

Efficient Indoor Localization Based on Geomagnetism

HANG WU, ZILIANG MO, JIAJIE TAN, SUINING HE, and S.-H. GARY CHAN,
The Hong Kong University of Science and Technology, People's Republic of China

Geomagnetism is promising for indoor localization due to its omnipresence, high stability, and availability of magnetometers in smartphones. Previous works often fuse it with pedometer via particles, which are computationally intensive and make strong user behavior assumptions. To overcome that, we propose Magil, an approach leveraging geomagnetism for indoor localization. To our best knowledge, this is the first piece of work using geomagnetism for smartphone localization without the need of pedometer or user walking model. Magil is applicable to any open or complex indoor environment. In the offline phase, Magil collects and stores geomagnetic fingerprints while surveyors walk indoors. In the online phase, it employs a fast algorithm to match the geomagnetic variation of the target with the stored fingerprints. Given closely matched segments, Magil constructs user trajectory efficiently with a modified shortest path formulation by selecting and connecting these matched segments.

To further improve accuracy and deployability, we propose MagFi, which extends **Magil** by fusing it with **Wi-Fi**. MagFi further collects opportunistic Wi-Fi RSSI for fingerprint construction. We have implemented both Magil and MagFi and conducted extensive experiments in our campus. Results show that our schemes outperform state-of-the-art schemes by a wide margin (often cutting localization error by 30%).

CCS Concepts: • **Networks** → **Location based services**; • **Human-centered computing** → **Ubiquitous and mobile computing**;

Additional Key Words and Phrases: Indoor localization, geo-magnetism, smartphone-based, fusion, joint constraints, shortest path problem formulation, sequence matching

ACM Reference format:

Hang Wu, Ziliang Mo, Jiajie Tan, Suining He, and S.-H. Gary Chan. 2019. Efficient Indoor Localization Based on Geomagnetism. *ACM Trans. Sen. Netw.* 15, 4, Article 42 (August 2019), 25 pages.
<https://doi.org/10.1145/3342517>

1 INTRODUCTION

There has been much recent interest in indoor localization, mainly due to the ubiquity of smart devices and advances in their sensing capabilities [7]. Fingerprinting emerges as a promising approach, because it is applicable to complex indoor environments without the need for any model

This work is supported, in part, by Guangdong Provincial Department of Science and Technology (GDST16EG04), Guangdong-Hong Kong joint innovation project (2016A050503024), Natural Science Foundation of Guangdong Province (2014A030313154), and the Sustainable and Smart Campus Project of HKUST (F0713).

Authors' address: H. Wu (corresponding author), Z. Mo, J. Tan, S. He, and S.-H. G. Chan, Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, People's Republic of China; emails: hwuav@cse.ust.hk, zmo@connect.ust.hk, {jtana, shea, gchan}@cse.ust.hk.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2019 Association for Computing Machinery.

1550-4859/2019/08-ART42 \$15.00

<https://doi.org/10.1145/3342517>

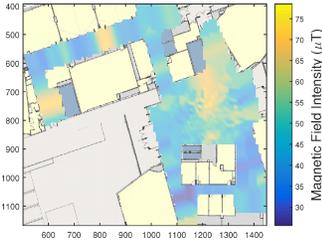


Fig. 1. The magnetic fingerprint (intensity in μT) in a typical indoor environment.

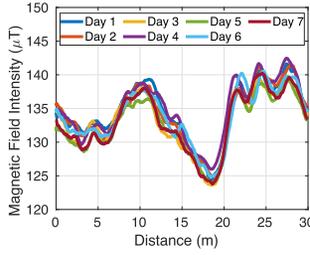


Fig. 2. Magnetic field intensity along the same trajectory in different days.

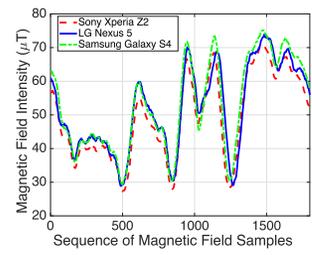


Fig. 3. Magnetic field intensity of different devices along the same trajectory.

assumption and line-of-sight signal measurements [6, 14, 35]. In the offline (survey) phase of fingerprint-based localization, signal strengths at predefined locations called reference points (RPs) are collected by surveyors. The signal strengths together with their corresponding locations form the fingerprints.¹ In the subsequent online phase, user location is obtained by matching his/her collected signals with the stored fingerprints.

Among all the sensed signals explored, geomagnetic field holds much promise, because it exhibits strong local variations caused by electrical appliances and building materials [11]. This means that an alteration in indoor layouts often has only local impacts without affecting the signal map significantly over a wide scope. It has also been shown that the impact of mobile objects and smartphone placements on the magnetic field is minimal [3, 11, 21, 34]. Figure 1 illustrates such local variations in a typical indoor environment, which can be used as fingerprints to identify target location.²

Moreover, the geomagnetic field is omnipresent, so there is no need for any extra infrastructure support. It can be efficiently collected at a high rate *while* walking using commonly available mobile devices. The field is more stable with much lower random fluctuation as compared with other signals, leading to better performance [11]. To verify its reliability, we conduct an experiment that collects the field intensity along a corridor in a week. Figure 2 shows the magnetic field at different location each day. We can observe that magnetic fields have similar distributions within the week. Such temporal invariance and spatial-distinguishable characteristics provide opportunities to employ magnetic fields as fingerprints for pervasive indoor localization.

However, even though different devices in the same location may have different magnetic field readings (i.e., device heterogeneity), constant offsets are observed among these devices. Figure 3 illustrates magnetic field data collected using different devices along the same trajectory. Clearly, although the reading varies for different devices, the gaps between their readings remain rather constant. Such a property leads to simplicity in design and applicability of algorithms across different device platforms.

In this article, we devise novel indoor localization techniques based on geomagnetic signals. Previous work in the area has been predominantly on fusing with pedometer (step counter) using recursive Bayesian filters such as Hidden Markov Models (HMMs) [15] and particle filters [20, 34]. They estimate the location based on maximum joint probability or particle convergence, given the magnetic readings and walking distance of the target. Some models also assume the availability

¹Unless otherwise stated, in this article we refer to RSSI as Wi-Fi RSSI.

²The x -axis and y -axis in all the maps in this article are in pixels, and 40 pixels in the maps correspond to 3m in the physical world.

of initial locations and accurate user walking model. Despite promising results, these approaches are often computationally intensive for a mobile device. Furthermore, they work the best for partitioned environment characterised by narrow corridors where the degree of freedom of the target is low. For large spacious settings, they hardly converge to the target location. Some other works make strong assumptions on user behaviors or rely heavily on a meticulously tuned pedometers (to accurately estimate steps, walking distances and directions). In practice, the pedometer is subject to random user movement, and errors tend to accumulate over time [2].

To overcome these weaknesses, we propose **Magil**, a novel approach leveraging **m**agnetic fields for indoor **l**ocalization. Without using any step counter or error-prone sensors such as a gyroscope (which suffers from angular drift), Magil is hence robust against arbitrary user behaviors. It is applicable to any indoor environment, including large open settings. To the best of our knowledge, this is the first work on using geomagnetism for smartphone-based indoor localization without the domain knowledge of a pedometer, initial position, or user walking model.

In its offline phase, Magil designs survey paths to collect magnetic field signals as *fingerprints*. Through discretizing the indoor space, target movements can be represented as a sequence of matched segments in fingerprints (and thus positions in the real world). For online location estimation, Magil first collects a sequence of magnetic signals along the walking path of the user (or target). Then it finds the trace of fingerprinted signals that best match with that sequence. It then selects and orders the matched segments properly to yield the target locations over time.

The crux of Magil is trajectory matching, which consists of two subproblems: how to identify the fingerprint segments best match the target observations (the so-called *sequence matching* problem), and, given the closely matched segments, how to select and connect those matched segments together to form a user path (the so-called *path construction* problem). We formulate them as optimization problems. For sequence matching, we propose a novel modified Smith-Waterman algorithm [22] that matches sequences obtained under different walking speeds. For path construction, we model and solve it as a modified shortest path problem. Our algorithms are computationally efficient.

While Magil typically works well, its accuracy can be further enhanced by fusing with Wi-Fi Received Signal Strength Indicator (RSSI) [1, 35, 36]. This is motivated by the facts that Magil alone may not work well when users stand still, because the magnetism readings may not show much variation. Geomagnetic signal also lacks unique location identifiers over a broad scope, because remote locations may possess similar geomagnetic field patterns. RSSI, however, shows distinct spatial variation due to complex indoor structure, with nearby locations sharing similar radio environments [5, 20].

Inspired by the above observations, we propose MagFi, a scheme fusing **Magil** with Wi-Fi RSSI fingerprints (MagFi). In its offline phase, MagFi collects both geomagnetic and Wi-Fi RSSI samples and constructs an RSSI fingerprint database automatically using geomagnetic signals. In the online stage, MagFi filters away those signal traces whose geomagnetic fields are not consistent with user observed RSSIs, hence improving the computational efficiency and accuracy. Compared with Magil, MagFi is much more robust in open areas, where geomagnetic field patterns can be quite similar. To summarize, MagFi broadens the scope of application to more general indoor environments and user behaviors and is more accurate and robust as validated by our experiments.

We have implemented both Magil and MagFi in mobile phones and conducted extensive experiments. The experimental results confirm the implementability and efficiency of our algorithms. Both systems outperform existing and state-of-the-art algorithms by a large margin (often cutting the location error by more than 30%).

We note in passing that the site survey of Magil can be conducted in parallel, i.e., multiple surveyors can conduct the site survey simultaneously and their fingerprints can be combined to

reduce the total survey time. Compared with the traditional Wi-Fi RSSI site survey procedure where signals vectors have to be collected by standing at length at different fixed points in the area, Magil can save time, because it collects fingerprint signals while the surveyor is walking. Furthermore, crowdsourced geomagnetism database can be integrated in our proposed system seamlessly and the survey cost can be significantly reduced. Fingerprint update can also be achieved using crowdsourcing methods [38].

The remainder of the article is organized as follows. After discussing related work in Section 2, we overview the system framework for Magil and MagFi in Section 3. In Section 4, we show how to construct the geomagnetic database for both Magil and MagFi and the RF RSSI fingerprint database for MagFi. Magil is then presented in Section 5, followed by MagFi in Section 6. In Section 7, we present illustrative experimental results and conclude with future works in Section 8.

2 RELATED WORK

We review related works on mobile localization in this section. There have been extensive studies on using ambient signals for localization, e.g., Wi-Fi [1, 31], geomagnetic field [11, 20], vision [9, 13], and so on. Among them, fingerprint-based schemes [1, 18, 20, 24, 27, 36] are promising, because they do not require the knowledge of signal source or propagation. As consecutive magnetic observations indicate much precise spatial movement information while Wi-Fi provides unique location identifiers over a broader scope, our work applies fingerprint techniques to fusing Wi-Fi on top of magnetic field to provide a much more robust and adaptive localization system.

