Tightly-coupled Magneto-Visual-Inertial Fusion for Long Term Localization in Indoor Environment

Jade Coulin[1,2], Richard Guillemard[1], Vincent Gay-Bellile[1], Cyril Joly[2], and Arnaud de La Fortelle[2]

***Abstract*—We propose in this paper a tightly-coupled fusion of visual, inertial and magnetic data for long-term localization in indoor environment. Unlike state-of-the-art Visual-Inertial SLAM (VISLAM) solutions that reuse visual map to prevent drift, we present in this paper an extension of the Multi-State Constraint Kalman Filter (MSCKF) that takes advantage of a magnetic map. It makes our solution more robust to variations of the environment appearance. The experimental results demonstrate that the localization accuracy of the proposed approach is almost the same over time periods longer than a year.**

***Index Terms*—Localization, Sensor Fusion, Visual-Inertial SLAM, Indoor magnetic field, MSCKF.**

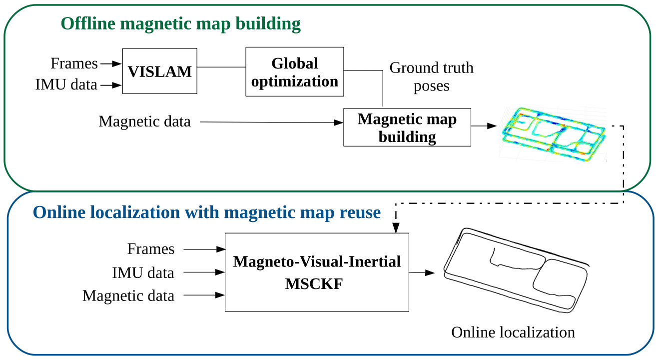

Fig. 1. Overview of the proposed magneto-visual-inertial MSCKF framework. Offline a magnetic map is built from magnetic data and ground truth positions. Online, the magnetic map is reused in the MSCKF for drift-free localization in the long term.

I. INTRODUCTION

VISUAL-INERTIAL SLAM algorithms (VISLAM) provide accurate and robust localization that made them state-of-the-art for indoor localization [1], [2]. The addition of the IMU resolves the lack of robustness of visual-only SLAM algorithms. It also helps to reduce drift over time but does not completely eliminate it. To solve this issue a prebuilt visual map of the environment is usually reused [1]. Its 3D points are matched with the key points of the images through a viewpoint recognition algorithm [3]. This results in a drift-free localization algorithm as long as a sufficient number of 3D points are successfully recognized. However, the performance of viewpoint recognition algorithms degrades over time due to illumination variations and changes in the environment appearance (moving furniture, decoration change, etc).

Recently, different localization methods taking advantage of magnetic field stability over time have been proposed. In [4], magnetic data are loosely coupled with a VISLAM algorithm through a particle filter. Their solution results in a localization with limited drift but with an increase of local inaccuracy (jitter). The magnetic data are also underused since they are not involved in the estimation of the VISLAM state.

We propose in this paper a tightly coupled fusion of magnetic, visual and inertial data that results in a locally accurate and drift-free localization in indoor environment. An extension of the Multi-State Constraint Kalman Filter (MSCKF) used initially for visual-inertial fusion is presented. Thus, the contributions of this paper are:

- A tightly-coupled magneto-visual-inertial MSCKF reusing magnetic map (see Figure 1), whose localization accuracy is not affected by appearance changes of the environment.
- Online estimation of the magnetometer bias in the MSCKF, avoiding recalibration of the magnetometer each time it is used.

The proposed algorithm is evaluated on a data set we created, including acquisitions of visual, inertial and magnetic data covering a time span of more than one year. Results demonstrate that the proposed solution keeps almost the same localization accuracy over time contrary to state-of-the-art VISLAM method [1] reusing visual map. Furthermore, the addition of magnetic data may reduce computation time since fewer images are required to achieve an accurate localization.

The paper is structured as follows. After a review of the existing literature in Section II, Section III provides some preliminaries on magnetic map building. Then, the proposed magneto-visual-inertial MSCKF algorithm is described in Section IV. In Section V, its performances in terms of localization accuracy in the long term are evaluated and compared to those of a state-of-the-art method that reuses a visual map.

II. RELATED WORK

A. Long-term localization

VISLAM algorithms, which reuse visual map for drift-free localization such as [1], become less accurate over time due

---

Manuscript received: September, 9th, 2021; Revised November, 3rd, 2021; Accepted December, 6th, 2021.

This paper was recommended for publication by Editor Sven Behnke upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by CEA List.

