MagLoc-AR: Magnetic-based Localization for Visual-free Augmented Reality in Large-scale Indoor Environments

Haomin Liu*, Hua Xue*, Linsheng Zhao, Danpeng Chen, Zhen Peng, and Guofeng Zhang†, *Member*, *IEEE*

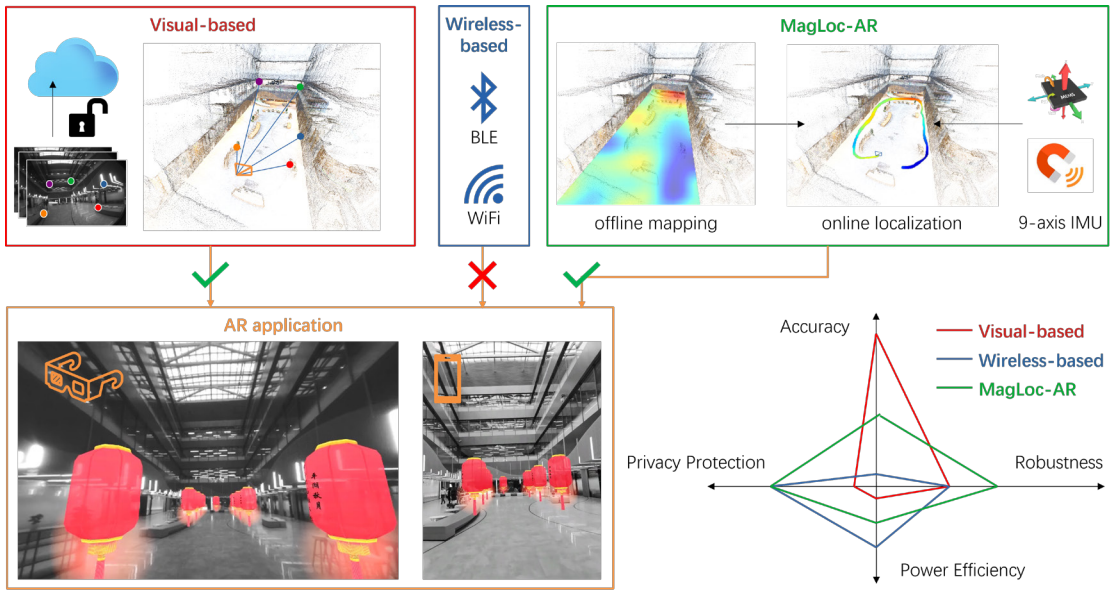

Fig. 1: The illustration of different types of localization methods. Visual-based methods are sufficiently accurate for AR, but they pose privacy risks, and suffer from robustness issues and high power consumption. Wireless-based methods are visual-free, but not accurate enough for AR. The proposed MagLoc-AR is visual-free for privacy protection, meets the accuracy requirement of AR navigation, and has advantages over visual-based methods in terms of robustness and power efficiency.

Abstract—Accurate localization of a display device is essential for AR in large-scale environments. Visual-based localization is the most commonly used solution, but poses privacy risks, suffers from robustness issues and consumes high power. Wireless signal-based localization is a potential visual-free solution, but its accuracy is not enough for AR. In this paper, we present MagLoc-AR, a novel visual-free localization solution that achieves sufficient accuracy for some AR applications (e.g. AR navigation) in large-scale indoor environments. We exploit the location-dependent magnetic field interference that is ubiquitous indoors as a localization signal. Our method requires only a consumer-grade 9-axis IMU, with the gyroscope and acceleration measurements used to recover the motion trajectory, and the magnetic measurements used to register the trajectory to the global map. To meet the accuracy requirement of AR, we propose a mapping method to reconstruct a globally consistent magnetic field of the environment, and a localization method fusing the biased magnetic measurements with the network-predicted motion to improve localization accuracy. In addition, we provide the first dataset for both visual-based and geomagnetic-based localization in large-scale indoor environments. Evaluations on the dataset demonstrate that our proposed method is sufficiently accurate for AR navigation and has advantages over the visual-based methods in terms of power consumption and robustness. Project page: https://github.com/zju3dv/MagLoc-AR/

Index Terms—Indoor localization, Augmented reality, Inertial navigation system

1 Introduction

Augmented Reality (AR) allows users to interact with virtual content and information across vast physical spaces, creating immersive and engaging experiences. Nowadays, the application of AR has been extended to large-scale environments, creating a broader range of user experiences. For example, AR navigation improves the navigation experience compared to traditional floorplan-based navigation by providing virtual content to guide and inform users more intuitively. The key is to accurately localize the display device in the environment, particularly in the indoor environment where GPS is unavailable.

The most commonly used solution is visual-based localization. For example, there are currently commercialized services such as Google ARCore Geospatial, Apple ARKit Location Anchors, and Microsoft Azure Spatial Anchors. In the offline phase, a visual map of the environment is reconstructed using structure-from-motion (SfM) [37]. In the online phase, the user captures images of their surroundings, which are then matched against the visual map to recover the 6DoF (6 Degree

- *Haomin Liu is with Peking University and Sensetime Research. E-mail: liuhaomin@sensetime.com.*
- *Hua Xue, Linsheng Zhao, and Zhen Peng are with Sensetime Research. E-mail: {xuehua, zhaolinsheng, pengzhen1}@sensetime.com.*
- *Danpeng Chen is with the State Key Lab of CAD&CG, Zhejiang University, Sensetime Research and Tetras.AI. E-mail: chendanpeng@tetras.ai.*
- *Guofeng Zhang is with the State Key Lab of CAD&CG, Zhejiang University. E-mail: zhangguofeng@zju.edu.cn.*

* *Equal contribution.*
† *Corresponding author.*

Manuscript received 25 March 2023; revised 17 June 2023; accepted 7 July 2023. Date of publication 2 October 2023; date of current version 31 October 2023. This article has supplementary downloadable material available at https://doi.org/10.1109/TVCG.2023.3321088, provided by the authors.
Digital Object Identifier no. 10.1109/TVCG.2023.3321088

of Freedom) pose of the device in the environment [36]. The 6DoF pose is tracked over time using visual SLAM [5], and periodically re-located to correct for accumulating errors. For large-scale environments, performing visual localization on resource-constrained mobile devices is impractical. Users have to upload images to a cloud server, revealing potentially confidential user information [19, 43]. Furthermore, the camera-related image processing and visual SLAM consume significant power on the mobile device, making it unsuitable for long-term AR experiences. Additionally, current visual-based methods still face robustness challenges such as repetitive structure, poor texture, occlusions, large viewpoint changes, low light, etc [24]. Wireless signal-based localization is a potential visual-free solution. Traditional methods require deployment of signal transmitters in the environment, such as WiFi Access Point (AP), or Bluetooth Beacon. The additional equipment, as well as the cost of deployment and maintenance, impedes the large-scale application of these methods. On the other hand, the location-dependent variation of the ambient magnetic field resulting from magnetic materials in building structures can also serve as a useful wireless signal for indoor localization, eliminating the need for signal transmitters. However, the accuracy of all existing wireless signal-based methods can only support a rough 3DoF localization on the floorplan, which is insufficient for AR.

In this paper, we present MagLoc-AR, a visual-free 5DoF localization solution for AR in large single-level indoor environments, where the user is moving at a constant height. As illustrated in Fig. 1, our method requires only a consumer-grade 9-axis IMU, which has advantages over existing visual-based methods in terms of privacy protection, power consumption, and robustness under challenging situations for visual-based methods. To meet the accuracy requirement of AR, we propose a mapping method to reconstruct a globally consistent magnetic field of the environment, and a localization method fusing the biased magnetic measurements with the network-predicted motion, which significantly improves localization accuracy over existing wireless signal-based methods. The main contributions are:

- We present a novel visual-free localization solution that achieves sufficient accuracy for AR navigation in large single-level indoor environments as far as we know.
- We propose a rectification method and a relative observation model to handle the non-negligible magnetic bias of the consumer-grade IMU for mapping and localization respectively.
- We propose a fusion framework leveraging a deep network to predict human motion during AR experience for accurate localization.
- We provide the first dataset for both visual-based and geomagnetic-based localization in large-scale indoor environments, which is also used to evaluate and verify the effectiveness of the proposed method.

We organize this paper as follows : Sec. 2 briefly reviews related works. Sec. 3 gives an overview of the proposed MagLoc-AR. The details of offline mapping and online localization are elaborated in Sec. 4 and Sec. 5 respectively. Finally, we evaluate our proposed method in Sec. 6 and conclude this work in Sec. 7.

2 Related Work

In this section, we review existing visual-based and wireless signal-based localization methods respectively.

2.1 Visual-based Localization

Visual SLAM is one of the most commonly used visual-based localization techniques for AR. It estimates the 6DoF pose of moving camera with respect to the local map which is simultaneously reconstructed. Early works only used visual measurements [22, 30, 32, 45], which suffered from robustness issues in poor texture or fast motion scenarios that frequently happen in practice. Recent works combine the complementary inertial measurements from IMU to achieve great improvement for these challenging cases [5, 26, 34]. However, if there are no reliable visual measurements for a long time, the motion tracking based on IMU alone will still suffer from serious drift problem [8].