There has been some work on geomagnetism-based localization [3, 4, 7, 16]. Unloc [25] proposes using sparse magnetic disturbances as landmarks for indoor localization. Different from their work, we use geomagnetic field to build the signal map. FOLLOWME [21] leverages walking patterns of earlier travelers to navigate the following users. However, it requires that there should be a reference trace with the same origin and destination for both earlier travelers and following users. Magil does not have this requirement and thus is more flexible. Inspired by the observation that geomagnetism changes around pillars or gates, LocateMe [23] maps the target location to the landmarks with the similar trend of signal change. The above works are best applied in narrow corridors where pedestrian walking patterns are constrained. Magil is not based on these approaches and is more applicable in complex environments.

IMU-assisted schemes (a.k.a. dead-reckoning) has been well studied [12, 21, 26]. These systems use readings from different sensors (accelerometer, gyroscope, etc.) and floor plans to constrain localization error. In contrast, Magil does not rely on gyroscope or accelerometer, which are error-prone and require meticulous calibration. Therefore, we achieve consistently robust localization results without assumptions on user behavior.

Fusion of fingerprint signals and sensors has been recently studied [8, 29, 32]. MapCraft [32], Mapel [29], and the work in Reference [19] utilize the conditional random field (CRF) to localize users. In these works, sequential motion and sensor readings are fed to a graphical model for location estimation. GROPING [38] fuses geomagnetic fingerprints with gyroscope readings to crowdsource fingerprint database and localize users. Magil advances from them by modeling the problem as a signal matching problem, which achieves higher localization accuracy without relying on user walking direction estimations and error-prone step counters [2, 39] and gyroscopes.

Particle-filter-based fusion has been widely studied in recent years for geomagnetic field localization. Magicol [20] considers a two-way particle filter to improve the fusion of Wi-Fi fingerprint and magnetic fields. MaLoc [33, 34] implements a novel augmented particle filter to address motion estimation errors. MaWi [37] fuses Wi-Fi with geomagnetism based on particle filter to estimate user velocity and locations. The work in Reference [8] approximates the floor map by connected line segments. All these works require either a meticulously tuned pedometer or a walking model

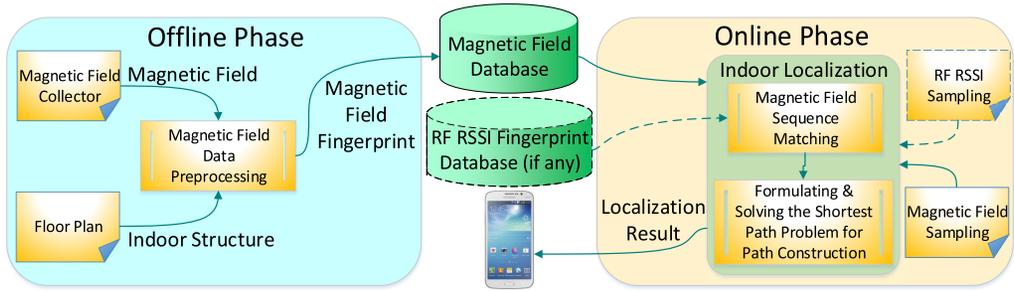


Fig. 4. System framework for Magil and MagFi.

to estimate user movements including walking distances, velocities, and directions. Unlike these works and others [2, 37, 39], Magil is not based on pedometers and walking models. Instead, it transforms the localization problem into a signal matching and path construction problem. Furthermore, we modify the typical Smith-Waterman algorithm by introducing a new cost function to align two sequences. The advantage of our approach is that the Smith-Waterman algorithm allows gaps while aligning, which is not allowed in dynamic time warping. After that, our substrings matching and selection algorithm is applied to generate all favored local alignments.

A preliminary version of this work has been reported in Reference [30]. We further advance from it in the following major ways: (1) *Significant computation efficiency improvement*: The previous trajectory matching algorithm still suffers from high computational overhead in the online phase. We improve it to achieve approximately $2.3\times$ speedup. (2) *Advanced fusion and accuracy enhancement*: We propose MagFi, which fuses Magil with RF fingerprints (Wi-Fi RSSI) to improve localization accuracy. (3) *Extensive experiment studies*: We conduct more extensive experimental evaluations to validate the performance of Magil and MagFi by comparing them with other state-of-the-art algorithms.

3 SYSTEM FRAMEWORK FOR MAGIL AND MAGFI

We show in Figure 4 the system framework for Magil and MagFi. It consists of two phases, the *offline* phase and the *online* phase. Note that Magil only has solid lined modules while MagFi has both solid lined modules and dotted modules. In the offline phase, surveyor(s) carries a smartphone while walking through the area. While walking, *geo-magnetic field* signals and their corresponding timestamps are fed to the module *magnetic field data preprocessing*. Given the map information, the module generates the fingerprint signal map and stores it in the *database* for online use. Furthermore, in MagFi, if Wi-Fi sensing results are available during the offline phase, then they will also be processed to construct a Wi-Fi RSSI fingerprint database.

In the online phase, a user continuously measures the geo-magnetic field, which is fed to the *magnetic field indoor localization* module. The module consists of two steps. First, in the module *magnetic field sequence matching*, we use our proposed modified Smith-Waterman algorithm [22] to find those segments whose fingerprint variations best match the target observations. If RSSI fingerprints are available, then MagFi also leverages them to further filter the matched segments to eliminate the segments whose RSSI fingerprints are not similar to the corresponding user observations. Second, in the module *formulating and solving the shortest path problem*, we formulate a novel shortest path problem. By solving this problem, we reorder the matched segments properly and find the trajectories and locations of the user. The results are then returned to the user.

4 FINGERPRINT DATABASE CONSTRUCTION FOR GEOMAGNETISM AND WI-FI

In this section, we first discuss how to collect magnetic fields using smartphones and identify magnetic field features for fingerprints in Section 4.1. Given that, in Section 4.2 we describe how to construct the magnetic signal fingerprint database for both Magil and MagFiand a Wi-Fi RSSI fingerprint database for MagFi.

4.1 Extracting Geomagnetism Features

The magnetic field vector \mathbf{B}_p can be measured by a smartphone's magnetometer [11]. However, the raw magnetic readings are under the smartphone's coordinate system. We hence need to transform the readings into the one under the Earth's coordinate system by the yaw ψ , pitch θ , and roll ϕ of the smartphone, i.e.,

$$\mathbf{B}_p = \mathbf{R}_x(\theta)\mathbf{R}_y(\phi)\mathbf{R}_z(\psi)\mathbf{B}_e, \quad (1)$$

where \mathbf{B}_e is the magnetic field vector at the same location in terms of the Earth's coordinate system, and $\mathbf{R}_x(\theta)$, $\mathbf{R}_y(\phi)$, $\mathbf{R}_z(\psi)$ are corresponding *rotation matrices* w.r.t. the three axes of the smartphone [11]. Then we obtain

$$\mathbf{B}_e = \mathbf{R}_z^{-1}(\psi)\mathbf{R}_y^{-1}(\phi)\mathbf{R}_x^{-1}(\theta)\mathbf{B}_p. \quad (2)$$

In this way, we can measure the magnetic field \mathbf{B}_e irrespective of the dynamic smartphone headings. However, we do not use \mathbf{B}_e directly as the observation, because a smartphone's heading estimation is error-prone [39] and these errors would be amplified on \mathbf{B}_e . Some works [11, 20] suggest using the magnitude of \mathbf{B}_p as fingerprints, because it is a rotation-invariant scalar quantity and quite stable. However, there is only one fingerprint dimension, which reduces the uniqueness of the fingerprints. However, most smartphones have been equipped with gravity sensors that sense the direction of gravity and are stable with location and time [11]. Therefore, we can retrieve both the vertical and horizontal components (w.r.t. gravity), B_v and B_h , of \mathbf{B}_p and combine them with the magnitude of \mathbf{B}_p to generate an observation at location o , i.e.,

$$\mathbf{B}_o = (\|\mathbf{B}_p\|, B_v, B_h). \quad (3)$$

4.2 Database Construction

In the offline phase, we conduct magnetic field data collection and database construction as follows:

1. *Survey path planning*: We preplan several survey paths from the floor map we obtain in the map preprocessing step. Typically, these paths should be along the corridors, across lobbies, and at the peripheries of obstacles and cover various walking directions. For an extremely large open space, we can add more paths that altogether form a grid with diagonal lines to ensure the coverage of positions and walking directions of interest. An example is shown in Figure 5, where the corridor is covered by one survey path and a large indoor area is covered by many survey paths in any potential directions. Note that we can adaptively align the survey paths according to the site shape. For example, in constrained corridors we align the trajectories along the accessible paths, while in open spaces we may align them in intersecting parallels. Note that for each survey path, we only need to walk in one direction to collect the fingerprint data.
2. *Magnetometer and motion sensor measurement*: A surveyor(s) walks along paths while his/her smartphone records magnetic field data and corresponding timestamps during the walk. Note that the surveyor(s) should walk at a reasonably constant speed for the better data-trajectory matching purpose, which is introduced in the following.

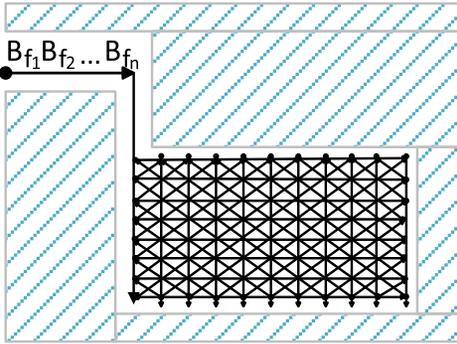


Fig. 5. An example of survey paths and corresponding walking directions. One fingerprint magnetic field data vector $Seq_{fingerprint} = \langle B_{f_1} B_{f_2} \dots B_{f_n} \rangle$ is also shown in the figure.

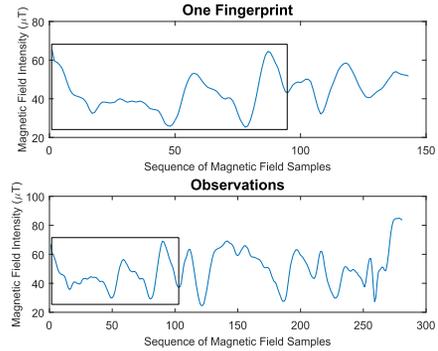


Fig. 6. Magnetic field sample matching between fingerprint and observations. Note that only the parts in the boxes match each other well.

3. *Data-trajectory matching and data preprocessing*: An actual trace is mapped against a pre-planned path based on data timestamps and turns. We calculate the locations of intermediate magnetic field data proportionally according to their timestamps and the overall time interval between two turns. The physical location of each magnetic field data collected is interpolated into segments of preplanned paths proportional to their timestamp differences. If Wi-Fi RSSI readings are available, then we also bind each RSSI vector to the magnetic field data with the closest timestamp and interpolate its location accordingly. This is because Wi-Fi data sampling frequency in most COTS smartphones (roughly 0.25–2Hz) is much lower than the magnetic field sampling frequency. After this, Piecewise Aggregate Approximation (PAA [10]) is performed separately on the magnetic field data of each survey path to reduce the data dimensionality. More specifically, the length of the data is denoted as n . Then the data are partitioned into $\frac{n}{10}$ equally sized frames, within each of which we calculate the mean value of the magnetic field data. A vector of these mean values from all frames becomes the dimension-reduced representation of the magnetic field data and is stored in the database. Note that as a by-product, a Wi-Fi RSSI fingerprint database can be constructed in the mean time.

