MagPIE: A Dataset for Indoor Positioning with Magnetic Anomalies

David Hanley\textsuperscript{1}, Alexander B. Faustino\textsuperscript{2}, Scott D. Zelman\textsuperscript{2}, David A. Degenhardt\textsuperscript{2}, and Timothy Bretl\textsuperscript{2}

Department of Electrical and Computer Engineering\textsuperscript{1}
Department of Aerospace Engineering\textsuperscript{2}
University of Illinois at Urbana-Champaign

Email: \{hanley6,afausti2,szelman2,degenha2,tbretl\}@illinois.edu

Abstract—In this paper, we present a publicly available dataset for the evaluation of indoor positioning algorithms that use magnetic anomalies. Our dataset contains IMU and magnetometer measurements along with ground truth position measurements that have centimeter-level accuracy. To produce this dataset, we collected over 13 hours of data (51 kilometers of total distance traveled) from three different buildings, with sensors both handheld and mounted on a wheeled robot, in environments with and without changes in the placement of objects that affect magnetometer measurements ("live loads"). We conclude the paper with a discussion of why these characteristics of our dataset are important when evaluating positioning algorithms.

Index Terms—Magnetic Localization, Indoor Localization, Dataset, Comparison of Methods

I. INTRODUCTION

Magnetic anomalies are variations in the Earth’s magnetic field that can be used for indoor positioning of both smartphones and robots. In indoor environments, the absence of a global navigation satellite system necessitates alternative positioning methods. Some indoor localization strategies for smartphones (such as Qualcomm’s Lumicast \cite{1}) require infrastructure specifically for the purpose of positioning. In the case of Lumicast, a set of light fixtures must broadcast positioning signals using visible light communication (VLC). Wi-Fi based positioning, beyond requiring the presence of a number of access points, can also require knowing the position of these access points and can suffer inconsistent fingerprints due to blockage from users or other pedestrians \cite{2}. In many cases, a localization solution that does not require such infrastructure is desired.

Several sensors used in robot or smartphone localization that do not require infrastructure like VLC lights or the use of access points include magnetometers, cameras, and LiDAR. Magnetometers offer several advantages to both camera and LiDAR localization. In robotic applications, there are some scenarios where cameras may not be allowed due to security and privacy concerns \cite{4}. Performance of camera-based positioning can also be significantly degraded due to low lighting or smoke \cite{4}. Magnetic field-based positioning systems, consisting of an inertial measurement unit (IMU) and magnetometer, have a cost advantage over other sensors commonly used for localization \cite{5}. Table I compares the cost of a magnetometer that can be found on a robot or smartphone with an IMU, camera, or LiDAR system. Magnetometers, moreover, are ubiquitous on mobile robot platforms and smartphones.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Example</th>
<th>Cost</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smartphone IMU</td>
<td>InvenSense MPU-6500</td>
<td>$3.50</td>
<td>10</td>
</tr>
<tr>
<td>Magnetometer</td>
<td>NXP MAG3110</td>
<td>$1.46</td>
<td>7</td>
</tr>
<tr>
<td>Camera</td>
<td>mvBlueFOX-MLC200w</td>
<td>$310.00</td>
<td>17</td>
</tr>
<tr>
<td>LiDAR</td>
<td>Velodyne PUCK VLP-16</td>
<td>$9999.00</td>
<td>18</td>
</tr>
</tbody>
</table>

The use of magnetometers to localize a smartphone or robot indoors is becoming common. For example, companies such as IndoorAtlas have developed indoor positioning solutions using magnetic anomalies \cite{10}. With respect to localization algorithms, methods exist to localize smartphones in a building using pedestrian dead reckoning (PDR) and a magnetic field \cite{11}–\cite{14}. There also exist methods for localizing smartphones using Wi-Fi and magnetic maps together \cite{2}, \cite{15}–\cite{18}. For example, when comparing methods for Wi-Fi surveying and localization with a magnetometer-based approach, Gao and Harle \cite{16} found that magnetic signals were particularly useful for detecting loop closure with smartphones. Magnetometers have also been used to position cleaning robots, cars, and unmanned ground vehicles (UGVs) \cite{4}, \cite{5}, \cite{19}–\cite{21}.

Considering the growing popularity of positioning with magnetic anomalies, it is necessary to establish publicly available datasets so that these methods can be compared. The first dataset presented for the purpose of evaluating positioning algorithms based on magnetic anomalies was the UJIIndoorLoc-Mag dataset of Torres-Sospedra et al \cite{12}. This dataset collected magnetic field and IMU measurements from a smartphone in a 15 × 20 meter office space. Another dataset, presented by Barsocchi et al \cite{22}, was developed to compare approaches to positioning methods using inertial sensors and Wi-Fi to methods using inertial sensors and magnetometers.