Visual localization by matching a query image to a pre-built map of the environment is an effective way to correct errors [4]. More importantly, for many AR applications, localization is required to be with respect to the pre-built map where the virtual contents are created. Traditionally, it is done by extracting hand-crafted features [31, 35] from query images, which are matched against features in the pre-built map. Recent works resort to learning-based features [10, 13] to improve the robustness against motion blurs, illumination changes and viewpoint variations. However, inherent limitations of visual-based methods still remain, such as day-night and across seasons changes, repetitive structure, poor texture, occlusions, low light, etc. These challenges are still far from being solved [24, 36]. Another drawback of the visual-based methods is the risk of privacy exposure. For large-scale environments, it is impractical to perform visual localization on resource-constrained mobile devices. Users have to upload images to a cloud server, which may disclose confidential information related to the captured environment. This is the case even when only extracted features are uploaded, as they can be used to reconstruct the query images [12, 33]. Recent works propose to lift 2D/3D feature points to random lines in order to conceal the geometry of the query image [19, 43]. However, they do not address the privacy issue when a series of query images are sent with high frame rate, or when the sensitive content is static [19].

2.2 Wireless signal-based Localization

Wireless signal-based techniques for indoor localization have been extensively explored in the past few decades. WiFi is one of the most commonly used signals. Earlier works estimate locations from the received signal strength indicator (RSSI) of WiFi APs. They either directly match the RSSI to a pre-constructed database [2, 50], or model the RSSI as a function of the distance [9]. Localization error of those methods can be up to several meters and even worse when the distribution of WiFi APs is sparse. Recent studies [23, 48] have shown the possibility of achieving sub-meter accuracy by estimating time-of-flight (ToF) and angle-of-arrival (AoA) with detailed physical layer information. However, they either work for only a few specific devices, or require complex hardware customization. Bluetooth Low Energy (BLE) is another technology that is widely considered for indoor localization. Methods using RSSI are proposed [7, 11, 14], which also suffer from large error. Although the research community is making progress continuously, due to the dependency on environmental infrastructures, WiFi and BLE-based technologies are far from large-scale commercial deployment.

Indoor magnetic field shows stable distinctions at different locations and has been proved to have the capability of localizing commercial mobile devices [27]. Some existing methods [21, 38, 39] avoid orientation estimation by only using the sequence of magnetic field strength as location features, which do not fully utilize the distinction of magnetic field at different locations and thus result in low accuracy and weak generalization ability. Although the work [41] takes 3-axis magnetic field values as different features, it uses a simple constant-velocity dynamic model, which introduces large error. As for the mapping strategies for magnetic-based methods, all the above works use pedestrian dead reckoning (PDR) based method to register magnetic measurements, which introduces large errors in mapping. The work [3] uses a robotic platform with an elaborately calibrated magnetic sensor to reconstruct the magnetic field, which has high accuracy but introduced high cost, making it difficult to scale.

There are also methods using other wireless signals like ultra-sound [28], visible light [49], ultra-wide band (UWB) [51], mmWave [42]. They all require specialized hardware and thus have difficulty to be widely deployed.

3 Method Overview

In this section, we give an overview of the proposed MagLoc-AR and explain how the components work together. First, we describe the notation used in this paper. The reference system for the vector coordinates

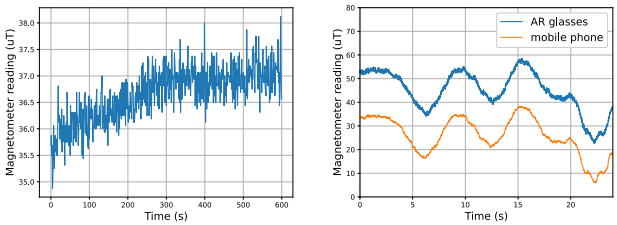

Fig. 2: Magnetic measurements by the consumer-grade MEMS magnetometer: (a) from an mobile phone placed on the table for 10 minutes in a static environment; (b) from a mobile phone and a pair of AR glasses moving along the same trajectory.

is shown by the left superscript. ${}^B_A T$ represents the transformation from coordinate A to B:

$$ {}^B X = {}^B_A T \circ {}^A X. \tag{1}$$

${}^G_I T_t$ denotes the transformation from IMU coordinate to the global map coordinate at timestamp t. It is also called the IMU pose, and is comprised of a rotation matrix R_t and a translation vector p_t. The transformation can be rewritten as

$$ {}^G X = R_t \, {}^I X + p_t. \tag{2}$$

For IMU pose, we omit the prescripts for R_t and p_t for simplicity. The measurement from the 9-axis IMU at timestamp t is denoted as $({}^I\omega_t, {}^I a_t, {}^I m_t)$, where ${}^I\omega_t$, ${}^I a_t$, and ${}^I m_t$ represent gyroscope, acceleration and magnetic measurement respectively, expressed in the IMU local coordinate. For the consumer-grade MEMS IMU, these measurements contain time-varying biases. The bias in ${}^I\omega_t$ and ${}^I a_t$ has been well investigated and modeled in the field of visual-inertial SLAM. However, the bias in ${}^I m_t$ has not been studied in the field of magnetic-based localization. As shown in Fig. 2(a), during a period of 10 minutes of standing in a static environment, we observe that the magnetometer bias changes at a constant rate for the first 5 minutes, and remains almost unchanged for the next 5 minutes. This indicates that the random walk model, commonly used for gyroscope and acceleration biases, is no longer suitable for magnetometer for the entire period of time. In addition, we also find that the magnitude of magnetometer bias varies significantly among different devices, as shown in Fig. 2(b). In the framework of MagLoc-AR, we propose innovative methods to address these challenges.

The framework is illustrated in Fig. 3. It consists of two phases, namely *offline mapping* and *online localization*. In the phase of *offline mapping*, we reconstruct the magnetic field of the environment. To do this, we walk around the site holding a panoramic camera Insta360 Pro 2[1] that can capture 360° images of the environment and record the 9-axis IMU measurements as well. We first reconstruct the global map using visual-inertial SfM and register each magnetic measurement to it, as detailed in Sec. 4.1. Then we propose a method to rectify the bias in the registered magnetic measurements, which is detailed in Sec. 4.2. Finally, we reconstruct the magnetic field map by generalizing the discrete magnetic measurements into a continuous and smooth magnetic field, as detailed in Sec. 4.3. For the purpose of efficiency, the magnetic field map is represented as a 2D grid, which prevents the subsequent localization from estimating changes in height. As a result, MagLoc-AR can only support 5DoF localization, assuming that the user is moving within a single level at a constant height.

In the phase of *online localization*, MagLoc-AR fuses IMU measurements ${}^I\omega_{1...t}$, ${}^I a_{1...t}$ and ${}^I m_{1...t}$ from the AR device to estimate the IMU pose ${}^G_I Tt$. Using ${}^I\omega_{1...t}$ and ${}^I a_{1...t}$, it can recover the motion trajectory, and the trajectory can be registered to the global map by matching ${}^I m_{1...t}$ along the trajectory against the magnetic field map. Since the magnetic field map contains significant ambiguities among different locations, we propose to use a particle filter for state estimation. The particle filter-based framework takes two steps at each iteration – the

[1] https://www.insta360.com/product/insta360-pro2

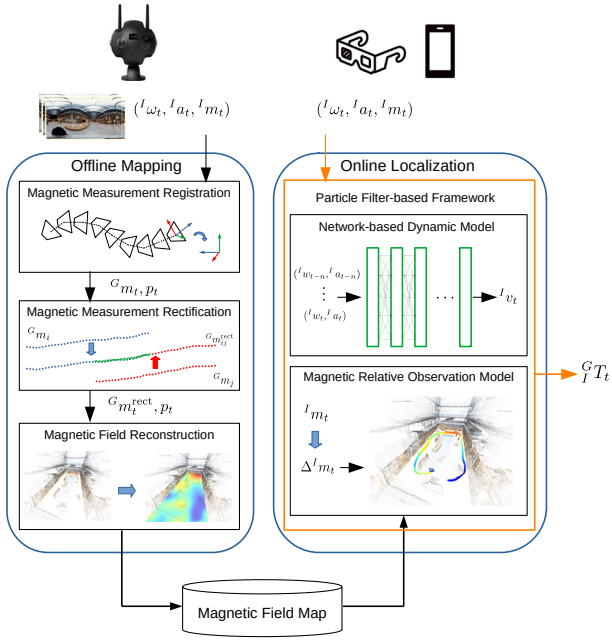

Fig. 3: The framework of MagLoc-AR.

prediction step with a dynamic model and the update step with an observation model. In the prediction step, we propose a network-based dynamic model which takes a sliding window of ${}^I\omega_{t-n...t}$ and ${}^I a_{t-n...t}$ to predict the instant velocity ${}^I v_t$, as detailed in Sec. 5.1. In the update step, we propose a magnetic relative observation model to handle the bias effect in the magnetic measurement, as detailed in Sec. 5.2.