It is worth noting that, in practice, we find that geomagnetic fields may vary dramatically within a small region. Hence we require all the survey paths to cover the whole area of interest to capture the geomagnetism variation accurately, which increases the database and localization accuracy in turn. Clearly, the above procedures can be conducted in parallel (i.e., multiple surveyors can conduct the site survey in the meantime and combine the fingerprints together to reduce the total survey time). Compared with the traditional Wi-Fi RSSI site survey procedure where signals vectors have to be collected by standing at length at different fixed points in the area, our system can save time, because it collects fingerprint signals while surveyor(s) is *walking*. Moreover, the more dense the survey paths are, the more time it may take to complete the site survey task. To balance the performance of our system, we require the physical distance between two parallels to be 1m.

Note that Magil and MagFi also work seamlessly in cooperation with existing crowdsourced fingerprint databases, which can reduce the site survey effort. However, since our article focuses on locating users given geomagnetism (and opportunistic Wi-Fi signals), interested readers are referred to References [28, 38] for crowdsourcing fingerprints.

Table 1. Major Symbols Used in Magil

Notations	Definitions
\mathbf{B}_p	Geomagnetism reading under phone coordinate system
\mathbf{B}_o	Geomagnetism observation under Earth coordinate system
$\langle \mathbf{B}_{o_1} \mathbf{B}_{o_2} \dots \mathbf{B}_{o_m} \rangle$	User observation sequence $Seq_{observe}$ whose length is m
$\langle \mathbf{B}_{f_1} \mathbf{B}_{f_2} \dots \mathbf{B}_{f_n} \rangle$	A fingerprint magnetic field data vector $Seq_{fingerprint}$ whose length is n
G	The corresponding graph generated by transforming all matched parts of fingerprints and user observations
N	The number of vertexes in G
M	The number of edges in G

5 MAGIL: INDOOR LOCALIZATION BASED ON GEOMAGNETISM

In this section, we present the problem formulations and solutions for sequence matching and path construction in Magil. The important symbols used are summarized in Table 1.

The user's smartphone first samples a sequence of magnetic field data, and Magil performs the similar PAA [10] on the geomagnetism data to obtain a shorter sequence of magnetic field observations. After that, we need to address the problem of how to determine the user location based on the signal sequence given the magnetic fingerprint database. In geomagnetism-based localization, the following two subproblems arise before the target can be effectively located.

From Figure 3, we note that along the same trajectory, the shapes of magnetic field sequences among devices are similar. As each sample is collected by a user somewhere in the survey path, the first subproblem, the so-called *sequence matching* problem, is as follows: Given a vector with each component being an observation sampled by the user and a magnetic field fingerprint data vector obtained from the database, how do we determine whether and, if so, where they are similar?

Furthermore, a user may traverse any place, and thus different observations may match different magnetic field fingerprint data vectors. Therefore, the second subproblem, the so-called *path construction* problem, is as follows: How can we perform "concatenation" on the matchings to obtain the whole estimated walking trajectory?

To address the above two subproblems, we propose a novel geomagnetism-based localization algorithm. To match the geomagnetism sequence, we modify the Smith-Waterman algorithm [22] to determine the similarity between the signal observations from the user and the observations from survey paths (Section 5.1). Given the matching results, we convert them into vertexes and a graph, formulate a shortest path problem (Section 5.2), and solve it efficiently via a modified Dijkstra algorithm (Section 5.3). For practical deployment of Magil, we introduce several improvements in Section 5.4. We end by summarizing the computational efficiency (Section 5.5).

5.1 Sequence Matching for Geomagnetic Field

We first match the user magnetic observations against the fingerprint. Figure 6 shows a matching example between a observation sequence and the fingerprint. We can clearly see that the parts in the boxes match each other well. The problem here is how to identify the similar parts among sequences.

We devise a magnetic field sequence matching scheme by modifying the Smith-Waterman algorithm [22]. The Smith-Waterman algorithm is a dynamic programming algorithm for performing local sequence alignment that has been widely used in bioscience (e.g., to determine similar regions between two strings of nucleotides or protein sequences).

We make modifications to the original Smith-Waterman algorithm to support the magnetic field sequence matching. We denote an observed sample (after performing PAA) as \mathbf{B}_o and the user observation sequence as $Seq_{observe} = \langle \mathbf{B}_{o_1} \mathbf{B}_{o_2} \dots \mathbf{B}_{o_m} \rangle$, where m is the sequence length. Note that each magnetic field observation \mathbf{B}_o can be viewed as a point in 3D space. We also denote a magnetic field fingerprint data vector as $Seq_{fingerprint} = \langle \mathbf{B}_{f_1} \mathbf{B}_{f_2} \dots \mathbf{B}_{f_n} \rangle$, where \mathbf{B}_{f_i} is the i th fingerprint magnetic field data (which is also an observation). However, we cannot determine whether \mathbf{B}_{o_i} matches \mathbf{B}_{f_j} by simply calculating their Euclidean distance, since different devices may yield different magnetic field values even if samples are collected at the same place. Figure 3 also shows this characteristic. To address this, our key observation is that the differences between readings among different devices are almost the same despite locations. Let $dist(\cdot, \cdot)$ be the Euclidean distance between two samples. Based on this observation, we can make use of it and determine \mathbf{B}_{o_i} and \mathbf{B}_{f_j} match each other if and only if $g(o_i, f_j)$, the average Euclidean distance between adjacent samples of \mathbf{B}_{o_i} and \mathbf{B}_{f_j} after the mean removal, is less than a certain threshold, i.e.,

$$g(o_i, f_j) < Threshold, \quad (4)$$

where

$$g(o_i, f_j) = C \cdot \sum_{t=-window}^{window} dist(\mathbf{B}_{o_{i+t}} - Mean(o_i), \mathbf{B}_{f_{j+t}} - Mean(f_j)), \quad (5)$$

$$Mean(o_i) = C \cdot \sum_{t=-window}^{window} \mathbf{B}_{o_{i+t}},$$

$$Mean(f_j) = C \cdot \sum_{t=-window}^{window} \mathbf{B}_{f_{j+t}}, \quad (6)$$

$$C = 1/(2 \cdot window + 1).$$

In practice, we empirically set $window = 5$ and $Threshold = 2\mu T$, which balances the robustness and reasonable computation time. We assign a score 1 for matches and the penalty cost for mismatches is set to an empirical value of -1 , which yields competitive performances in practice. We perform this modified Smith-Waterman sequence matching algorithm on each original magnetic field fingerprint data vector $\langle \mathbf{B}_{f_1} \mathbf{B}_{f_2} \dots \mathbf{B}_{f_n} \rangle$ and its corresponding reversed data vector $\langle \mathbf{B}_{f_n} \mathbf{B}_{f_{n-1}} \dots \mathbf{B}_{f_1} \rangle$, since the user can walk in two different directions along the same path. Note that user observation sequence and fingerprint data vector may be collected under different walking speeds even walking through the same path (hence their length may be different), and our algorithm can perform alignment for sequences with different lengths.

After running the sequence matching algorithm for a magnetic field fingerprint data vector and a user observation sequence, we can obtain several possible matched substrings with high matching scores. Note that a substring of the user observation sequence can match many substrings of the magnetic field fingerprint data vector due to the similarities in magnetic field data, and a substring of fingerprint magnetic field data vector can match many substrings of the user observation sequence, because the user can walk around the same place many times. Another characteristic is that if substring \vec{a} matches \vec{b} with a high score, then another substring \vec{c} whose position is near \vec{a} can also match \vec{b} with a relatively high score. We call this phenomenon *pattern repetition*. To select those substrings with high matching scores and reduce pattern repetition (thus reduce the time and space complexity), we introduce a new substring selection algorithm (Algorithm 1).

We use a parameter R in Algorithm 1 to determine the number of pattern repetitions. A large R will reduce pattern repetition, but it can also remove some correctly matched substrings. A small R can lead to both more pattern repetitions and more substrings. In our system we choose $R = 10$,

ALGORITHM 1: Matching Substrings Selection Algorithm

Input: Fingerprint magnetic field data vector $Seq_{fingerprint}$, user observation sequence $Seq_{observe}$, matching score matrix $M_{m \times n}$ obtained from the previous magnetic field sequence matching algorithm

Output: Matched substrings

```

1:  $D \leftarrow \mathbf{0}_{m \times n}$ 
2: while True do
3:   Select an entry  $(a_1, b_1)$  with the highest score  $M(a_1, b_1)$  satisfying  $D(a_1, b_1) = 0$ ; if impossible, break the while loop
4:   if  $M(a_1, b_1) < Threshold_1$  then
5:     Break
6:   end if
7:   Backtracking from  $(a_1, b_1)$  to reconstruct the optimum alignment  $\langle (B_{f_{a_1}}, B_{o_{b_1}}), \dots, (B_{f_{a_t}}, B_{o_{b_t}}) \rangle$  between two sequences by using  $M_{m \times n}$ 
8:   if  $t < Threshold_2$  then
9:     Break
10:  end if
11:  for all  $1 \leq p \leq m, 1 \leq q \leq n$  do
12:    if  $\sqrt{((a_1 - p)^2 + (b_1 - q)^2)} < R$  then
13:       $D(p, q) \leftarrow 1$ 
14:    end if
15:  end for
16:  Calculate the maximum length of platform of  $\langle a_1 a_2 \dots a_t \rangle$  as  $Plat_{\langle a_1 a_2 \dots a_t \rangle} = \max\{|p - q| \mid a_p = a_{p+1} = \dots = a_q\}$ . Similar calculation is also performed for  $\langle b_1 b_2 \dots b_t \rangle$ .
17:  if  $Plat_{\langle a_1 a_2 \dots a_t \rangle} \leq Threshold_3 \wedge Plat_{\langle b_1 b_2 \dots b_t \rangle} \leq Threshold_3$  then
18:    Yield these two substrings as matched substrings
19:  else
20:    Discard these two substrings
21:  end if
22: end while

```

an empirical value that balances the time complexity and the overall performance. Note that we also use several thresholds in the algorithm. The first one, $Threshold_1$, is to avoid the substrings whose scores are too low (thus not similar). The second one, $Threshold_2$, is to avoid the substrings too short to match to increase robustness. The last one, $Threshold_3$, is to avoid the substrings in which too many deletions and insertions occur during matching. This is because when a user walks along or near a survey path in the same walking direction and the same path as the surveyor, the observations are also order-preserving; we cannot skip too many magnetic field samples during matching, since they walk through the same place. We introduce some empirical values of these thresholds and analyze the system's performance under different thresholds in Section 7.

5.2 Shortest Path Formulation for Path Construction

Since a user may traverse many places, matches between the user observation sequence and one magnetic field fingerprint data vector may cover only a fraction of the whole observation sequence. We need an algorithm to combine matches among different fingerprints of magnetic field data vectors to obtain a path that covers the whole observation sequence.