[1] Jade Coulin, Richard Guillemard and Vincent Gay-Bellile are with Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France `jade.coulin@cea.fr`

[2] Cyril Joly and Arnaud de La Fortelle are with MINES ParisTech, PSL University, Center for robotics, 60 Bd St Michel, 75006 Paris, France `cyril.joly@mines-paristech.fr`

Digital Object Identifier (DOI): see top of this page.

to variations of the visual appearance of the environment. Some works try to improve their robustness to illumination conditions by learning illumination invariant descriptors [5] or recognition modules for different illumination conditions [6]. However, these approaches lead to an increase in computation time without dealing with all kinds of appearance variations.

There exist other alternatives to limit VISLAM drift. In [7], a dense LIDAR map is aligned with an online dense stereo reconstruction. This approach has only been evaluated in outdoor scenarios, as its performance may be degraded by the lack of texture and geometric information of indoor environments. In [8] Wi-Fi sensing signal and in [9] ultra-wideband range measurements are tightly coupled with the visual and inertial measurements. They require instrumentation of the environment and can therefore not be applied in all contexts.

B. Magnetic-Based Localization

Another approach is to exploit the magnetic field for indoor localization. Existing solutions are based on a magnetic map used in a sensor fusion framework with a magnetometer calibration as preprocessing.

1) Magnetic map building: The stability over time of the magnetic field is studied in [10]. Two sets of magnetic data are recorded at same locations with an interval of three months. The correlation of the measured values is high. Distortions created by electronic objects produce only negligible changes in the magnetic field over time: its variation at 30 cm of a computer is less than 0.3 μT (inside a building, standard norm of the magnetic field is above 30 μT). Only large metal objects, such as lifts, cause significant variations of the magnetic field over 2 or 3 meters.

Magnetic maps may be obtained by a complete exploration of a zone [11] (i.e measuring the magnetic field at each position of the map). Most of the time less magnetic data are acquired and the complete map is built by linear interpolation [12] or Gaussian Process regression [13]. These approaches do not consider that the magnetic field verifies Maxwell's equations. In [14], a unique GP on the potential of the magnetic field is learned rather than one GP for each of its 3 axes. This approach improves the quality of the magnetic map. More details on magnetic map building by a Gaussian Process regression are given in Section III.

2) Localization in a magnetic map: Magnetometer data are loosely-coupled with wheel odometry in 2D [13] and with VISLAM algorithm in 3D [4] through particle filtering. In [4] the magnetometer measures are compared with a magnetic map built online by Gaussian Process regression. They present accuracy improvement compared to VISLAM for trajectory that regularly cross the same places. However, the localization suffers from local deteriorations (jitter) due to the loosely-coupled fusion. In [15], magnetometer and visual data are combined to form a descriptor learned through a Neural Network. The map is composed of a visual-magnetic descriptor per cell of size 60×60 cm. A particle filter that compares the current descriptor with those constituting the map performs online localization. The accuracy reached by this approach is limited to the size of a cell. Furthermore, their descriptors are highly dependent to the line of sight.

3) Magnetometer calibration: Standard magnetometer calibration methods require the local magnetic field to be constant. In outdoor environments, this is typically a reasonable assumption. Therefore, outdoor calibration of the magnetometer has been widely studied [16], [17]. The most common approaches are to rotate the magnetometer in all possible directions [16] or to move a metal object around it [17]. Indoors, where magnetic field perturbations occur, calibration of magnetometers is more challenging. In [18], a Helmholtz coil is used to create a uniform magnetic field. In [19], inertial sensors are used to improve the magnetometer calibration. Although this approach has been validated indoors, the calibration must be performed relatively far away from any magnetic materials.

C. Tightly-coupled magneto-visual-inertial fusion

In [20], a tightly coupled localization algorithm with camera, IMU and magnetometers is presented. However, this approach is different than the one we propose since they use a sensor composed of four magnetometers to estimate the velocity by measuring locally the spatial gradient of the magnetic field. The robustness of the algorithm without visual information is improved but some localization drift still remains. Moreover, their sensor composed of four magnetometers is protected by a patent and is challenging to miniaturize. Our solution uses a single magnetometer and a magnetic map to constraint the position of a MSCKF-based algorithm for drift-free localization in the long term.

III. PRELIMINARIES ON GAUSSIAN PROCESS REGRESSION FOR MAGNETIC MAP BUILDING

The magnetic map used in the magneto-visual-inertial MSCKF presented in Section IV is obtained by the Gaussian Process (GP) regression algorithm described in [14]. We give a brief description of this algorithm below.