4 Offline Mapping

The map is fundamental in any global localization system and its quality directly impacts the performance of localization. In the case of mapping for magnetic-based localization using consumer-grade devices, practical problems are not well-studied and solved. In this section, we identify those problems and provide our solutions to them as building blocks in the offline mapping phase of MagLoc-AR.

4.1 Magnetic Measurement Registration

In order to register each magnetic measurement to the global map, traditional methods use pedestrian dead reckoning (PDR) with pre-measured landmark locations as constraints to estimate the IMU poses [21], which is less accurate and difficult to scale.

To break these limitations, we propose to use SfM [37] to recover the IMU poses. Specifically, 2D features are extracted for each image and matched throughout the image sequence to constrain the camera poses. We select the deep learning-based feature SuperPoint [10] for its superior robustness over traditional handcrafted features.

Taking these feature correspondences, SfM recovers the camera pose ${}^G_C T_i$ for each image frame i by jointly recovering the 3D coordinate ${}^G X_j$ for each feature point j. The recovered camera poses and 3D points need to satisfy the projection constraint, which means that each 3D point ${}^G X_j$, when projected into image i by camera pose ${}^G_C T_i$, must coincide with its 2D observation x_{ij} in the image. The process can be formulated as

$$ \underset{{}^G_C T_i, {}^G X_j}{\arg\min} \sum_{i,j} ||\pi({}^G_C T_i^{-1} \circ {}^G X_j) - x_{ij}||^2, \tag{3}$$

where $\pi({}^C X)$ is the projection function for a point expressed in the camera coordinates to the image plane by the camera intrinsics, which is pre-calibrated by Kalibr [18].

The image-only SfM cannot recover the true scale. Scaling all ${}^G_C T_i$ and ${}^G X_j$ will result in the same re-reprojection error. It also suffers from robustness issues in texture-less environments. We address these

problems by incorporating gyroscope and acceleration measurements to Eq. (3), resulting in the formulation of visual-inertial SfM:

$$\arg\min_{\mathscr{I}_i,{}^G X_j} \sum_{i,j} ||\pi({}^C_I T \circ {}^G_I T_i^{-1} \circ {}^G X_j) - x_{ij}||^2 + \sum_i h(\mathscr{I}_i, \mathscr{I}_{i+1}, \mathbb{Z}_{i,i+1}), \quad (4)$$

where $\mathscr{I}_i = \{{}^G_I T_i, {}^G v_i, {}^I b_i\}$ are the IMU motion parameter at image frame i, comprised of the IMU pose ${}^G_I T_i$, the velocity ${}^G v_i$, and the bias ${}^I b_i$ of gyroscope and acceleration. ${}^C_I T$ is the fixed transformation from IMU to camera coordinates, which is pre-calibrated by Kalibr [18]. $\mathbb{Z}_{i,i+1}$ is the set of IMU measurements between consecutive frames $(i, i+1)$, and $h(\cdot)$ is the cost function measuring the difference between IMU states $(\mathscr{I}_i, \mathscr{I}_{i+1})$ and IMU measurements $\mathbb{Z}_{i,i+1}$, calculated by IMU pre-integration [16]. Compared to Eq. (3), the additional cost function $h(\cdot)$ in Eq. (4) eliminates the ambiguity of scale by leveraging the acceleration measurements in the true scale, and constrains the poses between consecutive frames even in the featureless environments. With the IMU pose ${}^G_I T_i$ recovered, the local magnetic measurement ${}^I m_i$ is converted to the global coordinate by ${}^G m_i = R_i {}^I m_i$ with the associated location p_i.

4.2 Magnetic Measurement Rectification

Due to hardware imperfections, MEMS magnetometers have non-negligible time-varying bias as described in Sec. 3. SfM typically requires loops on the acquisition path to eliminate the drift, and the time-varying bias will result in inconsistent magnetic measurements at loop closure locations across time. Such inconsistency of measurements can pose unexpected uncertainty in the reconstruction of the magnetic field and finally cause large errors in online localization. To address this issue, we propose a method called magnetic measurement rectification.

At any time instance, the magnetic measurement in the IMU coordinate can be formulated as

$${}^I m_t = {}^I m_t^{\text{true}} + b_t, \quad (5)$$

where ${}^I m_t^{\text{true}}$ is the actual magnetic vector in the IMU coordinate and b_t is the time-varying bias. We omit the Gaussian white noise for simplicity. Given IMU poses at two time steps t and t' in a short period of time, where $t > t'$, the relative change of magnetic measurements under the global coordinate can be represented as

$$\Delta {}^G m_{t,t'} = R_t {}^I m_t^{\text{true}} - R_{t'} {}^I m_{t'}^{\text{true}} + R_t b_t - R_{t'} b_{t'}. \quad (6)$$

During a short time period, we assume that the bias follows the model of Brownian motion with white noise [16]

$$\dot{b}_t = \eta \sim \mathcal{N}(0, \sigma_b^2). \quad (7)$$

Integrating over the time interval $[t', t]$ obtains

$$b_t = b_{t'} + \eta_d. \quad (8)$$

If rotations R_t and $R_{t'}$ are approximately the same, we can substitute Eq. (8) into Eq. (6) to get

$$\Delta {}^G m_{t,t'} \approx R_t {}^I m_t^{\text{true}} - R_{t'} {}^I m_{t'}^{\text{true}} + R_t \eta_d, \quad (9)$$

in which the original terms related to b_t and $b_{t'}$ are replaced with a Gaussian noise, thus the relative measurement is approximately independent of the time-varying bias. By using the bias-independent relative measurements to reconstruct the magnetic field, we effectively counter the effect of time-varying bias.

Inspired by the observations above, MagLoc-AR makes the registered magnetic measurements consistent by minimizing the error of relative measurements in the global coordinate. First, we divide the mapping space using 2D grids with the shape of $0.5m \times 0.5m$. Then, within each grid, we choose the registered magnetic measurement which is closest to the center to represent the measurement of that grid. After that, along the acquisition trajectory, we select grid magnetic measurements with the difference of rotations between $(R_t, R_{t'})$ less than 5 degrees to form relative measurements in the global coordinate. Finally, we estimate the consistent magnetic measurements in the global coordinate of different grids by minimizing error of all the relative measurements as

$$\arg\min_{{}^G m_i} \sum_{i,j} ||\Delta {}^G m_{i,j} - ({}^G m_i - {}^G m_j)||^2, \quad (10)$$

where ${}^G m_i$ is the estimated magnetic vector at the i-th grid and $\Delta {}^G m_{i,j}$ is the relative measurement between the i-th grid and the j-th grid.

It is a standard linear least-square form which can be easily solved. However, since we only use relative measurements, the problem is ill-posed. To make the solution unique, we assign absolute measurements to eliminate the freedom. Specifically, we convert the 2D grid to a graph with nodes representing selected grid cells and two nodes are connected by an edge if there is a relative measurement between them. We first find all connected components of the graph. Then, we solve Eq. (10) for each connected component separately. For each connected component, we manually assign one absolute measurement to make the problem well-posed. We denote obtained consistent magnetic measurements as ${}^G m_i^{\text{rect}}$, which will be used for magnetic field reconstruction.

4.3 Magnetic Field Reconstruction

Given the magnetic measurements $(p_i, {}^G m_i^{\text{rect}})$ along the acquisition path, we reconstruct the entire magnetic field by Gaussian Process Regression (GPR) [47] that is widely used for magnetic field modeling [1, 40, 41]. For the purpose of completeness, we briefly introduce GPR here.

Assuming that the x/y/z-components of the 3D magnetic field are independent and can be reconstructed separately, we define the measurement set as $\mathscr{D} = \{(x_i, y_i) | i = 1 \dots n\}$, where x_i is the 2D location of p_i on the grid, and y_i is the corresponding 1D component of ${}^G m_i^{\text{rect}}$. GPR assumes the magnetic field obeys the Gaussian process function $f(x)$ at each location x:

$$\begin{aligned} f(x) &\sim \mathcal{GP}(\mu_0, k(x, x')) \\ y_i &= f(x_i) + \varepsilon \end{aligned}, \quad (11)$$

where μ_0 is the mean value of y_i in $\mathscr{D}$ and $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$ with hyper-parameter σ_n. $k(x, x')$ is the covariance between two locations (x, x'):

$$k(x, x') = \delta_f^2 \exp(-\frac{\|x - x'\|^2}{2l^2}), \quad (12)$$

where δ_f and l are hyper-parameters. The magnetic field reconstruction is formulated as predicting the posterior distribution $\mathcal{N}(\mu_*, \sigma_*^2)$ of magnetic intensity at each grid cell x_*:

$$\begin{aligned} \mu_* &= \mu_0 + \mathbf{k}_*^T (\mathbf{K} + \sigma_n^2 I)^{-1} \bar{\mathbf{y}} \\ \sigma_*^2 &= k(x_*, x_*) - \mathbf{k}_*^T (\mathbf{K} + \sigma_n^2 I)^{-1} \mathbf{k}_* \end{aligned}, \quad (13)$$

where $\mathbf{k}_*$ is the n-dimensional vector with i-th element $k(x_i, x_*)$. $\mathbf{K}$ is the $n \times n$ matrix with $\mathbf{K}_{ij} = k(x_i, x_j)$. $\bar{\mathbf{y}}$ is the n-dimensional vector with $\bar{\mathbf{y}}_i = y_i - \mu_0$. The hyper-parameters σ_n, δ_f and l are trained by maximizing the marginal likelihood from $\mathscr{D}$. Details are referred to [47].