More specifically, we denote that for each magnetic field fingerprint data vector $Seq_{fingerprint_i}$ we find a set of substrings $S_i = \{s_{i,1}, s_{i,2}, \dots, s_{i,k}\}$ using Algorithm 1. Each $s_{i,j} = \langle \mathbf{B}_{f_{i,a_j,1}} \mathbf{B}_{f_{i,a_j,2}} \dots \mathbf{B}_{f_{i,a_j,t}} \rangle$ ($1 \leq j \leq k$) matches some part of user observations, say, $\langle \mathbf{B}_{o_{b_1}} \mathbf{B}_{o_{b_2}} \dots \mathbf{B}_{o_{b_t}} \rangle$. We note that each substring $s_{i,j}$ can also be viewed as a corresponding physical route, since each sample can be mapped to a unique physical location. In the following, we use the words substring and physical route interchangeably for convenience. Our target is to find an ordered sequence of substrings Seq that together can cover all user observations in order and *physically connected*. Here *physically connected* means that every two consecutive substrings in Seq are connected head-to-tail in physical locations. This constraint is necessary, because user movement is continuous.

We approach this subproblem by transforming it into a shortest path problem. We consider all substrings as vertexes in a graph G . For any two vertexes, or two substrings s_i and s_j , there is a directed edge between them if there exists some physical location LOC such that both substrings go through the nearby LOC . Say that the nearest sample in s_i to LOC is $\mathbf{B}_{f_{i,a_x}}$, and the nearest sample in s_j to LOC is $\mathbf{B}_{f_{j,a_y}}$. We assign two weights, b_x and b_y , to the edge, where $\mathbf{B}_{o_{b_x}}$ aligns to $\mathbf{B}_{f_{i,a_x}}$ in s_i and $\mathbf{B}_{o_{b_y}}$ aligns to $\mathbf{B}_{f_{j,a_y}}$ in s_j . The cost of the edge is the absolute value of the difference between b_x and b_y . Note that b_x can be smaller or bigger than b_y due to the noise in magnetic field samples and alignment, and there may be many edges between two vertexes, as there can be infinitely many potential LOC s if the user walks through the same place multiple times. To limit the number of edges between two vertexes, we only select those LOC s that are not so physically close, say at least 0.4m in Euclidean distance between any two LOC s. Note that LOC s are determined online, and the graph is built based on the substrings.

Finally, we need to align the beginning and the end of user observations. We add two new vertexes to G : $START$ connects to each of the substrings matching the beginning of user observations with two weights: 0 and the index of the first matched user observation (also the matched user observation with the smallest index), while END is connected by each of the substrings matching the end of user observations with two weights: the index of the last matched user observation (also the matched user observation with the largest index) and m , the length of the user observation sequence. Note that due to the noise in the measurements, the beginning and the end of user observations may not be matched. Therefore, we propose that $START$ connects to the substrings with the smallest index of the matched user observation no greater than R_1 , and END connects to the substrings with the largest index of the matched user observation no smaller than $m - R_1$. A large R_1 can tolerate more noise but introduce more edges that increases the time complexity, while a small R_1 requires a stricter matching of the sequences but reduces the number of edges. Here we set $R_1 = 6$, which balances the performance of our system. We require that a *valid path* is a simple path where two consecutive edges, e_1 with weights b_{1x} and b_{1y} , e_2 with weights b_{2x} and b_{2y} , satisfy that:

$$b_{1y} \leq b_{2x}. \quad (7)$$

This is because we can only align samples in time order.

We illustrate an example in Figure 7, where each colored arrow represents a survey path with walking direction (and thus one corresponding fingerprint vector). Suppose the user walks along the trace $A-LOC1-LOC2-B$. Then after applying the magnetic field sequence matching and selection (Algorithm 1), we can obtain several matched substrings. In Figure 7, each colored number represents the index of user observations matching the fingerprint with the same color. For example, the first 15 samples of user observations $\langle \mathbf{B}_{o_1} \mathbf{B}_{o_2} \dots \mathbf{B}_{o_{15}} \rangle$ match a part of the blue trace from location A to $LOC1$ (so there is a blue number 1 and a blue number 15 near the blue magnetic field fingerprint data vector $Seq_{fingerprint_1}$, corresponding to the beginning and the end of the matched substring, respectively). Additionally, $\langle \mathbf{B}_{o_{13}} \mathbf{B}_{o_{14}} \dots \mathbf{B}_{o_{19}} \rangle$ matches a part of the fingerprint $Seq_{fingerprint_4}$ of the

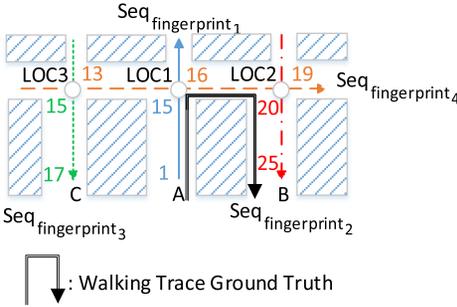


Fig. 7. An example of location inference.

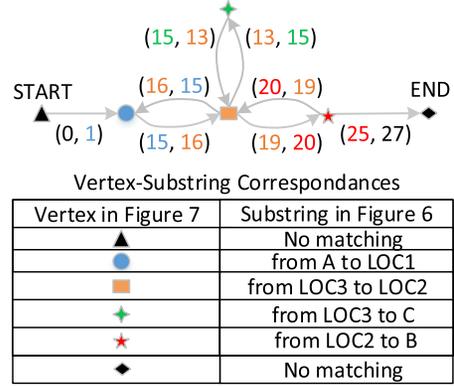


Fig. 8. The corresponding graph G after transforming the example in Figure 7 into a shortest path problem.

orange trace from location $LOC3$ to $LOC2$, and so on. Note that the index of matched user observations around $LOC1$, $LOC2$, and $LOC3$ can be different due to the noise in measurement.

After transforming this example into a shortest path problem, we obtain a graph G shown in Figure 8 containing four vertexes, each of which represents a matched substring (a part of colored arrow), besides $START$ and END . Each vertex corresponds to the same-corresponding-color matched substring in Figure 7. We can see that the corresponding simple path from $START$ to END (which is also the only available simple path) (the term “path” means “simple path” in this article) has edge weights $(0, 1)$, $(15, 16)$, $(19, 20)$, and $(25, 27)$, where 27 is the length of the user observation sequence. We can clearly see that the weights satisfy the condition mentioned before, thus the path is a valid path that is preferable.

Note that the corresponding path of trace $A-LOC1-LOC3-C$ is not a valid path, since $(15, 16)$ and $(13, 15)$ are weights belonging to two adjacent edges in the corresponding path, and they do not satisfy the constraint (7). Thus this path is not to be considered matching. If we walk along the trace $A-LOC1-LOC3-C$ in reality, then we will align the reversed fingerprint sequence corresponding to the orange arrow instead. Recall that the constraint (7) ensures that samples are aligned in the correct time order. Also, we note that there is no edge connecting the green vertex (corresponding to a matched substring in the magnetic field fingerprint data vector $Seq_{fingerprint_3}$) to vertex END . This is because in the corresponding substring, the largest index of the matched user observation is only 17, which is smaller than $27 - R_1 = 21$ (thus not close to the end of the user observation vector). Therefore, the user does not walk in the direction that the green arrow shows.

We denote the cost of a path as the sum of the weights of all edges in the path. For instance, in the above example, the cost of the valid path from $START$ to END is $|0 - 1| + |15 - 16| + |19 - 20| + |25 - 27| = 5$. Then all *valid paths* from $START$ to END can be viewed as possible user trajectories, and the *minimum cost valid path* corresponds to the estimated user trajectory where substrings can match as many user observations in the correct order as possible. In this context, the minimum cost path can be viewed as the shortest path from $START$ to END interchangeably if we view the cost of an edge as the distance.

5.3 Effective Solution for the Modified Shortest Path Problem

We denote N and M as the number of vertexes and edges in G , respectively. To solve the above shortest path problem, we need to convert the constraint (7) into a more usable one. If we add a

new edge e_0 to G that only connects to $START$ with both weights 0, and add another new edge e_{end} to G that only connects with END with both weights INT_MAX (a sufficiently large integer), and view edges in G as new “vertexes” in \hat{G} , and convert the weights of edges into weights of “vertexes” in \hat{G} , and add new “edges” into \hat{G} between two adjacent “vertexes” e_1 and e_2 (here adjacent means that the corresponding edges of the “vertexes” are connected head-to-tail in G) if a *valid path* passes through those “vertexes” (or equivalently, $b_{1y} \leq b_{2x}$), then the original subproblem is then equivalent to finding the shortest path (w.r.t. minimum sum of node weights instead of edge weights) from the “vertex” corresponding to e_0 to the “vertex” corresponding to e_{end} in \hat{G} . However, a simple implementation of the above using traditional Dijkstra’s algorithm together with a min-priority queue can run in $O(N^3 + N^2 \log N)$ in the worst case, which is computationally expensive, not to mention there may be a large number of “edges” that consumes much memory to store. To efficiently reduce both time and space complexity, in reality we do not construct such \hat{G} . Instead we maintain the shortest path from e_0 to other edges separately. In each step we extract the unexplored edge with the minimum distance and update its neighbors’ distances if possible. The pseudocode is as follows (Algorithm 2), which is adapted from the standard Dijkstra’s algorithm.

In Algorithm 2 we use $Out(e)$ to represent the set of outgoing edges of e . We note a key observation that in Algorithm 2 we remove some edges from $Out(e_1)$ after relaxation to further reduce the time complexity (line 20). This is because for each edge e_2 , if its $dist$ value is updated from e_1 the first time, then after that its $dist$ value cannot be updated from other e_3 connecting to e_2 . Dijkstra’s algorithm is a greedy algorithm that updates distances in an increasing manner, thus for any e_3 connecting to e_2 that updates $dist$ later than e_1 , if $b_{3y} \leq b_{2x}$, then Dijkstra’s algorithm guarantees that $dist[e_3] \geq dist[e_1]$; if $b_{3y} > b_{2x}$, then e_3 cannot further update $dist[e_2]$ due to the constraint in line 16 in Algorithm 2. In this way we can see that each edge in G can be updated at most once, and the overall time complexity can be reduced to $O(M \log M)$, or roughly $O(N^2 \log N)$ after using our algorithm. Finally, we can backtrack the shortest path from END to $START$ (line 24–29) and obtain the estimated user trajectory by mapping the matched parts of the magnetic field fingerprint data vector back to the real world coordinates.