Our primary contribution is to present a new, publicly available dataset—the Magnetic Positioning Indoor Estimation (MagPIE) dataset\cite{1}—that can be used to further benchmark and study approaches to magnetic and inertial positioning. This...
dataset provides magnetometer, IMU, and ground truth position and orientation measurements collected in 723 separate trials over a total duration of more than 13 hours, a total distance of more than 51 kilometers, and a total area of more than 960 square meters. We describe the system, methods, and experiments used to produce this dataset in Sections II-IV.

The MagPIE dataset has three characteristics that differentiate it from what has been available previously:

- Data were collected from multiple buildings with a large total test area (960 square meters of floor space in three different buildings compared to, for example, 185 square meters in one building [22]). We show in Section V-A that methods of indoor positioning may perform quite differently in different buildings.

- Data were collected with sensors that were both handheld (as in prior datasets) and mounted on a wheeled robot (new to our dataset). We show in Section V-B that indoor positioning algorithms are, by necessity, different in these two cases, because of the presence or absence of well-defined steps into which IMU measurements can be segmented.

- Data were collected both with and without changes in the placement of objects that may affect magnetometer measurements. Prior datasets were based on the assumption that all magnetic anomalies are constant over time (i.e., are “dead magnetic loads”), while our dataset is based on the assumption that certain magnetic anomalies are only temporary (i.e., are “live magnetic loads”). We show evidence in Section V-C that supports this assumption.

Our dataset can provide a benchmark for testing magnetometer map surveying, localization, SLAM, and loop closure approaches. It can also be used to improve algorithms for PDR or gravitation tracking (such as Sarkka et al [23]).

II. DATA COLLECTION: SYSTEM

A. Sensing Platforms

As our sensing platform, we used a Motorola Moto Z Play smartphone collecting magnetometer data at approximately 50 Hz and accelerometer and gyroscope data at approximately 200 Hz. The Allan standard deviation plots—which describe the angle and velocity random walk, bias instability, and rate random walk—of accelerometers and gyroscopes on the smartphone are shown in Figures 1 and 2. These plots were created from a 3908 second static test, and data used to produce them have been provided on the dataset website. The Allan standard deviation plots shown can be used to tune Kalman filtering approaches to positioning with this dataset (see El-Sheimy et al [24]). The norm of magnetometer measurements on the smartphone were found to have a standard deviation of 0.495 μT. The Motorola Moto Z Play was used for both pedestrian and UGV tests.

B. Ground Truth

The position and orientation estimates produced by an application using Google’s Tango API on a Lenovo Phab 2 Pro were taken to be the ground truth measurements for our dataset. As a way to assess our ability to use Google’s Tango as a source of ground truth for this application, we mapped and then localized our Phab 2 Pro inside a motion capture arena. We then computed the absolute trajectory error (ATE), as described by Sturm et al [25]. The root mean square absolute trajectory error over five different trials is shown in Table II. Figure 3 shows our two smartphones arranged for the UGV and walking case.

III. DATA COLLECTION: METHODS

Each trial of our dataset was collected according to the pipeline outlined in Figure 4. First, our test areas were mapped using the Google Tango application. Subsequently,

<table>
<thead>
<tr>
<th>Trial</th>
<th>RMS ATE (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.92</td>
</tr>
<tr>
<td>2</td>
<td>3.13</td>
</tr>
<tr>
<td>3</td>
<td>4.20</td>
</tr>
<tr>
<td>4</td>
<td>4.19</td>
</tr>
<tr>
<td>5</td>
<td>3.52</td>
</tr>
</tbody>
</table>

Fig. 1. Smartphone accelerometer Allan standard deviation, which describes the sensor’s velocity random walk, bias instability, and rate random walk.

Fig. 2. Smartphone gyroscope Allan standard deviation, which describes the sensor’s angle random walk, bias instability, and rate random walk.
we collected data on our smartphone and UGV outside to account for magnetometer biases. Immediately, prior to collecting all relevant magnetometer, accelerometer, gyroscope, and ground truth data, we kept our Motorola Moto Z Play still for approximately thirty seconds for coarse quasi-stationary initialization to be performed. While in between obtaining data for individual trials, we occasionally collected additional data for magnetometer calibration. Offline, we compared all magnetometer calibrations to ensure that significant biases were not inadvertently introduced during testing.