5 Online Localization

In this section, we give a detailed description of techniques used in the online localization phase. The online localization works under the traditional particle filter fusion framework with a network-based dynamic model and a relative observation model.

The particle filter is a well-known and powerful iterative state estimation framework which does not rely on the certainty of initial states. The online Bayesian state estimation problem of MagLoc-AR under the

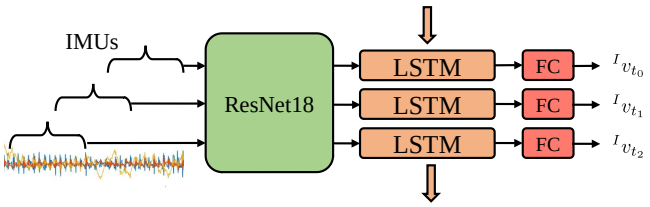

Fig. 4: The structure of the network.

assumption of Markov dynamic model can be formulated in an iterative manner as

$$p({}^G_I T_t|S_t) \propto p({}^I m_t|{}^G_I T_t)p({}^G_I T_t|{}^G_I T_{t-1},{}^I\omega_t,{}^I a_t)p({}^G_I T_{t-1}|S_{t-1}), \tag{14}$$

where S_t represents measurements $({}^I\omega_{1...t},{}^I a_{1...t},{}^I m_{1...t})$ from the 9-axis IMU up to the timestamp t. The particle filter framework uses weighted samples to represent the posterior distribution and takes two steps at each iteration. First, it predicts ${}^G_I T_t$ by drawing samples from the distribution $p({}^G_I T_t|{}^G_I T_{t-1},{}^I\omega_t,{}^I a_t)$ which is defined by the dynamic model. Then it updates weights of those samples using $p({}^I m_t|{}^G_I T_t)$ which is defined by the observation model. The sampled expectation of the posterior distribution is finally used as the estimate of the IMU pose.

5.1 Network-based Dynamic Model

The dynamic model describes the way state transits between continuous timestamps. Traditional dynamic models for magnetic-based localization systems either assume low degree-of-freedom [38,39,41] or require motion sensors with high accuracy [17]. In the case of MagLoc-AR, the dynamic model is required to be accurate under high degree-of-freedom motions in AR applications. To meet the requirement, we resort to network-based method. For MagLoc-AR, the dynamic model can be formulated as

$$\begin{bmatrix} R_t & p_t \end{bmatrix} = \begin{bmatrix} R_{t-1} & p_{t-1} \end{bmatrix} \begin{bmatrix} \Delta^I R_t & {}^I v_t \Delta t \\ 0 & 1 \end{bmatrix}, \tag{15}$$

where Δt is the time passed from timestamp $t-1$ to t and $\Delta^I R_t$ is the corresponding rotation in the IMU local coordinate. We obtain $\Delta^I R_t$ through a traditional extended Kalman filter (EKF) with IMU measurements.

The estimation of ${}^I v_t$ is not trivial. Recently, many deep network-based inertial navigation methods [6, 8, 20] have achieved good positioning accuracy and robustness. They directly regress velocity from IMU data without the need for complex initialization processes. We adopt the overlapping IMU window method proposed in [8] to regress a high frequency velocity. As shown in Fig. 4, we input IMU data from windows with some overlap into the ResNet18 network to obtain motion hidden variables in time sequence. Since motion is regular and continuous in general, we use the LSTM network to fuse the time sequence motion hidden variables and finally use a fully connected layer to regress the corresponding velocity. The data preprocessing and training process is similar to the method [8]. We train network models for different devices separately. For the mobile phone model, we use data provided in [8]. For the AR glasses model, we collect data by ourselves in different environments with typical motion patterns.

In the prediction step of the particle filter, we first draw samples from the distributions of $\Delta^I R_t$ and ${}^I v_t$, then predict the state through Eq. (15).

5.2 Magnetic Relative Observation Model

The observation model in our case describes the distribution of magnetic measurements given the IMU poses. For each predicted state ${}^G_I T_t$, we first match the translation p_t to the closest grid in the magnetic field map to obtain the corresponding mean μ_{p_t} and covariance Σ_{p_t} of the magnetic measurement ${}^G m_t$ in the global coordinate, then use the rotation matrix R_t to further get the distribution of ${}^I m_t$. However, due to the time-varying bias as we mentioned in Sec. 3, the magnetic measurement is not always consistent with the distribution.

To conquer the problem introduced by bias, similar to Sec. 4.2, we take the relative change between measurements as our observation. Denote the states of one particle at two time steps t and t' $(t > t')$ as ${}^G_I T_t$ and ${}^G_I T_{t'}$ respectively, and the corresponding magnetic measurements as ${}^I m_t$ and ${}^I m_{t'}$. The relative observation of MagLoc-AR at the time step t with respect to t' is defined as

$$\Delta^I m_{t,t'} = {}^I m_t - {}^I m_{t'}. \tag{16}$$

Combining Eq. (16), Eq. (5) and Eq. (8), we can see that when t and t' are close, the slowly-changing bias approximately disappears in the relative observation. We assume that magnetic measurements are independent at different times and locations. The distribution of $\Delta^I m_{t,t'}$ is then a linear combination of two independent Gaussian distributions as

$$\Delta^I m_t \sim \mathcal{N}(R_t^{-1}\mu_{p_t} - R_{t'}^{-1}\mu_{p_{t'}},\ R_t^{-1}\Sigma_{p_t}R_t + R_{t'}^{-1}\Sigma_{p_{t'}}R_{t'}). \tag{17}$$

In order to make the relative observation more distinct across different trajectories, MagLoc-AR maintains a sliding window for each particle. This sliding window keeps memory of all most recent magnetic measurements and corresponding states for the particle in a pre-defined time duration. Multiple relative observations between different time steps within the window are combined to calculate the probability of magnetic measurements. Specifically, suppose the window includes states and measurements at sequential time steps $(t_1, t_2, t_3, \ldots, t_n)$, MagLoc-AR combines relative observations between t_n and all the other previous time steps to update the weight of the particle from t_{n-1} to t_n as

$$w_{t_n} = w_{t_{n-1}} \prod_{i=1}^{n-1} p(\Delta^I m_{t_n,t_i}), \tag{18}$$

where w_t indicates the weight at time step t.

6 Results

In this section, we conduct performance evaluation of MagLoc-AR through various experiments under several typical indoor environments. As far as we know, there is no available dataset containing visual-inertial data (images and IMU), magnetic measurements and Bluetooth signals (RSSI) in large-scale indoor environments for the evaluation. Therefore, we collect the first dataset of this kind, and use it to compare MagLoc-AR with both visual-based and wireless-based baselines in terms of accuracy, robustness and efficiency. We also conduct an ablation study to verify the effectiveness of the proposed components.

6.1 Dataset

We select four typical indoor environments to analyze the localization performance of different methods. The first two are visual-friendly, and the last two are visual-challenging.

- **Medium office (MO)**: An office room of 500m^2, with sufficient light and rich textures.
- **Large office (LO)**: A very large office of 4000m^2, with sufficient light and rich textures.
- **Spacious hall (SH)**: A very spacious hall inside the office building, with an area of 900m^2 and a height of 15m. Walking in the spacious hall, most contents in the field of view are more than 10 meters away unless looking at the ground.
- **Parking lot (PL)**: A parking lot of 12000m^2, with dim light and repetitive scenes in different areas, which is very challenging for visual-based methods.