5.4 Further Improvements

In our implementation, we also note several approaches to speed up the calculation:

- By noting that the matching score matrix $M_{m \times n}$ is updated only when a new user observation is generated and the Smith-Waterman algorithm fills the matching score matrix row by row, we can leverage results in the last row to help update the matrix and find matched substrings. More specifically, we use a max heap of fixed size to keep the matched substrings with high scores (we only keep the top $H = 100$ matched substrings). After filling the last row of M , we only need to check the newly matched substrings ending in the last row with high scores, since the others ending in the previous rows are already kept in the heap. After finding the new matched substrings, we only keep the top H matched substrings in the heap, and discard the matched substrings whose scores are relatively low. This can help reduce the amortized complexity. Here we also note that there is no need to backtrack the whole alignment until its beginning, since if checking the backtracking path for an alignment in M , then we can see that the path will eventually go through the previous row at some location l . Hence, to obtain the whole alignment we can simply concatenate the part of alignment from its end to l and the alignment starting from l , which can be updated column by column when a new user observation is generated.
- Since the length of alignments can become very large if the user walks for a long time, we also consider removing the old data periodically. After every minute, we prune all the

ALGORITHM 2: Finding Minimum Cost Path Algorithm**Input:** Constructed graph G .**Output:** Minimum cost path from $START$ to END

```

1: Add a new edge  $e_0$  to  $G$  that only connects to  $START$  with both of weights 0
2: Add another new edge  $e_{end}$  to  $G$  that only connects with  $END$  with both of weights  $INT\_MAX$ ,
   a sufficiently large integer
3:  $dist[e_0] \leftarrow 0$ 
4: for edge  $e \in G - \{e_0\}$  do
5:    $dist[e] \leftarrow \infty$ 
6: end for
7: for edge  $e \in G$  do
8:    $back[e] \leftarrow NULL$ 
9: end for
10:  $Q \leftarrow$  all edges in  $G$ 
11: while  $Q \neq \emptyset$  do
12:   Extract  $e_1 \in Q$  with the minimum distance  $dist[e_1]$ 
13:   if  $e_1 == e_{end}$  then
14:     Break
15:   end if
16:   for  $e_2 \in Out(e_1), b_{1y} \leq b_{2x}$  do
17:     if  $dist[e_2] > dist[e_1] + cost(e_2)$  then
18:        $dist[e_2] \leftarrow dist[e_1] + cost(e_2)$ 
19:        $back[e_2] = e_1$ 
20:       remove  $e_2$  from  $Out(e_1)$ 
21:     end if
22:   end for
23: end while
24:  $Backtrack \leftarrow back[e_{end}]$ 
25:  $MinCostPath \leftarrow \emptyset$ 
26: while  $Backtrack \neq e_0$  do
27:    $MinCostPath \leftarrow MinCostPath \cup \{Backtrack\}$ 
28:    $Backtrack \leftarrow back[Backtrack]$ 
29: end while

```

alignments to remove the samples with timestamps earlier than 1 minute before the current time. In this way, we also reduce the space complexity efficiently.

5.5 Time Complexity

Denote the number of magnetic field fingerprint data vectors as K , and the number of RSSI fingerprints as P . The total time complexity of the magnetic field sequence matching and selection (Algorithm 1) is

$$O\left(Kmn\left(1 + \frac{1}{R^2}\right)\right), \quad (8)$$

where m is the length of user observation sequence, n is the maximum length of magnetic field fingerprint data vector, and R is the parameter mentioned in Algorithm 1. This is due to the time complexities for matching and substring selection being $O(Kmn)$ and $O(K\frac{mn}{R^2})$, respectively. Note that after the substring selection algorithm we may select at most $O(K\frac{mn}{R^2})$ substrings in total.

After the above procedures, we convert the localization problem into a shortest path problem in $O(N^2)$, which can be solved in roughly $O(N^2 \log N)$ by using Algorithm 2, where N is the total number of matched magnetic field substrings. Thus the overall time complexity of Magil is

$$O\left(\frac{K^2 m^2 n^2}{R^4} \log\left(\frac{Kmn}{R^2}\right)\right). \quad (9)$$

The amortized time complexity is $O(K(n+H)\log(H))$, where H is the size of the heap mentioned in Section 5.4. If RSSI fingerprints are given, then the amortized time complexity becomes $O(R+K(n+H)\log(H))$, since we only need to calculate the estimated physical location of each user observation only once (Section 6).

6 MAGFI: FUSING MAGIL WITH WI-FI RSSI

Given the wide deployment of Wi-Fi, we can extend Magil to include Wi-Fi RSSI signals to achieve much better localization performance and overcome the requirement of user movement. We may regard the magnetic field as a kind of global signal that lacks specific location indicator, while the RF signal as a location-specific signal, because different areas would observe different RF signals due to complex indoor structures. This inspires us to further refine Magil to fuse it with existing RSSI fingerprints effectively.

Generally speaking, our key idea is to leverage RSSI fingerprints first to constrain the search scope in a relatively small area and perform our sequence matching algorithm in such area afterward. In such way, we fully utilize both the global location information of RF signals and detailed local position information of geomagnetic fields.

Specifically, we revise the criterion for determining whether a user observation \mathbf{B}_{o_i} matches a magnetic field fingerprint data \mathbf{B}_f (see Equation (4)). If there is a Wi-Fi sample with its timestamp close to the that of \mathbf{B}_{o_i} , then we first leverage RSSI fingerprints to calculate a roughly estimated physical location of \mathbf{B}_{o_i} . Specifically, we check the RSSI fingerprints similar to the user Wi-Fi sample based on Euclidean distance and then apply the K-NN method ($K = 5$ in our case). We then compare it with the physical location of \mathbf{B}_f . If the distance between these two locations is too far (larger than 5m in our implementation), then we consider these two samples as unmatched. Otherwise, we leverage Equation (4) to calculate whether these two samples match each other. This can effectively filter those matched substrings whose magnetic field fingerprints are quite similar to user observations but the RSSI readings are different from user observations. Thus MagFi improves the localization robustness and efficiency of Magil.

Besides, we note that when a user stands still, his/her magnetic field observation is almost fixed and stable, and we cannot achieve localization without offline device magnetometer calibration. To combat this weakness, in practice, if we detect that the variance of the user geomagnetic observations obtained within recent 5s is less than an empirical threshold $1\mu\text{T}^2$, then we directly apply the above K-NN method to the latest user Wi-Fi sample and return the estimated physical location to the user. This consideration can also improve the system performance of MagFi.

7 ILLUSTRATIVE EXPERIMENTAL RESULTS

In this section, we evaluate the accuracy and efficiency of Magil and MagFi through implementation and experiments. In Section 7.1, we introduce experimental settings and comparison schemes. We present illustrative experimental results for Magil and MagFi in Section 7.2.

7.1 Experimental Settings and Comparison Schemes

We implement our system on different mobile phones, including Sony Xperia X2, Samsung Galaxy S4, and LG Nexus 5. We use Sony Xperia X2 to perform site survey tasks. They all run Android

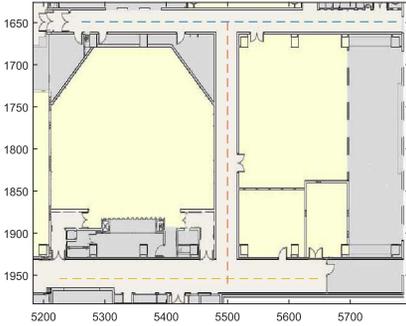


Fig. 9. Survey paths (dashed lines) in the university corridors.

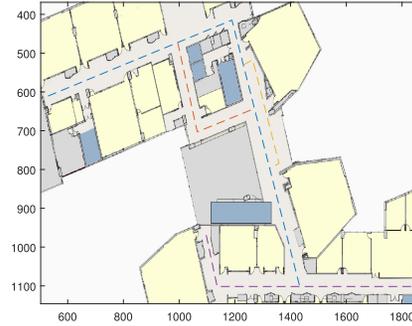


Fig. 10. Survey paths (dashed lines) in a whole floor.

operating systems of later than 4.0 version and are all equipped with magnetometers. We implement the server on a Dell PC with a 3.6GHz processor and 16G RAM, running Windows 8.1. During localization, the mobile client performs continuous background inertial sensor (magnetometer) sampling. Sampling frequency is set to be 25Hz. From the experiment, we observe that just a few steps of walking traces are sufficient to localize the user. Therefore, in practice we only need to keep a small buffer for the walking trace and discard the earlier part. Unless otherwise stated, we use the following thresholds: $Threshold_1 = 9$, $Threshold_2 = 15$, and $Threshold_3 = 10$. Later, we analyze the performance of our system under different thresholds.

In our experiments, we compare Magil with another two state-of-the-art magnetic field-based localization algorithms, and the detailed algorithms and parameter settings refer to their works.

- MaLoc [34]: utilizes a particle filter together with INS to measure the user location. Candidate locations with the best magnetic field matching are selected, then the particle filter further reduces the weights of incorrect locations and performs localization. Specifically, we implement MaLoc by setting the initial number of particles to be 2,000.
- Magical [20]: measures not only magnetic field values but also relative trends of magnetic field changes when a user is walking through an indoor environment. By utilizing Dynamic Time Warping (DTW) and particle filter, Magical maps target to locations with the best trend matching while filtering away incorrect locations. We implement Magical by setting the initial number of particles to be 3,000 as suggested in their work.

To evaluate Wi-Fi-geomagnetism fusion performance, we further set each device to its fastest Wi-Fi sampling frequency (roughly 1Hz). We compare MagFi with (1) Magical, which supports fusion with Wi-Fi signals based on two-pass bidirectional particle filtering (TBPF) and (2) Radar [1], which only uses RSSI readings and matches the incoming RSSI vector measurement against the RSSI fingerprints. Note that we use the Euclidean distance as the similarity metric and the K-NN method ($K = 5$) to estimate the location.

We have conducted extensive experiments to validate our localization algorithm in four different typical indoor environments: a small indoor environment containing three short corridors ($1,091\text{m}^2$, see Figure 9), a larger indoor environment (a whole floor) containing several long corridors ($5,909\text{m}^2$, see Figure 10), a large indoor environment containing open space and several corridors ($4,133\text{m}^2$, see Figure 11) and a large indoor environment with many connected corridors ($3,296\text{m}^2$, see Figure 12). In these figures, we also show the survey paths. Note that in a large open space, survey paths should cover various walking directions to achieve better localization accuracy. In addition, we also show the 53 Wi-Fi reference points as dots in Figure 12.

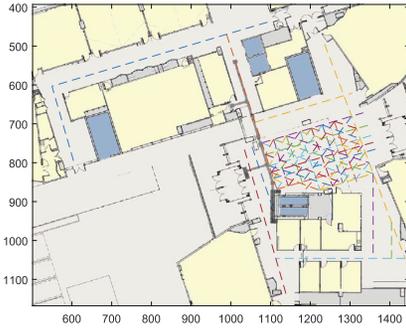


Fig. 11. Survey paths (dashed lines) in a large indoor environment.

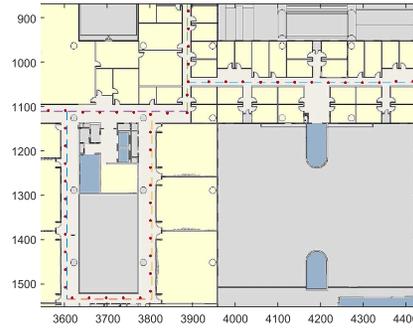


Fig. 12. Survey paths (dashed lines) and reference points (dots) in the environment with many connected corridors.

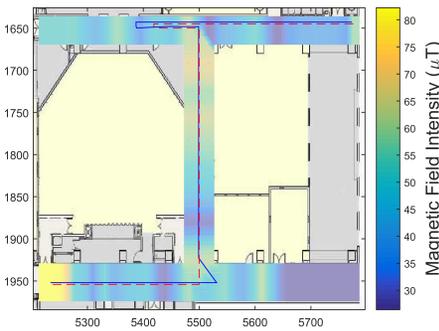


Fig. 13. The estimated walking trajectory of Magil in the corridors. The red dotted line is the ground truth and the blue line is the estimated trajectory.