A. Mapping with Tango

We used Google Tango’s API to create an application that can map an area and subsequently localize our Lenovo Phab 2 Pro in that map. In all cases, in order to ensure a higher quality map, we added visual features (in the form of fiducial markers called AprilTags [26]) to the test area prior to mapping. We explored each area and provided multiple opportunities for drift in the map to be corrected through the identification of loop closures. Each map of the environment was saved in an area description file. That file was then used to localize our test devices. Fiducial markers remained in the scene throughout the relevant area of the associated building in the dataset.

B. Calibrating Magnetometers

Prior to our tests, we collected data to compute the hard iron offset vector, \( b_m \), and soft iron matrix, \( \Lambda \), of our magnetometers. These offsets were applied to raw magnetometer measurements as described by [27]:

\[
m_c = \Lambda (m_{raw} - b_m).
\] (1)

A triad of uncalibrated and orthogonal magnetometers create an ellipsoid when spun about a point. The hard iron offset describes how far from the origin the center of that ellipsoid is, while the soft iron offset matrix describes the eccentricity of the ellipsoid. Equation 1 projects raw magnetometer measurements onto a sphere centered at the origin with a radius of approximately 53 \( \mu T \), since this is roughly the magnitude of Earth’s magnetic field in Urbana, IL.

C. Initialization

Prior to each test, we kept our test devices still for approximately thirty seconds to allow for the computation of the initial orientation of the device relative to the plumb-bob gravity vector, which we call quasi-stationary initialization. In particular, we computed coarse estimates of the initial orientation of the device by claiming that the direction cosine matrix \( C \) of the IMU’s frame of reference in terms of the local frame of reference, which is NED, as described by Savage [28]:

\[
C = \begin{bmatrix}
C_{(1,1)} & C_{(1,2)} & C_{(1,3)} \\
0 & \frac{C_{(2,2)}}{\sqrt{C_{(2,2)}^2 + C_{(3,2)}^2}} & \frac{-C_{(2,3)}}{\sqrt{C_{(2,2)}^2 + C_{(3,2)}^2}} \\
\left(\frac{a_{SF}}{\|a_{SF}\|}\right)_x & \left(\frac{a_{SF}}{\|a_{SF}\|}\right)_y & \left(\frac{a_{SF}}{\|a_{SF}\|}\right)_z
\end{bmatrix}
\] (2)

where

\[
C_{(1,j)} = C_{(2,2)} C_{(3,j)} - C_{(2,j)} C_{(3,3)}
\]

\[
C_{(1,2)} = C_{(2,3)} C_{(3,1)} - C_{(2,1)} C_{(3,3)}
\]

\[
C_{(1,3)} = C_{(2,1)} C_{(3,2)} - C_{(2,2)} C_{(3,1)}
\]

Here \( C_{(i,j)} \) refers to the element of \( C \) on the \( i \)th row and \( j \)th column, \( a_{SF} \) is the specific force vector measured by our accelerometers (in \( m/s^2 \)) and \((\cdot)_x, (\cdot)_y, \text{ and } (\cdot)_z\) are the \( x, y, \) and \( z \) components of the contained vector respectively.

IV. DATA COLLECTION: EXPERIMENTS

Leveraging the structure of previously used datasets (i.e. the UJJIndoorLoc-Mag dataset [12]), we separated our collection process into training and testing trials. Typically, we would anticipate that training trials are used to generate magnetic maps, while testing trials are used to assess how well proposed algorithms localize within these generated maps. To accommodate differing use cases, we performed tests for walking with a handheld phone (as if one was texting) and for a UGV. Finally, we conducted two different testing scenarios: one without any imposed changes to the magnetic field of the building (dead load cases) and a second set with objects (or live loads) added to the scene that change the magnetic field of the building.

A. The buildings and trial descriptions

We collected data in portions of three different buildings on the UIUC campus: the Coordinated Sciences Laboratory (CSL), Talbot Laboratory, and Loomis Laboratory. These buildings are depicted in Figure 5 and the portion of each building used in our dataset contain approximately 195, 295, and 470 square meters of area respectively. The number of training and testing trials for each building, platform, and test scenario are described in Table III. The table also reports the average density of magnetometer training measurements over the relevant area of the associated building in the dataset. Histograms of the time and distance traveled over all trials in the dataset are presented in Figure 6. In total, over 13 hours and 51 kilometers of data have been collected. On average, each test case was approximately 90 seconds and 101 meters long. The shortest test case by time was 21 seconds, and the shortest test case by distance traveled was 23 meters. In
Fig. 4. The pipeline for collecting data for the MagPIE dataset. We first map each test area with our Lenovo Phab 2 Pro. Then the magnetometers are calibrated. The test device is subsequently held stationary, so that an initial estimate of the device’s orientation can be computed. The data presented in the publicly available dataset is then collected. Throughout our tests, the magnetometer is calibrated again to ensure that biases were not inadvertently introduced.