We walked around these four environments holding a panoramic camera Insta360 Pro 2 that can capture 360° images of the environment and record the 9-axis IMU measurements as well. The reconstructed maps of the environments as well as the representative 360° images are

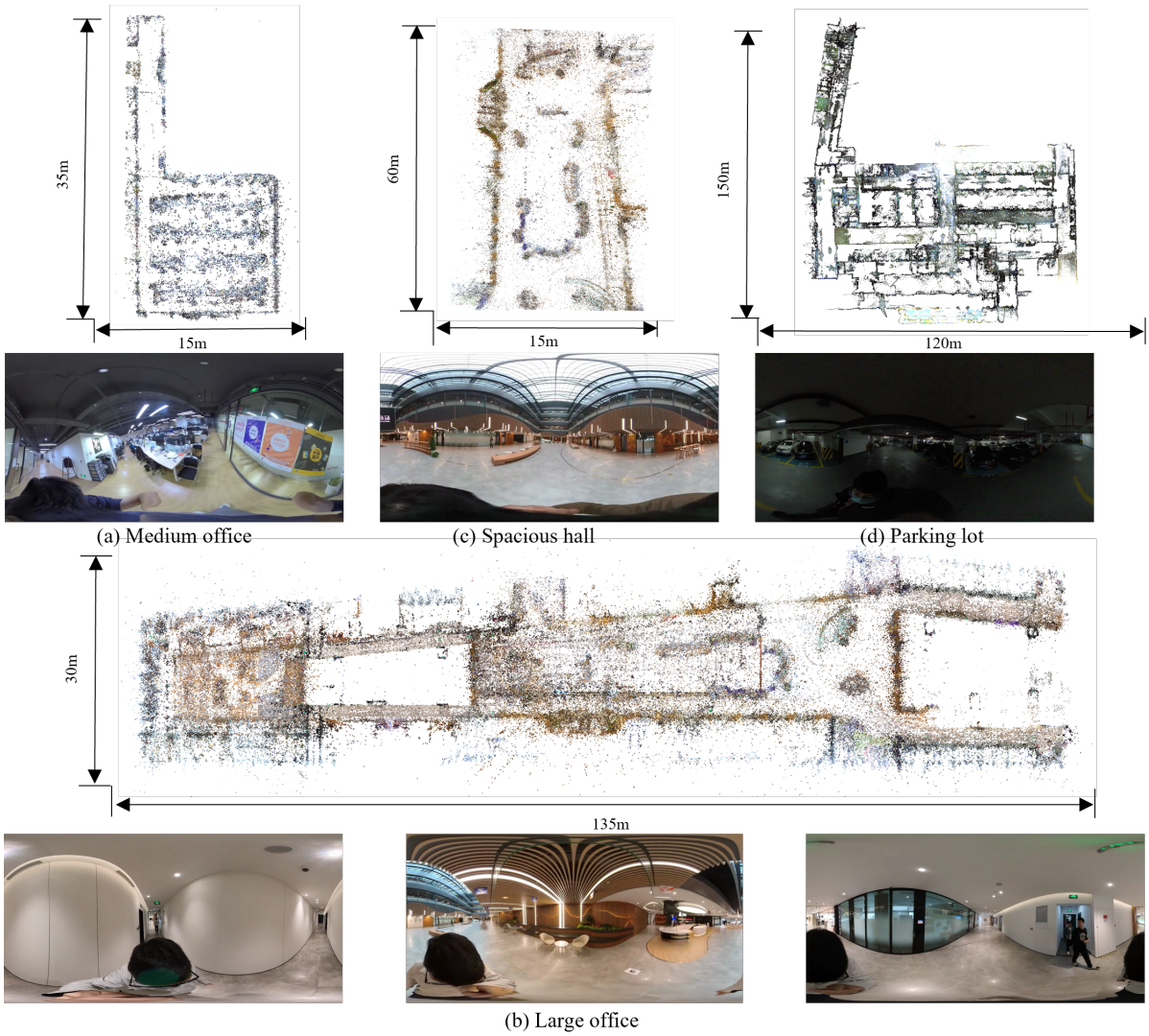

Fig. 5: Four indoor environments in the dataset.

shown in Fig. 5, and the full 360° videos are shown in the supplementary material. To analyze the performance of the BLE-based method, we installed Bluetooth beacons on the ceiling beams in the first three environments, but not in the last one because it was a public area where it was inconvenient to deploy beacons. We placed one beacon every 2 meters for the medium-scale office (MO), and every 10 meters for the large-scale office buildings (LO and SH).

We use AR glasses of Shadow Creator Action One Pro, and a mobile phone Huawei Mate30Pro respectively to record three sequences in each of the four environments, resulting in 24 sequences in total. Details of the three sequences in each environment are described in Table 1. Each sequence contains visual-inertial data, magnetic measurements, and Bluetooth signals for comparison among the visual-based and wireless signal-based methods. The visual data captured by the AR glasses and the mobile phone are stereo images and monocular images respectively. The video of the 24 sequences are shown in the supplementary material.

To perform a quantitative analysis of different localization methods, it is necessary to recover the ground truth pose for each sequence. Following previous works [24, 29, 44] that built benchmarks for visual-based localization in large-scale environments, we use SfM to recover the groundtruth poses. Specifically, we perform the SfM described in Sec. 4.1 on all the visual-inertial data from the panoramic camera, the AR glasses, and the mobile phone in the same environment. The recovered poses serve as groundtruth.

Table 1: Various types of motions in the dataset

Seq.	Motion
MO1	Along a large loop, looking ahead, slow
MO2	Along a large loop, looking around, slow
MO3	Along a large loop, looking around, fast
LO1∼3	In 3 different areas, looking ahead, slow
SH1	Along a loop, looking at the center, slow
SH2	Along a loop, looking at the outer wall, slow
SH3	Along a line and return, looking at the distance, slow
PL1∼3	In 3 different areas, looking around, fast

6.2 Performance Comparison

In this section, we first analyze the performance of the proposed MagLoc-AR with the visual-based and wireless signal-based baselines quantitatively and qualitatively in terms of accuracy and robustness, and then compare the computational efficiency.

6.2.1 Baselines

We select four representative localization methods as baselines. The first two are visual-based methods, and the next two are wireless signal-based methods.

- **VisLoc**: For each frame, SuperPoint features [10] are extracted, and matched against the global visual map to obtain a set of 2D-3D feature correspondences, from which the pose is recovered by

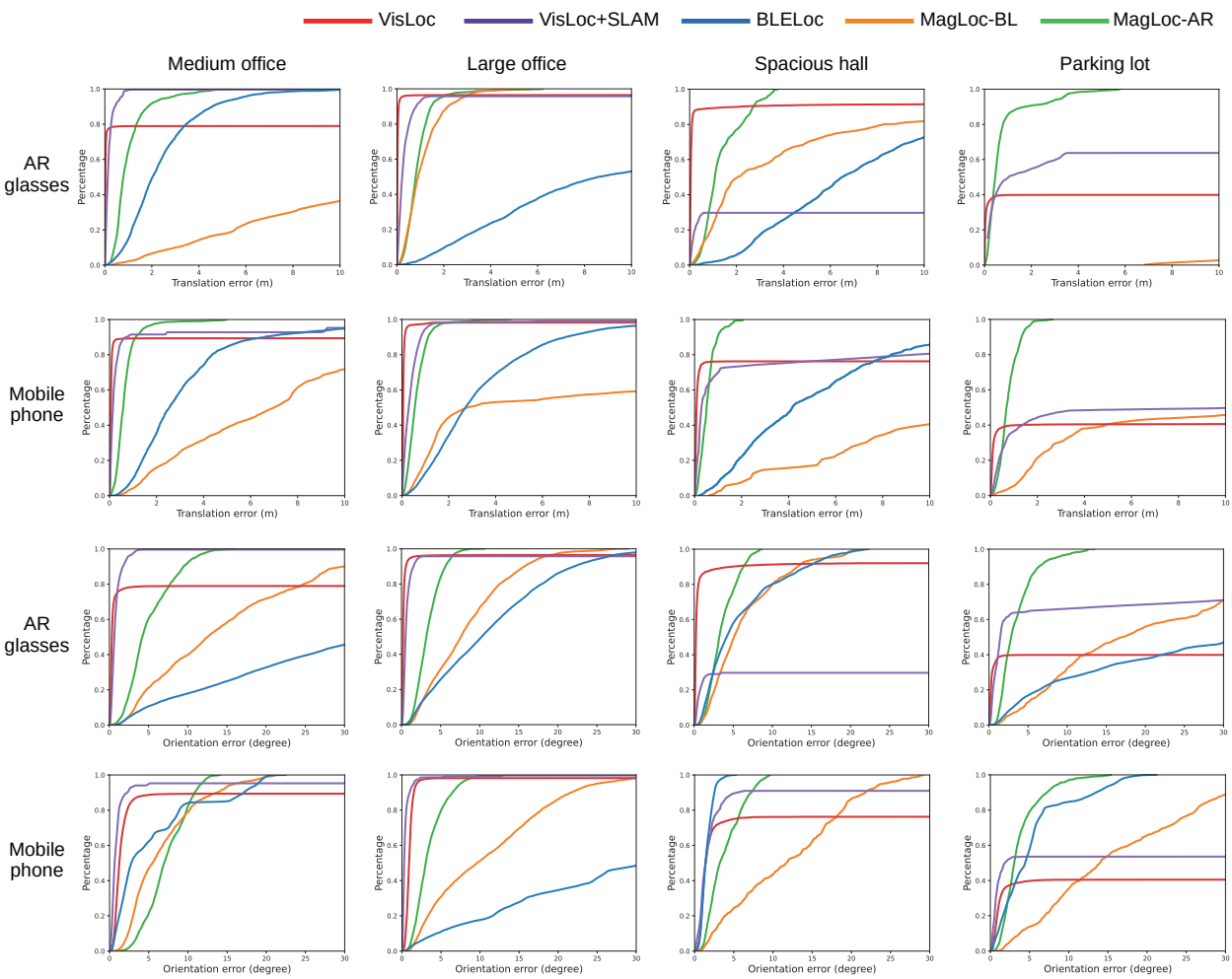

Fig. 6: CDF of the errors of different localization methods for quantitative comparison among visual-based and wireless-based methods. The color codes of different methods are shown in the top right corner. The four columns correspond to the four environments: Medium office, Large office, Spacious hall and Parking lot. The first and third rows correspond to the AR glasses data, and the second and fourth rows correspond to the mobile phone.