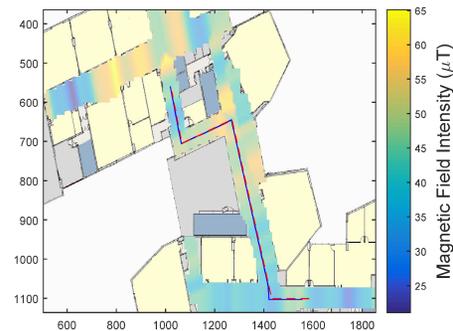


Fig. 14. The estimated walking trajectory of Magil throughout the whole floor. The red dotted line is the ground truth and the blue line is the estimated trajectory.

The performance metrics are presented as below. To obtain the ground truth of walking trajectories, we set many landmarks (say, the doors and pillars) and measure their locations in advance. During the experiment, users record the time they pass by those landmarks to evaluate the real-time location estimation. *Localization error* is calculated as the Euclidean distance between ground truths and estimated locations in historical trajectories, while the *real-time localization error* is calculated as localization error at each step. The *mean localization error* is calculated as the mean of all the estimation errors along a trace.

7.2 Illustrative Results

Geomagnetism-only (Magil) Localization Performance: Figures 13, 14, 15, 16, and 17 show the estimated walking traces compared with ground truths in corridors, a whole floor, a large indoor environment, and an environment with many connected corridors, respectively. We can see that estimations highly match ground truths.

We also compare Magil with other magnetic field-based localization systems such as Magicol and MaLoc. From Figures 18, 19, 20, and 21, we can clearly see that our system works well in both narrow spaces (such as corridors) and large open spaces (as shown in Figure 15) and outperforms other state-of-the-art systems with lower error and smaller deviation.

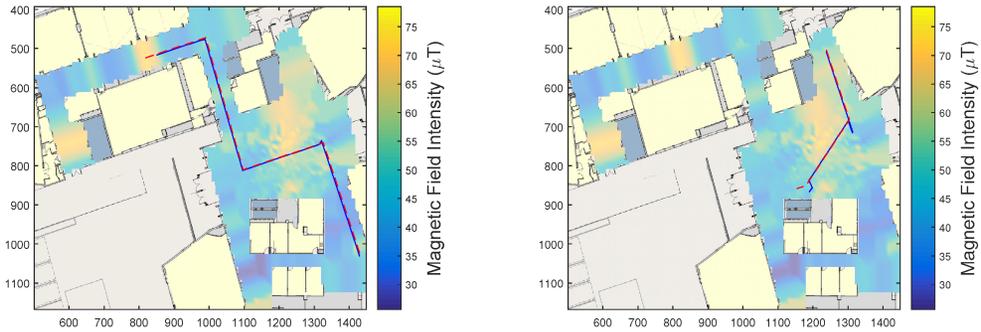


Fig. 15. Two estimated walking trajectories of Magil in the large indoor environment. The red dotted lines are the ground truths and the blue lines are the estimated trajectories.

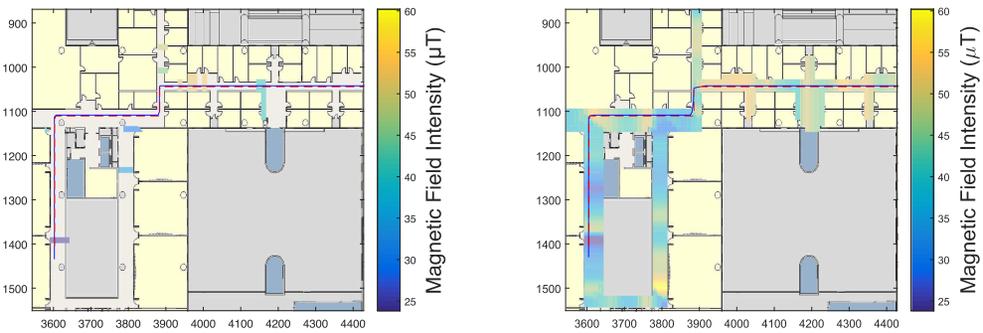


Fig. 16. The estimated walking trajectory of MagFi in the environment with many connected corridors. The left figure shows the results using only magnetic field while the right figure shows the localization performance of fusing both Wi-Fi and magnetic field. The red dotted lines are the ground truths and the blue lines are the estimated trajectories.

Localization Performance of RF-Geomagnetism Fusion (MagFi): For fusion with RSSI fingerprints, we demonstrate the localization accuracy among different systems such as Magicol and Radar in Figures 22, 23, 24, and 25. Specifically, in Figures 22 and 24 we compare MagFi using only geomagnetism with state-of-the-art Magicol and MaLoc. In Figures 23 and 25, we compare our geomagnetism-Wi-Fi fusion version with Wi-Fi-enabled Magicol and Radar. We can see that while Magil achieves comparable performances to Radar (Wi-Fi-based localization system) on its own without fusion, the combination of signals leads to a more accurate performance improvement than using any individual method. Specifically, MagFi with signal fusion further cuts the localization error by at least 15% compared with the previous one without fusion [30]. Furthermore, when compared with Wi-Fi-enabled Magicol, our system yields more robust and stable results with smaller deviation. As Magil with Wi-Fi fusion (MagFi) generally outperforms other state-of-the-art systems, in the following evaluations we focus on Magil without fusing Wi-Fi signals (hence does not require RSSI fingerprints).

System Performance under Different Settings: In the experiment, we also ask another eight subjects in addition to user 1 to walk along the ground-truth trajectory as shown in Figure 13 to validate Magil’s performance. In sum, there are seven males and two females, and their heights and weights vary from 1.68m to 1.87m, and from 45kg to 80kg, respectively. Figure 26 shows the mean error and standard deviation for each user. We can observe that Magil achieves high accuracy in localization among different users.

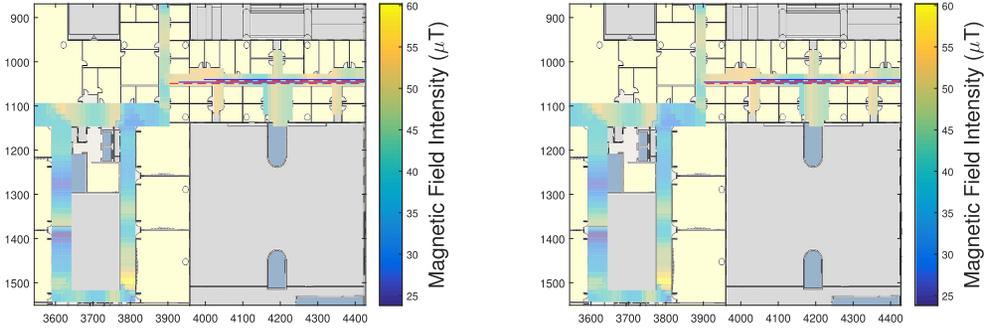


Fig. 17. Another estimated walking trajectory of MagFi in the environment with many connected corridors. The left figure shows the results using only magnetic field while the right figure shows the localization performance of fusing both Wi-Fi and magnetic field. The red dotted lines are the ground truths and the blue lines are the estimated trajectories.

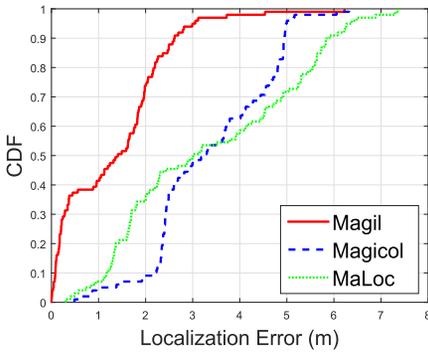


Fig. 18. CDF of indoor localization error in the corridors.

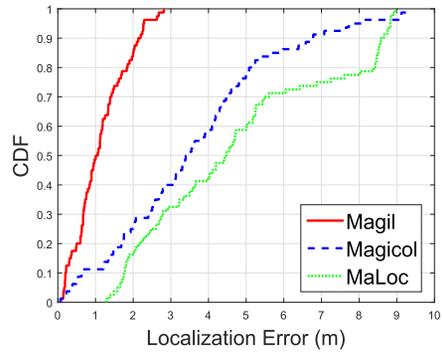


Fig. 19. CDF of indoor localization error throughout the whole floor.

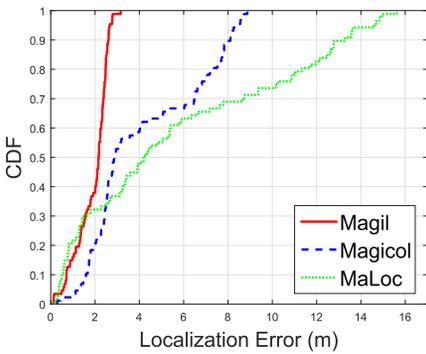


Fig. 20. CDF of indoor localization error in the large indoor environment (the left trajectory in Figure 15).

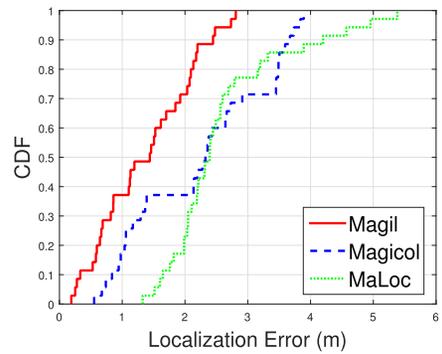


Fig. 21. CDF of indoor localization error in the large indoor environment (the right trajectory in Figure 15).

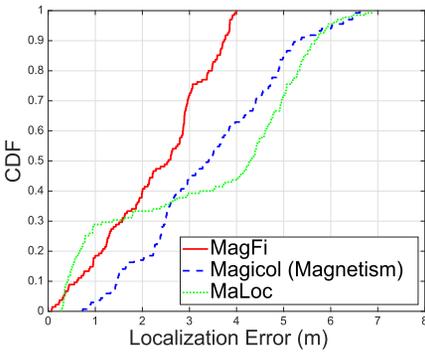


Fig. 22. CDF of indoor localization error in the environment with many connected corridors (the left trajectory in Figure 16).

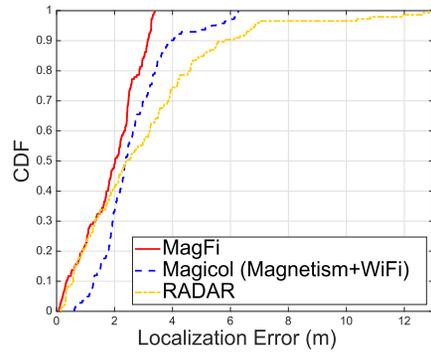


Fig. 23. CDF of fusion indoor localization error in the environment with many connected corridors (the right trajectory in Figure 16).

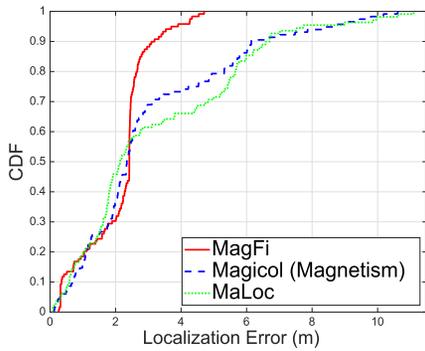


Fig. 24. CDF of indoor localization error in the environment with many connected corridors (the left trajectory in Figure 17).