Usually, in GP regression, the model functions $f(\mathbf{p})$ are assumed to be realizations from a Gaussian random process prior with zero mean and a covariance function $\kappa(\mathbf{p},\mathbf{p}')$ following the model:

$$\begin{aligned} f(\mathbf{p}) &\sim \mathcal{GP}(0, \kappa(\mathbf{p},\mathbf{p}')), \\ y_i = f(\mathbf{p}_i) + \varepsilon_i \ &, \ \varepsilon_i \sim \mathcal{N}(0, \sigma_m^2), \end{aligned} \tag{1}$$

where the observations y_i corrupted by a white Gaussian noise are measured at points of space $\mathbf{p}_i$ for $i \in \{1 \dots n\}$. The GP regression predicts the conditional mean and variance of f at any point of the space of interest $\mathbf{p}^*$ as follows:

$$\begin{aligned} \mathbb{E}[f(\mathbf{p}^*)] &= \mathbf{k}_*^\top (\mathbf{K} + \sigma_m^2 \mathbf{I}_n)^{-1} \mathbf{y}, \\ \mathbb{V}[f(\mathbf{p}^*)] &= \kappa(\mathbf{p}^*, \mathbf{p}^*) - \mathbf{k}_*^\top (\mathbf{K} + \sigma_m^2 \mathbf{I}_n)^{-1} \mathbf{k}_*, \end{aligned} \tag{2}$$

where $\mathbf{K}_{i,j} = \kappa(\mathbf{p}_i, \mathbf{p}_j)$, $\mathbf{k}_*$ is an n-dimensional vector with the i-th entry being $\kappa(\mathbf{p}_i, \mathbf{p}^*)$, and $\mathbf{y}$ a vector concatenating the n observations.

The magnetic map is a vector-field that associates at each point of space $\mathbf{p}^*$ its magnetic field $\mathbf{B}(\mathbf{p}^*)$ with $\mathbf{B} : \mathbb{R}^3 \to \mathbb{R}^3$. Applying Maxwell's equations and under the assumption that the free-current is negligible in most contexts [14], the magnetic-field is approximated as the gradient of a scalar potential $\varphi : \mathbb{R}^3 \to \mathbb{R}$:

$$\mathbf{B} = -\nabla\varphi. \tag{3}$$

It is thus possible to write the magnetic GP regression with a one-dimensional function by learning φ instead of $\mathbf{B}$:

$$\begin{aligned} \varphi(\mathbf{p}) &\sim \mathscr{GP}(0, \kappa(\mathbf{p},\mathbf{p}')), \\ \mathbf{m}_i &= -\nabla\varphi(\mathbf{p}_i) + \varepsilon_i, \\ \kappa(\mathbf{p},\mathbf{p}^*) &= \sigma_{lin}^2 \mathbf{p}^\top \mathbf{p}^* + \sigma_{SE}^2 \exp\left(-\frac{\| \mathbf{p} - \mathbf{p}^* \|^2}{2 l_{SE}^2}\right), \end{aligned} \tag{4}$$

where $\mathbf{m}_i$ is the magnetometer data at point $\mathbf{p}_i$, σ_{SE}^2 is the variance of the scalar potential at any points $\mathbf{p}$ and l_{SE} represents the distance above which a magnetic data no longer influences the predicted magnetic field. Since φ is not a zero-mean function, $\sigma_{lin}^2 \mathbf{p}^\top \mathbf{p}^*$ is added to model the linear part of the potential (corresponding to the Earth magnetic field).

The size of the complete covariance matrix κ is three times the number of data points. If their number is high (a few hundreds or more), GP becomes intractable. To overcome this problem, it is possible to approximate the expression of the covariance matrix so that it is independent of the number of data points:

$$\kappa(\mathbf{p},\mathbf{p}^*) \simeq \Phi(\mathbf{p})\Lambda\Phi(\mathbf{p}^*)^\top \tag{5}$$

with Λ a diagonal matrix (see [14] for its definition) and $\Phi(\mathbf{p}) = \begin{bmatrix}\mathbf{p}^\top & \phi_1(\mathbf{p}) & \phi_2(\mathbf{p}) & \dots & \phi_N(\mathbf{p})\end{bmatrix}$. The functions ϕ_j are the eigenfunctions of the negative Laplace operator, while verifying the Dirichlet boundary conditions:

$$\begin{cases} -\nabla^2\phi_j(\mathbf{p}) = \lambda_j^2 \phi_j(\mathbf{p}), & \mathbf{p} \in \Omega \\ \phi_j(\mathbf{p}) = 0, & \mathbf{p} \in \partial\Omega \end{cases} \tag{6}$$

where Ω is the domain over which the magnetic field is predicted and $\partial\Omega$ its boundaries (see [14] for the explicit expression of λ_j). N corresponds to the number of functions chosen: the greater N is, the more precise the approximation is.