EPnP [25] in the framework of RANSAC [15].

- **VisLoc+SLAM**: We select the visual-inertial SLAM system ORB-SLAM3 [5] for its well-known accuracy. VisLoc is called every 10 seconds to register the trajectory to the global map.
- **BLELoc**: We collect RSSIs of BLE beacons in the environment within a time window of 2 seconds and use the weighted k nearest neighbor (WKNN) [21] to estimate the locations with respect to a pre-build BLE database. It can only estimate the position but not the orientation. The orientation is obtained from the device's built-in API, which is usually estimated by [46].
- **MagLoc-BL**: We implement the method proposed in [41], which first reconstructs the map by directly performing GPR on the magnetic measurements registered in a traditional way, and then performs localization using a particle filter with the constant-velocity dynamic model and the absolute magnetic observations.

6.2.2 Quantitative Comparison

Evaluation criterion. Typically, there is a trade-off between the accuracy and robustness of a localization method. Different classes of methods may have different types of trade-offs between accuracy and robustness. While the same class of methods can be evaluated using individual metrics such as Root Mean Square Error (RMSE) and localization success rate to evaluate accuracy and robustness respectively, a more comprehensive evaluation criterion is needed to compare different classes of methods. We propose to use the Cumulative Distribution Function (CDF) as a comprehensive evaluation method to reflect both accuracy and robustness. The point (x, y) on the CDF curve means that the probability of having an error less than x is y. The faster y rises in the region of small x, the higher the accuracy; the closer y is to 1 in the region of large x, the better the robustness; y at the largest x indicates the localization success rate. If the curve of method A is higher than that of method B in the entire range of x, it indicates that A has better both accuracy and robustness; if A and B intersect, it indicates that one method has higher accuracy while the other one has better robustness. The CDF curves are shown in Fig. 6. Below we will use "environment-device" abbreviations to refer to a subfigure, such as "MO-ARG" for Medium office with AR glasses, "LO-MP" for Large office with Mobile phone. In addition, ORB-SLAM3 has stricter initialization requirements than other SLAM methods, as it mandates stable tracking for 15 seconds to be considered a successful initialization [5]. Our aim is to provide a general analysis that is less influenced by a specific algorithmic strategy. So we use data after 15 seconds from the beginning to calculate the CDF curve of errors.

Comparison with visual-based methods. In the visual-friendly environments MO and LO, the visual-based VisLoc and VisLoc+SLAM achieve moderately better accuracy than MagLoc-AR, while their robustness is comparable to MagLoc-AR. By combining SLAM, Vis-

Loc+SLAM effectively improves the robustness of VisLoc especially in MO-ARG where VisLoc has the lowest success rate. MagLoc-AR doesn't have a clear advantage in these cases, except for MO-MP where the robustness is slightly better than the visual-based methods. By contrast, in the visual-challenging environments SH and PL, MagLoc-AR shows a clear advantage in terms of robustness. In SH, the environment is too spacious that most contents in the field of view are far away, which makes VisLoc+SLAM not only fail to improve VisLoc's success rate, but also cause a drastic decrease in both accuracy and robustness. Although SH-MP does not show an obvious drop on the curve, there is still significant drift in some frames. About 15% of frames have errors larger than 5 meters by VisLoc+SLAM. Moreover in PL, the environment is dim, which poses a challenge for both VisLoc and SLAM. Even worse, the structure is similar in different regions, resulting in only 40% success rate of VisLoc. After combining SLAM, the success rate slightly increases, but still remains much lower than that in the visual-friendly environments, with about 64% and 46% success rates in PL-ARG and PL-MP respectively. One of the main reasons causing the low success rate is that ORB-SLAM3 did not initialize successfully in this dimly lit environment even after the first 15 seconds that we cut off. In contrast, MagLoc-AR maintains consistent accuracy and robustness, and outperforms the visual-based methods significantly in robustness, while being slightly inferior in accuracy. Only in SH-ARG, due to the spaciousness of the environment, the magnetic field has less variation at different locations, which leads to a slight decrease in accuracy than other environments, but still has a clear advantage over the visual-based methods.

Comparison with wireless-based methods. Unlike the visual-based methods, the wireless signal-based methods do not show a clear trade-off between accuracy and robustness. Their CDF curves barely intersect, with the higher curve indicating better performance in both aspects. For translation, MagLoc-AR outperforms BLELoc and MagLoc-BL significantly. For orientation, MagLoc-AR also surpasses BLELoc in most cases due to its ability to eliminate location-dependent magnetic interference. The only exception is SH-MP, where the magnetic interference is small in the spacious environment. In this case, BLELoc, which uses magnetic measurement as the Earth's magnetic field, has better orientation. Compared to MagLoc-BL based on the same principle as ours, the proposed MagLoc-AR is significantly better because it can handle the magnetic bias of the consumer-grade IMU and adopt a dynamic model that aligns with human motion better.

6.2.3 Qualitative Comparison

We qualitatively compare the localization results of different methods through AR effects. We manually place virtual objects along the groundtruth trajectory on the global map. Then we use the localization results of different methods to render these virtual objects from corresponding viewpoints, which are overlaid onto the background image. If the localization results are accurate, the virtual objects will align with the image. Otherwise, there will be jitter or drift. The results are included in the supplementary materials. The qualitative experiment results and the quantitative results are consistent. Compared to visual-based methods, MagLoc-AR exhibits moderately lower accuracy but significantly better robustness. Compared to wireless-based methods, MagLoc-AR demonstrates significantly higher accuracy and robustness, making it suitable for some AR applications (e.g. AR navigation). However, we also observed that the visual-based methods, despite their higher accuracy, still exhibit slight jitter in the AR effect generated by VisLoc, and moderate drift in SLAM causing noticeable jumps of virtual objects generated by VisLoc+SLAM, even in the visual-friendly environments. In the visual-challenging environments, the problem of jitter and drift becomes more severe. In contrast, MagLoc-AR does not encounter these issues, demonstrating its robustness advantage over visual-based methods.

6.2.4 Efficiency Comparison

We use Huawei Mate30Pro to compare the computational efficiency among different methods. We first compare power consumption. The results are listed in Table 2. VisLoc is not listed on the first row as an individual method. The reason is it is not practical to call VisLoc every frame or compute VisLoc on a mobile device. In practice, VisLoc is typically used in conjunction with SLAM by performing low-frequency VisLoc on the cloud while tracking by SLAM on the mobile device. In this regard, VisLoc only consumes the power of the mobile phone for uploading images and receiving location results every 10 seconds. We list the result of VisLoc as a component of VisLoc+SLAM and find that it does not consume much power. Regarding SLAM, since the purely software-implemented ORB-SLAM3 [5] does not represent the optimal power efficiency of SLAM, we select AREngine which has undergone sufficient power optimization as a representative of practical applications. We also evaluate the power consumption of the camera itself without running any other algorithms. We find that it accounts for half of the power consumption of VisLoc+SLAM, which is a fundamental problem of all vision-based methods.

Table 2: Power consumption in W

VisLoc+SLAM				BLELoc	MagLoc-BL	MagLoc-AR
Camera	VisLoc	SLAM	Toltal			
2.237	0.036	2.439	4.712	0.025	1.232	1.256

We evaluate the power consumption of wireless signal-based methods in the Medium office with the map fully loaded into the memory while running. For BLELoc, we perform localization using received beacon signals within a time window of 2 seconds. For magnetic-based methods, the particle filter is efficient to generate estimates in real-time. In comparison, all wireless signal-based methods have much lower power consumption, with BLELoc consuming almost no power. Although the magnetic-based methods consume much more power than BLELoc, they consume only one-fourth of the power consumption of VisLoc+SLAM.

Latency is also an important aspect for AR. Long latency will result in poor AR experiences. Since directly measuring the motion-to-photon latency is not an easy task, we compare the pose update rate instead to reflect the latency. Usually the faster the update rate, the smaller the latency is. The update rates are 30/0.5/500 Hz for AREngine/BLELoc/MagLoc-AR respectively.

6.3 Ablation Study

We conduct an ablation study by replacing the proposed components with traditional methods, resulting in five methods. The first two are for the offline mapping, and the last three are for the online localization.