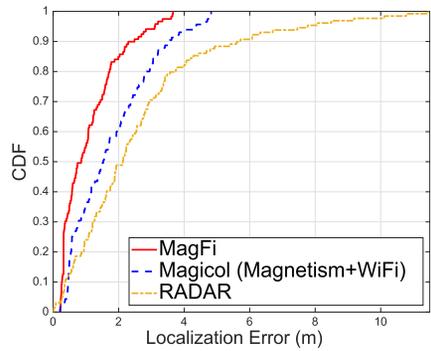


Fig. 25. CDF of fusion indoor localization error in the environment with many connected corridors (the right trajectory in Figure 17).

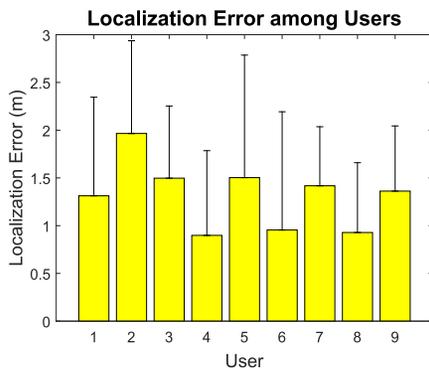


Fig. 26. Mean localization error (with error bars) among users.

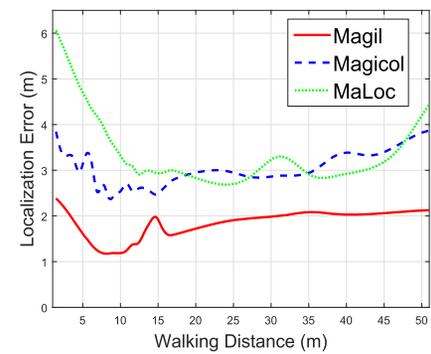


Fig. 27. Real-time localization error versus user walking distances.

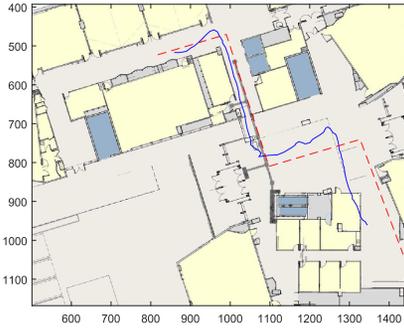


Fig. 28. Particle filter-based magnetic field localization results in the large indoor environment. The red dotted line is the ground truth and the blue line is the estimation. The black lines intersecting the ground truth indicate the accessibility design of the building, showing accessible paths for mobility/visual impaired people.

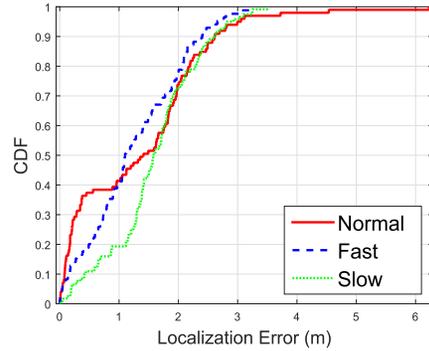


Fig. 29. CDFs of the localization error versus user walking speeds.

Figure 27 shows the average real-time localization error among different systems. For particle filter-based systems, we repeat the followings 10 times and show their average. In each experiment, we manually choose a random initial position near the accurate initial position and walk on the same trajectories afterward. We can see that our system converges more quickly and yields stabler localization results, while other particle filter-based systems cannot converge well. Note that traditional particle filter requires an explicit or accurate initial position of the target and relies heavily on accurate estimation of user movement such as steps, walking distances, and directions. Otherwise, given a coarse initial position, it can work with large localization error, as shown in Figure 28. Since particle filter tries to capture all possible step length and direction error information by means of particles, it cannot converge quickly in large open spaces or complex environments [8], which still applies in magnetic field fusion. As the number of dimensions of magnetic field observations is less than Wi-Fi or Bluetooth, the dispersion of particles can be severe in large open spaces. Magil elegantly avoids this dilemma by matching magnetic field observations instead of estimating actual user movements, which is error-prone [2]. Moreover, our system models user movements as transitions between several possible trajectories, which is computationally cheap.

We are also interested in the system performance under different user moving speeds. In this experiment a user walks along the trajectory shown in Figure 13 at three different speeds: walking at fast (roughly 1.8m/s), normal (roughly 1.4m/s), and slow (roughly 1.0m/s) speeds. Figure 29 illustrates that the overall performance is not influenced much under different speeds. This is because our modified Smith-Waterman algorithm (Algorithm 1) can align the shapes of magnetic field observations, which do not alter under different walking speeds along the same trajectory. Different walking speeds can only affect the length but not the shape of the magnetic field observation sequence, which indicates the effectiveness of Magil.

Note that we use several empirical thresholds in our system settings. However, different thresholds may also influence the system performance. Figures 30, 31, and 32 show the mean localization error and corresponding standard deviations versus different thresholds. In this experiment, the user walks along the trajectory shown in Figure 13. We can see that a too-large $Threshold_1$ or $Threshold_2$ will incur a large error, since it will skip more potentially useful substrings. Meanwhile,

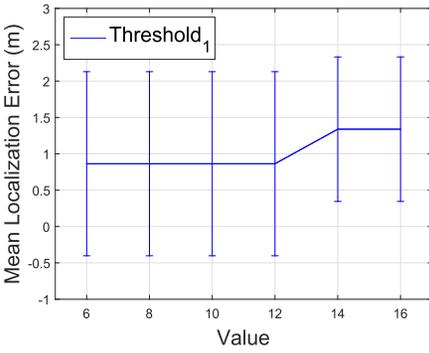


Fig. 30. Mean localization error versus $Threshold_1$.

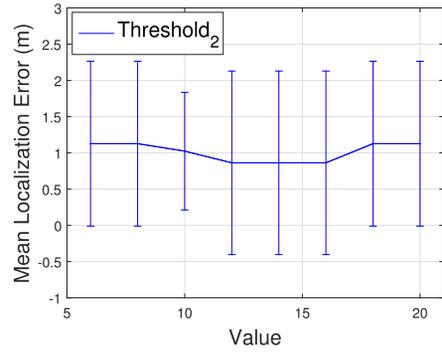


Fig. 31. Mean localization error versus $Threshold_2$.

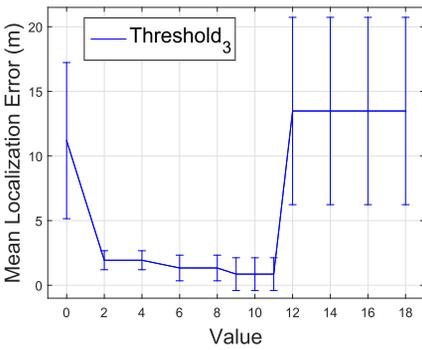


Fig. 32. Mean localization error versus $Threshold_3$.

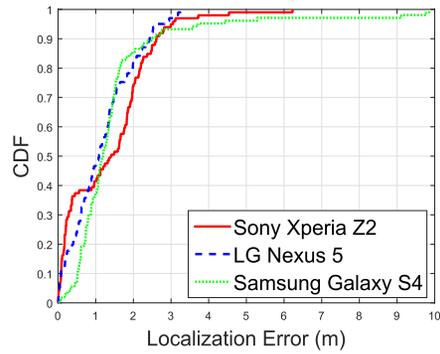


Fig. 33. CDF of localization error among different smartphones.

small $Threshold_1$ or $Threshold_2$ can introduce more substrings that have too-low matching scores or are too short, respectively, under which the system performance degrades. A large $Threshold_3$ will allow more insertions and deletions during matching, which can lead to more wrong matchings. However, a small $Threshold_3$ can also make the localization accuracy lower due to the stricter matching process. By setting those empirical thresholds mentioned above, we can maintain both accuracy and robustness of Magil.

Since the user may use different devices other than the one used for magnetic field fingerprint collection, we also evaluate Magil’s efficiency by using different smartphones. Figure 33 shows the CDFs of localization error among different smartphones. We can see that our system’s performance is good and stable, since we use Equations (5) and (6) to calculate similarity between observations instead of using the magnetic field fingerprint itself directly to avoid calibrating different magnetometers.

Next we investigate the cost of Magil on a mobile phone (Sony Xperia Z2). For a responsive and user-friendly pedestrian tracking system, the typical processing time for each step during localization should be less than 600ms. Any application with requirements exceeding this limit can result in an unpleasant user experience. Table 2 shows the execution time of various systems on the corridor’s trajectory as shown in Figure 13. Time is calculated as the average time it takes to process a single step over the trajectory. We can see that Magil outperforms other systems in online execution time, whilst obtaining the lowest localization error. Note that by using the

Table 2. Computation Time Comparison between Different Magnetic Field Localization Algorithms

System	MagFi	Magil	Magicol	MaLoc
Time (ms)	196.2	215.0	1,781.4	700.7

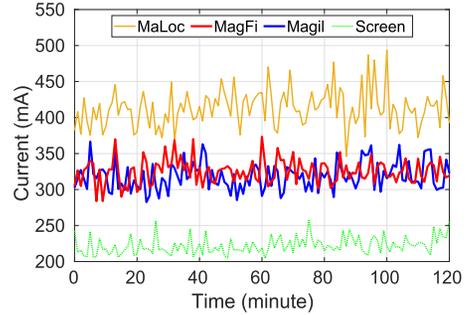


Fig. 34. Comparison of power consumption among different magnetic field localization algorithms.

speedup procedures introduced in Section 5.4, we achieve a $2.3\times$ speedup over the preliminary version of Magil [30] (roughly 498.1ms). It is mainly because we reduce the amortized complexity significantly. Note that MagFi runs faster than Magil, because the matching complexity is reduced due to constrained search scope of MagFi compared with Magil, and their computation time is much less than other schemes.

Finally, we compare energy consumption of different systems in terms of electric current. We run each algorithm for 2 hours on the same device (LG Nexus 5) and use Emmagee [17], an open-source Android-based system performance monitoring tool, to record the real-time current. Figure 34 presents the average current of each application tool over time. The average current is 412.81mA, 335.41mA, 321.43mA, and 224.25mA for MaLoc, MagFi, Magil, and screen energy consumption, respectively. We can observe that both MagFi and Magil have much lower energy consumption than MaLoc. The reason is that they get rid of two major power-hungry components, i.e., pedometer (which uses accelerometer and gyroscope dramatically) and particle filter (which requires more computations). However, MagFi does not consume too much extra energy compared with Magil ($\sim 4.3\%$). This is because Wi-Fi scanning frequency in MagFi is usually low (roughly 1Hz), which does not introduce much burden.

8 CONCLUSION AND FUTURE WORKS

In this article, we have proposed and studied Magil, a novel indoor localization scheme based on geomagnetic field without a pedometer. Magil models the localization as two optimization problems. It first finds several path segments in its database whose fingerprint variations best match the target observations (the so-called sequence matching problem). Given these segments, Magil employs a modified shortest path algorithm to select and order those matched parts to obtain the target locations (the so-called path construction problem). To further enhance its accuracy, computational efficiency, and deployability, we propose MagFi, an extension that fuses Magil with Wi-Fi RSSI fingerprints.

