithm based on an online sequential extreme learning machine," *Sensors*, vol. 15, pp. 1804–1824, 2015.

[10]   G. Lui, T. Gallagher, B. Li, A. G. Dempster and C. Rizos, "Differences in RSSI readings made by different Wi-fi chipsets: A limitation of WLAN localization," in *Proc. Int. Conf. on Localization and GNSS (ICL-GNSS)*, Tampere, Finland, pp. 53–57, 2011.

[11]   J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J. P. Avariento, T. J. Arnau *et al.,* "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems," in *Proc. IPIN*, Busan, Kore, pp. 261–270, 2014.

[12]   P. Barsocchi, A. Crivello, D. La Rosa and E. Palumbo, "A multisource and multivariate dataset for indoor localization methods based on WLAN and geo-magnetic field fingerprinting," in *Proc. IPIN*, Alcala de Henares, Spain, pp. 1–8, 2016.

[13]   J. Torres-Sospedra, R. Montoliu, G. M. Mendoza-Silva, O. Belmonte, D. Rambla *et al.,* "Providing databases for different indoor positioning technologies: Pros and cons of magnetic field and Wi-Fi based positioning," *Mobile Information Systems*, vol. 2016, pp. 1–23, 2016.

[14]   Z. Tóth and J. Tamás, "Miskolc IIS hybrid IPS: Dataset for hybrid indoor positioning," in *Proc. Int. Conf. Radioelektronika (RADIOELEKTRONIKA)*, Kosice, Slovakia, pp. 408–412, 2016.

[15]   D. Hanley, A. B. Faustino, S. D. Zelman, D. A. Degenhardt and T. Bretl, "A dataset for indoor positioning with magnetic anomalies," in *Proc. IPIN*, Sapporo, Japan, pp. 1–8, 2017.

[16]   Asahi Kasei Microdevices Corporation. "AK09918C," Https://www.akm.com/global/en/products/electronic-compass/ak09918c/. Online: accessed 19 April 2021.

[17]   STMicroelectronics, LSM6dsl. https://www.st.com/resource/en/datasheet/lsm6dsl.pdf. Online: accessed 21 April 2021.

[18]   Asahi Kasei Microdevices Corporation, AK09916C. https://www.akm.com/akm/en/file/datasheet/AK09916C.pdf. Online: accessed 20 April 2021.

[19]   Asahi Kasei Microdevices Corporation, AK09915C. https://www.akm.com/global/en/products/electronic-compass/ak09915c/. Online: accessed 19 April 2021.

[20]   BOSCH, BMI-160. https://www.bosch-sensortec.com/media/boschsensortec/downloads/datasheets/bst-bmi160-ds000.pdf. Online: accessed 18 April 2021.

[21]   TDK Corporation, TDK InvenSense. https://invensense.tdk.com/wp-content/uploads/2017/11/DS-000196-IAM-20680-v1.1-Typ.pdf. Online: accessed 19 April 2021.

[22]   Y. Shu, C. Bo, G. Shen, C. Zhao, L. Li *et al.,* "Magicol: Indoor localization using pervasive magnetic field and opportunistic WiFi sensing," *IEEE Journal on Selected Areas in Communications*, vol. 33, pp. 1443–1457, 2015.

[23]   B. Li, T. Gallagher, A. G. Dempster and C. Rizos, "How feasible is the use of magnetic field alone for indoor positioning?," in *Proc. IPIN*, Sydney, Australia, pp. 1–9, 2012.

[24]   K. P. Subbu, B. Gozick and R. Dantu, "LocateMe: Magnetic-fields-based indoor localization using smartphones," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 4, pp. 1–27, 2013.

[25]   I. Ashraf, S. Hur, M. Shafiq and Y. Park, "Floor identification using magnetic field data with smartphone sensors," *Sensors*, vol. 19, no. 11, pp. 2538, 2019.

[26]   Y. Zou, G. Wang, K. Wu and L. M. Ni, "SmartScanner: Know more in walls with your smartphone," *IEEE Transactions on Mobile Computing*, vol. 15, pp. 2865–2877, 2015.

[27]   I. Ashraf, S. Hur and Y. Park, "Enhancing performance of magnetic field based indoor localization using magnetic patterns from multiple smartphones," *Sensors*, vol. 20, no. 9, pp. 2704, 2020.

[28]   I. Ashraf, S. Hur and Y. Park, "BLocate: A building identification scheme in GPS denied environments using smartphone sensors," *Sensors*, vol. 18, no. 11, pp. 3862, 2018.

[29]   J. Haverinen and A. Kemppainen, "Global indoor self-localization based on the ambient magnetic field," *Robotics and Autonomous Systems*, vol. 57, no. 10, pp. 1028–1035, 2009.

[30]   H. Xie, T. Gu, X. Tao, H. Ye and J. Lu, "A Reliability-augmented particle filter for magnetic fingerprinting based indoor localization on smartphone," *IEEE Transactions on Mobile Computing*, vol. 15, no. 8, pp. 1877–1892, 2015.

[31] M. Frassl, M. Angermann, M. Lichtenstern, P. Robertson, B. J. Julian *et al.,* "Magnetic maps of indoor environments for precise localization of legged and non-legged locomotion," in *Proc. Int. Conf. on Intelligent Robots and Systems*, Tokyo, Japan, pp. 913–920, 2013.

[32] J. Chung, M. Donahoe, C. Schmandt, I. J. Kim, P. Razavai *et al.,* "Indoor location sensing using geo-magnetism," in *Proc. Int. Conf. on Mobile Systems, Applications, and Services*, Bethesda Maryland USA, pp. 141–154, 2011.

[33] I. Ashraf, S. Hur and Y. Park, "mPILOT:-magnetic field strength based pedestrian indoor localization," *Sensors*, vol. 18, no. 7, pp. 2283, 2018.