- **Mapping w/o SfM**: The proposed SfM introduced in Sec. 4.1 is replaced with a traditional method [21] to register magnetic measurement to the global map. Waypoints along the acquisition path are manually set and measured to provide positions of the magnetic measurements. Pedestrian Dead Reckoning (PDR) [21] is used for interpolation between consecutive waypoints. Orientations are obtained from the device's built-in API. Due to the high cost of this solution, we only experiment in the Medium office.
- **Mapping w/o rectification**: The magnetic measurement rectification proposed in Sec. 4.2 is disabled.
- **Localization with PDR**: The network-based velocity estimation proposed in Sec. 5.1 is replaced with a traditional method [39], which uses PDR [3] to estimate the velocity.
- **Localization with RoNIN**: The network-based velocity estimation proposed in Sec. 5.1 is replaced with RoNIN [20], which also performs relative motion estimation using data-driven neural networks.
- **Localization w/o relative observation**: The magnetic relative observation proposed in Sec. 5.1 is replaced with the traditional absolute observation [41].

The results of comparison are shown in Fig. 7. In summary, our proposed MagLoc-AR comprehensively outperforms all the compared methods. Specifically, in the phase of offline mapping, **Mapping w/o SfM** relies on the device's built-in API orientation, which suffers from magnetic interference and leads to lower accuracy than the proposed

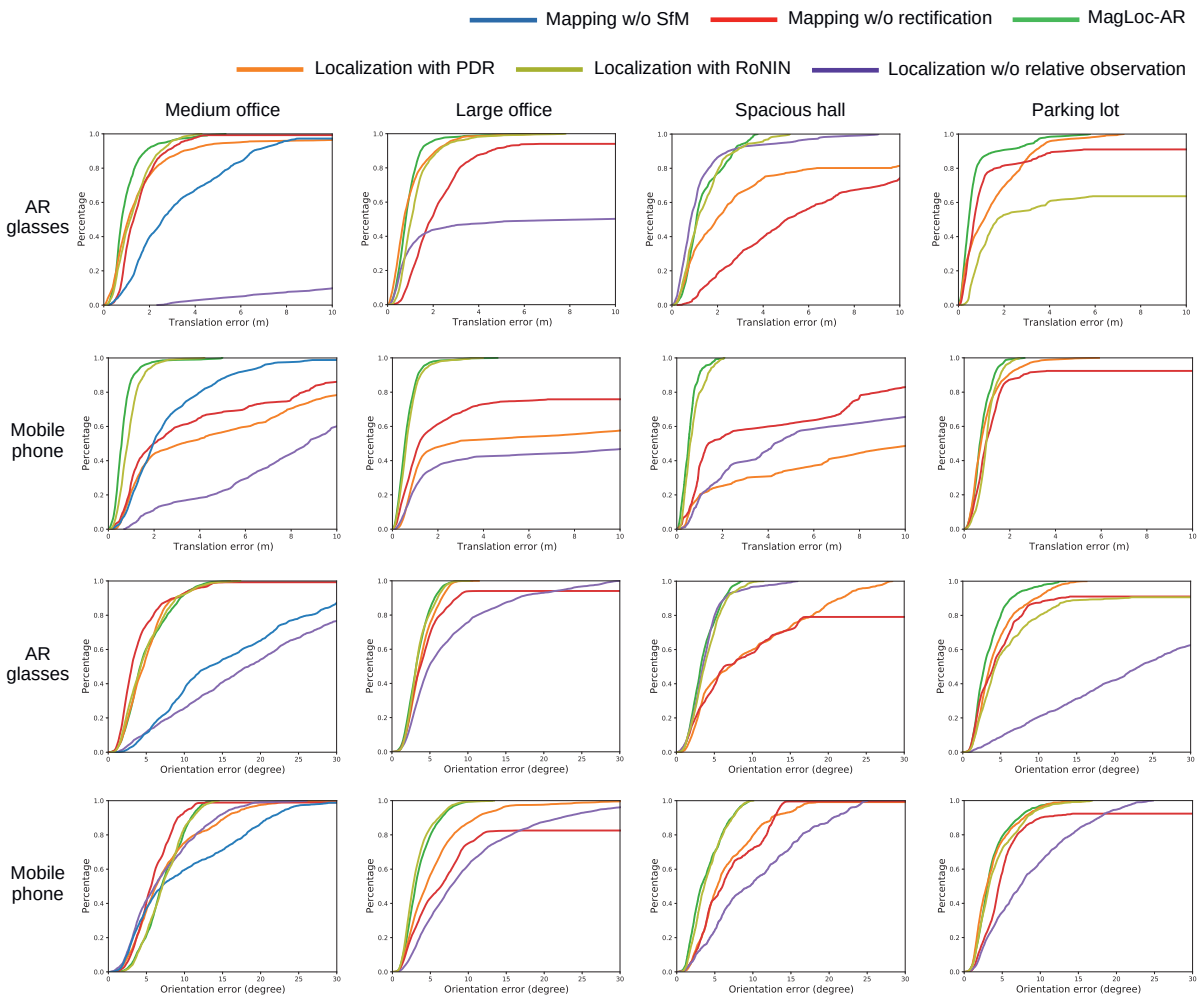

Fig. 7: CDF of the errors of different localization methods for ablation study. The color codes of different methods are shown in the top right corner. The four columns correspond to the four environments: Medium office, Large office, Spacious hall and Parking lot. The first and third rows correspond to the AR glasses data, and the second and fourth rows correspond to the mobile phone.

SfM. Compared with **Mapping w/o rectification**, MagLoc-AR also shows superior performance, demonstrating the effectiveness of handling the time-varying bias in magnetic measurements. In the phase of online localization, we find that the dynamic motion model based on PDR aligns better with the motion of head-mounted AR glasses than the motion of handheld mobile phone which is much more flexible. Therefore, MagLoc-AR's advantage over **Localization with PDR** is more significant on data of mobile phone than AR glasses. **Localization with RoNIN** has comparable performance to MagLoc-AR in most cases, demonstrating the ability of our method to support other network-based dynamic models. The only exception is PL-ARG where MagLoc-AR outperforms **Localization with RoNIN** significantly. The reason is RoNIN relies on the orientation obtained from the device's built-in API, which is erroneous in this case. Compared with **Localization w/o relative observation**, MagLoc-AR's advantage is the most remarkable among all methods, once again demonstrating the necessity of handling magnetic bias in consumer-grade IMU.

7 Conclusion

In this work, we present MagLoc-AR, a novel 5DoF visual-free localization solution that achieves sufficient accuracy for AR navigation in large single-level indoor environments. In the phase of offline mapping, we use 360° images and 9-axis IMU measurements from a panoramic camera to reconstruct the magnetic field of the environment. In the stage of online localization, we only use the 9-axis IMU measurements from AR devices to localize the device in the magnetic map. To meet the requirement of AR, we propose several methods to handle the non-negligible magnetic bias of the consumer-grade IMU in both phases of mapping and localization, as well as a method to better predict human motion during AR experience in the phase of localization. We also provide the first dataset for comparison between the visual-based and wireless-based methods. The evaluations demonstrate the visual-free MagLoc-AR not only meets the accuracy requirement of AR, but also has advantages over the visual-based methods in terms of robustness and power efficiency.

One of the main limitations of MagLoc-AR is its sensitivity to the variation of magnetic field. In the spacious environments where magnetic field exhibits less variation across different locations, the localization performance may degrade. Another limitation is that the magnetic field map is currently represented as a 2D grid for efficiency purposes, which prevents MagLoc-AR from localizing the height of the mobile device. This may be an important consideration for certain AR applications. We plan to overcome these limitations in future work. Additionally, we currently only use the magnetometer for localization. However, we plan to explore its potential for assisting in visual-based mapping and improving the robustness of SfM in poor-textured and repetitive environments.

Acknowledgments

This work was partially supported by China Postdoctoral Science Foundation, and NSF of China (No. 61932003).

REFERENCES

[1] N. Akai and K. Ozaki. Gaussian processes for magnetic map-based localization in large-scale indoor environments. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4459–4464. IEEE, 2015. 4

[2] P. Bahl and V. Padmanabhan. RADAR: an in-building RF-based user location and tracking system. In *The 19th International Conference on Computer Communications*, vol. 2, pp. 775–784 vol.2, March 2000. 2

[3] S. Beauregard and H. Haas. Pedestrian dead reckoning: A basis for personal positioning. In *The 3rd Workshop on Positioning, Navigation and Communication*, pp. 27–35, 2006. 2, 8

[4] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics*, 32(6):1309–1332, 2016. 2

[5] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós. ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM. *IEEE Transactions on Robotics*, 37(6):1874–1890, 2021. 2, 7, 8

[6] C. Chen, X. Lu, A. Markham, and N. Trigoni. IONet: Learning to cure the curse of drift in inertial odometry. In *AAAI Conference on Artificial Intelligence*, vol. 32, 2018. 5

[7] D. Chen, K. G. Shin, Y. Jiang, and K.-H. Kim. Locating and tracking BLE beacons with smartphones. In *The 13th International Conference on Emerging Networking EXperiments and Technologies*, p. 263–275. Association for Computing Machinery, New York, NY, USA, 2017. 2

[8] D. Chen, N. Wang, R. Xu, W. Xie, H. Bao, and G. Zhang. RNIN-VIO: Robust neural inertial navigation aided visual-inertial odometry in challenging scenes. In *IEEE International Symposium on Mixed and Augmented Reality*, pp. 275–283. IEEE, 2021. 2, 5