(a) Coordinated Science Laboratory

(b) Talbot Laboratory

(c) Loomis Laboratory

Fig. 5. We collect data for mapping and localization in portions of three buildings: the first floor of the UIUC’s Coordinated Sciences Laboratory, the third floor of Talbot Laboratory, and the first floor of Loomis Laboratory.

addition, the longest test case by time was approximately 245 seconds, and the longest test case by distance was 297 meters.

B. File Descriptions

We structured our dataset such that each test or training trial contained five files: Initial_DCM#.txt, Output_accel#.txt, Output_gf#.txt, Output_gyro#.txt, Output_mag#.txt. The initial DCM files were the result of the coarse quasi-stationary initialization. The first column of the accelerometer, ground truth, gyroscope, and magnetometer files have been made to contain the time stamp of each measurement in seconds. The accelerometer, gyroscope, and magnetometer data have been made to list, in their subsequent columns, measurements along

<table>
<thead>
<tr>
<th>Building and Platform</th>
<th>Num. Train Cases</th>
<th>Num. Dead Load Cases</th>
<th>Num. Live Load Cases</th>
<th>Mag Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSL UGV</td>
<td>93</td>
<td>11</td>
<td>10</td>
<td>1615.1</td>
</tr>
<tr>
<td>CSL WLK</td>
<td>100</td>
<td>11</td>
<td>11</td>
<td>530.1</td>
</tr>
<tr>
<td>Talbot UGV</td>
<td>100</td>
<td>11</td>
<td>11</td>
<td>944.5</td>
</tr>
<tr>
<td>Talbot WLK</td>
<td>99</td>
<td>11</td>
<td>11</td>
<td>417.6</td>
</tr>
<tr>
<td>Loomis UGV</td>
<td>99</td>
<td>11</td>
<td>11</td>
<td>1907.2</td>
</tr>
<tr>
<td>Loomis WLK</td>
<td>100</td>
<td>12</td>
<td>11</td>
<td>4397.0</td>
</tr>
</tbody>
</table>

Fig. 6. Histograms of the distance traveled and time span of all trials.
Fig. 7. Directory tree of the MagPIE dataset. Directories for walking trials are labeled WLK. Dead load test cases are labeled Outlier Free and live load test cases are labeled With Outliers.

the body-frame’s $x$, $y$, and $z$ axes (in that order and in $m/s^2$, $rad/s$, and $\mu T$ respectively). The ground truth data has been made to list position and quaternion information respectively in subsequent columns. Our data was arranged in the file structure shown in Figure 7.

V. DISCUSSION

Our dataset has three characteristics that differentiate it from what has been available previously: data were collected from multiple buildings with sensors that were both handheld and mounted on a wheeled robot both with and without changes in the placement of objects that affect magnetometer measurements. In this section, we discuss the ways in which indoor positioning algorithms depend on these three characteristics to motivate their inclusion in our dataset.

A. Differences Between Buildings

Previous datasets—such as the UJIIndoorLoc-Mag dataset or the dataset from Barsocchi et al [22]—only use data from one building in their dataset. Using the method presented by Li et al [2], we use training data and Gaussian process regression (GPR) to generate two sets of maps: Talbot UGV maps and Loomis UGV maps. Then, we compute the percent of the global maximum likelihood estimate (MLE) and local maximum a posteriori (MAP) estimate above a set of thresholds over all dead load test cases associated with each scenario. Finally, we compute the same estimates over all live load test cases for the Talbot UGV segment of the dataset. As a result, we see that there is a significant difference between two of our buildings, which motivates the inclusion of multiple buildings in our dataset.

1) Mapping with Gaussian Process Regression: We generate a map with GPR for each element of the magnetic field vector in the Google Tango frame of reference using our training magnetic field measurements and the ground truth locations and orientations of those measurements. We construct this map with the help of the GPML toolbox [29]. We take as our GPR kernel the squared exponential function and we also assume a constant mean function. Our Loomis UGV dataset contains more than 444,000 data points for training our GPR algorithm, and our Talbot UGV dataset contains more than 475,000 data points for training. Due to memory limitations, we compute hyperparameters and a heat map using 75,000 randomly selected training points with 2,220 induced points and a variational free energy approximation [30]. The resulting GPR approximation computes a map of the expectation of the measurement from a calibrated magnetometer at any given point on the map as well as its variance. The map of the expectation of calibrated magnetometer measurements for the Loomis UGV case is shown in Figure 8. From the figure, the range of magnetic field measurements modeled by the map can be seen. Clearly, magnetic anomalies are present in the scene and can be quite large relative to Earth’s magnetic field.
2) Localization: Using the framework from Li et al [2], we take as our state vector, \( \theta \), the two dimensional map position and an azimuth angle. Given calibrated magnetometer measurements, we search over a grid of states in the map to find the maximum likelihood estimate using Bayes’ formula