Our schemes do not require any pedometer or user walking model. Compared with the traditional particle filter approach, Magil and MagFi are much more implementable, computationally efficient, and robust against user random behaviors when deployed on smartphones. They achieve significantly better localization accuracy. Extensive experimental studies at our university campus have shown that Magil and MagFi outperform the state-of-the-art schemes by a large margin (often cutting localization error by more than 30%).

We discuss future works for Magil as follows:

- *Necessity of User Walking*: One limitation is that user needs to walk a distance before Magil can obtain enough magnetic field data to analyze the pattern and localize the user. This is because if a user stands still, then his/her magnetic field observation is almost fixed and stable, and we cannot achieve localization without offline device magnetometer calibration. While this problem has been overcome with MagFi, we are currently investigating how to infer location without the help of other signals.
- *Survey Path Planning*: Note that we require survey paths in large open space to cover various directions, as shown in Figure 11. This is because Magil performs pattern matching on geomagnetic fingerprints, which also limits our localization results to several predefined survey paths. Since the magnetic field is omnipresent in the whole space, our ongoing work is to study effective methods to localize users based on the entire geomagnetic fields of the site, including crowdsourced labeling [38] or other low-cost sensor fusion.

REFERENCES

- [1] P. Bahl and V. N. Padmanabhan. 2000. RADAR: An in-building RF-based user location and tracking system. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM'00)*, Vol. 2. 775–784. DOI: <https://doi.org/10.1109/INFOCOM.2000.832252>
- [2] Agata Brajdic and Robert Harle. 2013. Walk detection and step counting on unconstrained smartphones. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'13)*. 225–234.
- [3] Jaewoo Chung, Matt Donahoe, Chris Schmandt, Ig-Jae Kim, Pedram Razavai, and Micaela Wiseman. 2011. Indoor location sensing using geo-magnetism. In *Proceedings of the ACM International Conference on Mobile Systems, Applications, and Services (MobiSys'11)*. 141–154. DOI: <https://doi.org/10.1145/1999995.2000010>
- [4] M. Frassl, M. Angermann, M. Lichtenstern, P. Robertson, B. J. Julian, and M. Doniec. 2013. Magnetic maps of indoor environments for precise localization of legged and non-legged locomotion. In *Proceedings of the IEEE RSJ International Conference on Intelligent Robots and Systems (RSJ IROS'13)*. 913–920.
- [5] Dongsoo Han, Sukhoon Jung, Minkyu Lee, and Giwan Yoon. 2014. Building a practical Wi-Fi-based indoor navigation system. *IEEE Perv. Comput.* 13, 2 (Apr. 2014), 72–79. DOI: <https://doi.org/10.1109/MPRV.2014.24>
- [6] Suining He and S.-H. Gary Chan. 2016. Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons. *IEEE Commun. Surv. Tutor.* 18, 1 (2016), 466–490.
- [7] Suining He and Kang G. Shin. 2018. Geomagnetism for smartphone-based indoor localization: Challenges, advances, and comparisons. *ACM Comput. Surv.* 50, 6 (2018), 97.
- [8] Sebastian Hilsenbeck, Dmytro Bobkov, Georg Schroth, Robert Huitl, and Eckehard Steinbach. 2014. Graph-based data fusion of pedometer and WiFi measurements for mobile indoor positioning. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'14)*. 147–158.
- [9] A. Kendall, M. Grimes, and R. Cipolla. 2015. PoseNet: A convolutional network for real-time 6-DOF camera relocalization. In *Proceedings of the 2015 IEEE International Conference on Computer Vision*. DOI: <https://doi.org/10.1109/ICCV.2015.336>
- [10] Eamonn Keogh, Kaushik Chakrabarti, Michael Pazzani, and Sharad Mehrotra. 2001. Dimensionality reduction for fast similarity search in large time series databases. *Knowl. Inf. Syst.* 3, 3 (2001), 263–286.
- [11] Binghao Li, T. Gallagher, A. G. Dempster, and C. Rizos. 2012. How feasible is the use of magnetic field alone for indoor positioning? In *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN'12)*. 1–9.
- [12] Fan Li, Chunshui Zhao, Guanzhong Ding, Jian Gong, Chenxing Liu, and Feng Zhao. 2012. A reliable and accurate indoor localization method using phone inertial sensors. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'12)*. 421–430.
- [13] Mingkuan Li, Ning Liu, Qun Niu, Chang Liu, S.-H. Gary Chan, and Chengying Gao. 2018. SweepLoc: Automatic video-based indoor localization by camera sweeping. *Proc. ACM Interact. Mob. Wear. Ubiqu. Technol.* 2, 3, Article 120 (Sep. 2018), 25 pages. DOI: <https://doi.org/10.1145/3264930>
- [14] Dimitrios Lymberopoulos, Jie Liu, Xue Yang, Romit Roy Choudhury, Vlado Handziski, and Souvik Sen. 2015. A realistic evaluation and comparison of indoor location technologies: Experiences and lessons learned. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*. ACM, 178–189.
- [15] J. Ma, J. Qian, P. Li, R. Ying, and P. Liu. 2013. Indoor localization based on magnetic anomalies and pedestrian dead reckoning. In *Proceedings of the Institute of Navigation GNSS Conference (ION GNSS+ '13)*. 1033–1038.
- [16] Andrew Markham and Niki Trigoni. 2012. Magneto-inductive networked rescue system (MINERS): Taking sensor networks underground. In *Proceedings of the ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN'12)*. 317–328.

- [17] NetEase, Inc. 2015. Android performance test tool-CPU, memory, network traffic, starting time, battery current and status. Retrieved from <https://github.com/NetEase/Emmagee>.
- [18] Qun Niu, Mingkuan Li, Suining He, Chengying Gao, S. H. Gary Chan, and Xiaonan Luo. 2019. Resource-efficient and automated image-based indoor localization. *ACM Trans. Sens. Netw.* 15, 2 (Feb. 2019), 1–31. DOI : <https://doi.org/10.1145/3284555>
- [19] Jun-geun Park and Seth Teller. 2014. *Motion Compatibility for Indoor Localization*. Technical Report. MIT.
- [20] Yuanchao Shu, Cheng Bo, Guobin Shen, Chunshui Zhao, Liqun Li, and Feng Zhao. 2015. Magicol: Indoor localization using pervasive magnetic field and opportunistic WiFi sensing. *IEEE J. Select. Area Commun.* 33, 7 (Jul. 2015), 1443–1457.
- [21] Yuanchao Shu, Kang G. Shin, Tian He, and Jiming Chen. 2015. Last-mile navigation using smartphones. In *Proceedings of the ACM International Conference on Mobile Systems, Applications, and Services (MobiCom'15)*. 512–524.
- [22] Temple F. Smith and Michael S. Waterman. 1981. Identification of common molecular subsequences. *J. Molec. Biol.* 147, 1 (1981), 195–197.
- [23] Kalyan Pathapati Subbu, Brandon Gozick, and Ram Dantu. 2013. LocateMe: Magnetic-fields-based indoor localization using smartphones. *ACM Trans. Intell. Syst. Technol.* 4, 4, Article 73 (Oct. 2013), 27 pages. DOI : <https://doi.org/10.1145/2508037.2508054>
- [24] X. Tian, R. Shen, D. Liu, Y. Wen, and X. Wang. 2017. Performance analysis of RSS fingerprinting based indoor localization. *IEEE Trans. Mobile Comput.* 16, 10 (Oct. 2017). DOI : <https://doi.org/10.1109/TMC.2016.2645221>
- [25] He Wang, Souvik Sen, Ahmed Elgohary, Moustafa Farid, Moustafa Youssef, and Romit Roy Choudhury. 2012. No need to war-drive: Unsupervised indoor localization. In *Proceedings of the ACM International Conference on Mobile Systems, Applications, and Services (MobiSys'12)*. 197–210.
- [26] Oliver Woodman and Robert Harle. 2008. Pedestrian localisation for indoor environments. In *Proceedings of the 10th International Conference on Ubiquitous Computing*. ACM, 114–123.
- [27] Chenshu Wu, Jingao Xu, Zheng Yang, Nicholas D. Lane, and Zuwei Yin. 2017. Gain without pain: Accurate WiFi-based localization using fingerprint spatial gradient. *Proc. ACM Interact. Mobile Wear. Ubiqu. Technol.* 1, 2 (Jun. 2017), 29:1–29:19. DOI : <https://doi.org/10.1145/3090094>
- [28] Chenshu Wu, Zheng Yang, and Yunhao Liu. 2015. Smartphones based crowdsourcing for indoor localization. *IEEE Trans. Mobile Comput.* 14, 2 (2015), 444–457.
- [29] Hang Wu, Suining He, and S.-H. Gary Chan. 2017. Accurate and efficient localization with geomagnetism-pedometer graphical model fusion. In *Proceedings of the 14th International Conference on Mobile Ad-hoc and Sensor Systems (IEEE MASS'17)*.
- [30] Hang Wu, Suining He, and S.-H. Gary Chan. 2017. Efficient sequence matching and path construction for geomagnetic indoor localization. In *Proceedings of the International Conference on Embedded Wireless Systems and Networks (EWSN'17)*. 156–167.
- [31] Kaishun Wu, Jiang Xiao, Youwen Yi, Dihua Chen, Xiaonan Luo, and Lionel M. Ni. 2013. CSI-based indoor localization. *IEEE Trans. Parallel Distrib. Syst.* 24, 7 (Jul. 2013), 1300–1309.
- [32] Zhuoling Xiao, Hongkai Wen, A. Markham, and N. Trigoni. 2014. Lightweight map matching for indoor localisation using conditional random fields. In *Proceedings of the ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN'14)*. 131–142.
- [33] H. Xie, T. Gu, X. Tao, H. Ye, and J. Lu. 2016. A reliability-augmented particle filter for magnetic fingerprinting based indoor localization on smartphone. *IEEE Trans. Mobile Comput.* 15, 8 (Aug. 2016), 1877–1892.
- [34] Hongwei Xie, Tao Gu, Xianping Tao, Haibo Ye, and Jian Lv. 2014. MaLoc: A practical magnetic fingerprinting approach to indoor localization using smartphones. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'14)*. 243–253.
- [35] Weixing Xue, Weining Qiu, Xianghong Hua, and Kegen Yu. 2017. Improved Wi-Fi RSSI measurement for indoor localization. *IEEE Sens. J.* 17, 7 (2017), 2224–2230.
- [36] Moustafa Youssef and Ashok Agrawala. 2005. The Horus WLAN location determination system. In *Proceedings of the ACM International Conference on Mobile Systems, Applications, and Services (MobiSys'05)*. 205–218.
- [37] Chi Zhang, Jun Luo, and Jianxin Wu. 2014. A dual-sensor enabled indoor localization system with crowdsensing spot survey. In *Proceedings of the IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS'14)*. IEEE, 75–82.
- [38] Chi Zhang, Kalyan P. Subbu, Jun Luo, and Jianxin Wu. 2015. GROPING: Geomagnetism and crowdsensing powered indoor navigation. *IEEE Trans. Mobile Comput.* 14, 2 (2015), 387–400.
- [39] Pengfei Zhou, Mo Li, and Guobin Shen. 2014. Use it free: Instantly knowing your phone attitude. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*. ACM, 605–616.

Received August 2018; revised March 2019; accepted June 2019