[9] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan. Indoor localization without the pain. In *The Sixteenth Annual International Conference on Mobile Computing and Networking*, p. 173–184. Association for Computing Machinery, New York, NY, USA, 2010. 2

[10] D. DeTone, T. Malisiewicz, and A. Rabinovich. SuperPoint: Self-supervised interest point detection and description. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 224–236, 2018. 2, 3, 6

[11] P. Dickinson, G. Cielniak, O. Szymanezyk, and M. Mannion. Indoor positioning of shoppers using a network of Bluetooth Low Energy beacons. *International Conference on Indoor Positioning and Indoor Navigation*, pp. 1–8, 2016. 2

[12] A. Dosovitskiy and T. Brox. Inverting visual representations with convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4829–4837, 2016. 2

[13] M. Dusmanu, I. Rocco, T. Pajdla, M. Pollefeys, J. Sivic, A. Torii, and T. Sattler. D2-Net: A trainable CNN for joint description and detection of local features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8092–8101, 2019. 2

[14] R. Faragher and R. Harle. Location fingerprinting with Bluetooth Low Energy beacons. *IEEE Journal on Selected Areas in Communications*, 33(11):2418–2428, nov 2015. 2

[15] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 7

[16] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza. IMU preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation. Georgia Institute of Technology, 2015. 4

[17] M. Frassl, M. Angermann, M. Lichtenstern, P. Robertson, B. J. Julian, and M. Doniec. Magnetic maps of indoor environments for precise localization of legged and non-legged locomotion. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 913–920. IEEE, 2013. 5

[18] P. Furgale, J. Rehder, and R. Siegwart. Unified temporal and spatial calibration for multi-sensor systems. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1280–1286. IEEE, 2013. 3, 4

[19] M. Geppert, V. Larsson, P. Speciale, J. L. Schönberger, and M. Pollefeys. Privacy preserving Structure-from-Motion. In *The 16th European Conference on Computer Vision*, pp. 333–350. Springer, 2020. 2

[20] S. Herath, H. Yan, and Y. Furukawa. RoNIN: Robust neural inertial navigation in the wild: Benchmark, evaluations, & new methods. In *IEEE International Conference on Robotics and Automation*, pp. 3146–3152. IEEE, 2020. 5, 8

[21] Y. Hu, F. Qian, Z. Yin, Z. Li, Z. Ji, Y. Han, Q. Xu, and W. Jiang. Experience: Practical indoor localization for malls. In *The 28th Annual International Conference on Mobile Computing And Networking*, p. 82–93. Association for Computing Machinery, New York, NY, USA, 2022. 2, 3, 7, 8

[22] G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In *The 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 225–234. IEEE, 2007. 2

[23] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti. SpotFi: Decimeter level localization using WiFi. In *ACM Conference on Special Interest Group on Data Communication*. London, UK, 2015. 2

[24] D. Lee, S. Ryu, S. Yeon, Y. Lee, D. Kim, C. Han, Y. Cabon, P. Weinzaepfel, N. Guérin, G. Csurka, et al. Large-scale localization datasets in crowded indoor spaces. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3227–3236, 2021. 2, 6

[25] V. Lepetit, F. Moreno-Noguer, and P. Fua. EPnP: An accurate O(n) solution to the PnP problem. *International Journal of Computer Vision*, 81(2):155, 2009. 7

[26] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale. Keyframe-based visual–inertial odometry using nonlinear optimization. *The International Journal of Robotics Research*, 34(3):314–334, 2015. 2

[27] B. Li, T. Gallagher, A. G. Dempster, and C. Rizos. How feasible is the use of magnetic field alone for indoor positioning? *International Conference on Indoor Positioning and Indoor Navigation*, pp. 1–9, 2012. 2

[28] Q. Lin, Z. An, and L. Yang. Rebooting ultrasonic positioning systems for ultrasound-incapable smart devices. In *The 25th Annual International Conference on Mobile Computing and Networking*. Association for Computing Machinery, New York, NY, USA, 2019. 2

[29] H. Liu, M. Jiang, Z. Zhang, X. Huang, L. Zhao, M. Hang, Y. Feng, H. Bao, and G. Zhang. LSFB: A low-cost and scalable framework for building large-scale localization benchmark. In *IEEE International Symposium on Mixed and Augmented Reality Adjunct*, pp. 219–224. IEEE, 2020. 6

[30] H. Liu, G. Zhang, and H. Bao. Robust keyframe-based monocular SLAM for augmented reality. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1–10. IEEE, 2016. 2

[31] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 2

[32] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, 31(5):1147–1163, 2015. 2

[33] F. Pittaluga, S. J. Koppal, S. B. Kang, and S. N. Sinha. Revealing scenes by inverting structure from motion reconstructions. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 145–154, 2019. 2

[34] T. Qin, S. Cao, J. Pan, P. Li, and S. Shen. VINS-Fusion: An optimization-based multi-sensor state estimator, 2019. 2

[35] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. ORB: An efficient alternative to SIFT or SURF. In *International Conference on Computer Vision*, pp. 2564–2571. IEEE, 2011. 2

[36] T. Sattler, W. Maddern, C. Toft, A. Torii, L. Hammarstrand, E. Stenborg, D. Safari, M. Okutomi, M. Pollefeys, J. Sivic, et al. Benchmarking 6DoF outdoor visual localization in changing conditions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8601–8610, 2018. 2

[37] J. L. Schonberger and J.-M. Frahm. Structure-from-Motion revisited. In *IEEE Conference on Computer Vision and Pattern Recognition*, June 2016. 1, 3

[38] Y. Shu, C. Bo, G. Shen, C. Zhao, L. Li, and F. Zhao. Magicol: Indoor localization using pervasive magnetic field and opportunistic WiFi sensing. *IEEE Journal on Selected Areas in Communications*, 33(7):1443–1457, jul 2015. 2, 5

[39] Y. Shu, K. G. Shin, T. He, and J. Chen. Last-mile navigation using smartphones. In *The 21st Annual International Conference on Mobile Computing and Networking*, p. 512–524. Association for Computing Machinery, New York, NY, USA, 2015. 2, 5, 8

[40] A. Solin, M. Kok, N. Wahlström, T. B. Schön, and S. Särkkä. Modeling and interpolation of the ambient magnetic field by Gaussian processes. *IEEE Transactions on robotics*, 34(4):1112–1127, 2018. 4

[41] A. Solin, S. Särkkä, J. Kannala, and E. Rahtu. Terrain navigation in the magnetic landscape: Particle filtering for indoor positioning. In *European Navigation Conference*, pp. 1–9. IEEE, 2016. 2, 4, 5, 7, 8

[42] E. Soltanaghaei, A. Prabhakara, A. Balanuta, M. Anderson, J. M. Rabaey, S. Kumar, and A. Rowe. Millimetro: mmWave retro-reflective tags for accurate, long range localization. In *The 27th Annual International Con-*

ference on Mobile Computing and Networking, p. 69–82. Association for Computing Machinery, New York, NY, USA, 2021. 2

[43] P. Speciale, J. L. Schonberger, S. B. Kang, S. N. Sinha, and M. Pollefeys. Privacy preserving image-based localization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5493–5503, 2019. 2

[44] E. Spera, A. Furnari, S. Battiato, and G. M. Farinella. EgoCart: a benchmark dataset for large-scale indoor image-based localization in retail stores. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(4):1253–1267, 2019. 6

[45] W. Tan, H. Liu, Z. Dong, G. Zhang, and H. Bao. Robust monocular SLAM in dynamic environments. In *IEEE International Symposium on Mixed and Augmented Reality*, pp. 209–218. IEEE, 2013. 2

[46] N. Trawny and S. I. Roumeliotis. Indirect Kalman filter for 3D attitude estimation. *University of Minnesota, Department of Computer Science & Engineering, Technical Report*, 2, 2005. 7

[47] C. K. Williams and C. E. Rasmussen. *Gaussian processes for machine learning*, vol. 2. MIT press Cambridge, MA, 2006. 4

[48] Y. Xie, J. Xiong, M. Li, and K. Jamieson. mD-track: Leveraging multi-dimensionality for passive indoor Wi-Fi tracking. In *The 25th Annual International Conference on Mobile Computing and Networking*. Association for Computing Machinery, New York, NY, USA, 2019. 2

[49] Z. Yang, Z. Wang, J. Zhang, C. Huang, and Q. Zhang. Wearables can afford: Light-weight indoor positioning with visible light. In *The 13th Annual International Conference on Mobile Systems, Applications, and Services*, p. 317–330. Association for Computing Machinery, New York, NY, USA, 2015. 2

[50] M. Youssef and A. Agrawala. The Horus WLAN location determination system. In *The 3rd International Conference on Mobile Systems, Applications, and Services*, pp. 205–218. ACM, 2005. 2

[51] M. Zhao, T. Chang, A. Arun, R. Ayyalasomayajula, C. Zhang, and D. Bharadia. ULoc: Low-power, scalable and cm-accurate UWB-Tag localization and tracking for indoor applications. *The ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(3), 2021. 2