\[
p(\theta|m_c) = \frac{p(m_c|\theta)p(\theta)}{p(m_c)}
\]

where the probability generated by the maps enable the computation of \( p(m_c|\theta) \) using the multivariate Gaussian model

\[
p(m_c|\theta) = \frac{\exp\left(\frac{1}{2} [m_c - \mu(\theta)]^T Q^{-1}(\theta) [m_c - \mu(\theta)]\right)}{(2\pi)^{\frac{3}{2}} |Q(\theta)|^{\frac{1}{2}}}
\]

where \( \mu(\theta) \) are the mean values from the maps produced through Gaussian process regression and \( Q(\theta) \) are the associated covariances from these maps. As with Li et al [2], \( p(m_c) \) is a normalizing factor since it is not a function of the state. We compute a global MLE every half meter along all test trajectories. We then, following the same approach as Li et al [2], include a prior, \( p(\theta) \), at the true state with a standard deviation of \( 3m \) to compute a local MAP estimate for all the test trajectories.

We compute, for the Loomis and Talbot UGV cases, the percent of tested points with a global MLE and local MAP estimate beyond a set of given thresholds in Table IV. While it is important to point out that such a map can be improved (with a resulting improvement in corresponding MLE and MAP estimation results), it is clear from the table that for our given mapping approach, there does exist a difference between the results associated with two of the buildings in our dataset. This result motivated collecting data from multiple buildings in the MapPIE dataset.

### C. Differences Between Live Load and Dead Load Cases

In previous literature on localization and mapping based on magnetic anomalies, it has been assumed that the magnetic fields in buildings are constant over time. It is stated by Wang et al [3] that “anomalies are reasonably stable in a long period of time.” In addition, Li et al [2] state that there exist “long temporal persistence in terms of direction and intensity” of a building’s magnetic field. While Frassl et al [4] acknowledges that distortions in a building’s magnetic field may occur due to moving objects like elevators or furniture, the field in a building still “appears to be stable over long periods of time.” It is stated by Akai and Ozaki [21] that the “stability of the disturbed magnetic field over long periods of time has already been demonstrated.” Finally, Robertson et al [32] also states that magnetic fields in buildings are “temporally stable.”

Rather than assuming that the magnetic field in a building is static, we include test cases in our dataset wherein we move objects along the test trajectory that are meant to change the building’s magnetic field. Thus, a positioning method may compare performance in the presence of larger innovations. Toward that end, we select eight potential candidate “live magnetic loads” that may alter the magnetic field in a building. We then place our UGV on rails and drive down a hallway twenty times: ten times without any object in the hallway.
and then a second ten with a given object in the hallway. The results are shown in Figure 10. Based on this, we use a computer cart, an industrial fan, a table and chairs, and a toolbox to change the magnetic field in approximately half of our UGV test trials. Performing a similar routine for our walking case, we choose to include a table and chairs and a computer cart in approximately half of our walking test trials.

Relative to the size of the map or the number of points sampled along the test trajectories in our localization algorithm, the number of distortions produced by the live magnetic loads are few. Thus, because test trajectories between live load cases and dead load cases are not kept constant, it is not surprising that live loads do not dominate the results produced in Table IV. However, as shown in Figure 10, this segment of the dataset may be important for comparing positioning algorithms that look to reject such outliers. Thus, while we present evidence that these live magnetic loads cause differences in the magnetic field of a building, more work must be done to understand their impact on localization algorithms. Due to these results, and the need for further study, we include live loads in our dataset.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation Grant No. 14-46765 and 14-27111. The authors would like to thank Bo Liu and the UIUC Intelligent Robotics Laboratory for help and support with the UGV used in this work.

REFERENCES

Fig. 10. We consider a UGV proceeding down a hallway 20 times. Ten times without a potential “live magnetic load” present (red tube of sample paths) and ten times with a set of potential “live magnetic loads” in the hallway (blue tube of sample paths). The pictures and plots above display these selected candidate live magnetic loads and the corresponding norm of the magnetic field as a function of the distance down the hallway for the 20 trials. Several of these candidates clearly alter the magnetic field in the hallway.