From the approximated expression of κ (Equation 5), the conditional mean and the variance of B, gradient of φ, for all points $\mathbf{p}^*$, are estimated:

$$\begin{cases} \mathbb{E}[B(\mathbf{p}^*)] \approx \nabla\Phi_* \boldsymbol{\mu}_l \\ \mathbb{V}[B(\mathbf{p}^*)] \approx \nabla\Phi_* \Sigma_l \nabla\Phi_*^\top \end{cases} \tag{7}$$

$$\begin{aligned} \text{with } \boldsymbol{\mu}_l &= ([\nabla\Phi]^\top \nabla\Phi + \sigma_m^2 \Lambda^{-1})^{-1} [\nabla\Phi]^\top \mathrm{vec}(\mathbf{m}), \\ \Sigma_l &= \sigma_m^2 ([\nabla\Phi]^\top \nabla\Phi + \sigma_m^2 \Lambda^{-1})^{-1}. \end{aligned}$$

We denote above for simplification $\Phi_* = \Phi(\mathbf{p}^*)$, $\Phi = \Phi(\mathbf{p})$ and $\mathrm{vec}(\mathbf{m})$ a vectorization of all the measures ordered by their index. Equations 7 predict the value of $B(\mathbf{p}^*)$ and its incertitude on the domain Ω.

IV. MAGNETO-VISUAL-INERTIAL MSCKF

In this section, the proposed magneto-visual-inertial MSCKF is presented. The MSCKF [21] is a state-of-the-art visual-inertial odometry algorithm. It is a variant of the classical Extended Kalman Filter (EKF) designed to insert visual measurements inside the EKF framework. The EKF-based framework is particularly adapted for the addition of new sensors, such as the magnetometer.

First, the notations and the different coordinate frames are defined, then the model of the magnetometer is introduced. Finally, the modifications to add magnetic data in the visual-inertial MSCKF *i.e* the state vector including the magnetometer bias, its prediction and the magnetic measurement, are presented.

A. Notations and frames

For 3D transformations to frame F from frame E, the following notations are used:

$$\mathrm{P}_{FE} = \begin{bmatrix} \mathrm{R}_{FE} & \mathbf{t}_{FE} \\ 0_{1\times 3} & 1 \end{bmatrix} \in SE(3)$$

where $\mathrm{R}_{FE} \in SO(3)$ is the rotation matrix from frame E to frame F, $\mathbf{t}_{FE} \in \mathbb{R}^3$ the 3D coordinates of frame F in E.

Hereafter, two referentials will be used: the IMU referential I, and a global referential G. The magnetometer referential is supposed to be confounded with I. This hypothesis is often verified since many IMUs contain a magnetometer with its axis aligned on the IMU referential. The magnetic map is expressed in the Global referential G.

B. Model of the Magnetometer Sensor

The magnetometer measurements are corrupted by different sources of error [16]. Some are due to the sensor itself: a bias on the measures, $\mathbf{b}_{sensor}$, and a scale-misalignment matrix $\mathbf{A}_{sensor}$ that are directly linked to the physics of the sensor and its factory calibration. Moreover, the magnetic elements close to the sensor, such as ferromagnetic materials, modify the measured magnetic field. They are referred to as hard and soft iron effects. The former is due to permanent magnetization resulting in an offset $\mathbf{b}_{HI}$ whereas the latter depends on the orientation of the system towards the magnetic field and is modeled by a matrix $\mathbf{A}_{SI}$. Finally, the measures are deteriorated by a random noise, as for most sensors. The measurement of the magnetic field returned by the sensor is thus given by:

$$\begin{aligned} \hat{\mathbf{m}}_I &= \mathbf{A}_{sensor}(\mathbf{A}_{SI}\mathbf{m}_I + \mathbf{b}_{HI}) + \mathbf{b}_{sensor} + \boldsymbol{\eta}_m \\ \text{with } \boldsymbol{\eta}_m &\sim \mathscr{N}(\mathbf{0}, \sigma_m^2 \mathbf{I}_3) \end{aligned} \tag{8}$$

where $\hat{\mathbf{m}}_I$ is the sensor value, $\mathbf{m}_I$ the real magnetic field and $\boldsymbol{\eta}_m$ the sensor white Gaussian noise ([16]).

Without loss of generality, the model described in Eq. 8 can be summed up by the following model:

$$\begin{aligned} \hat{\mathbf{m}}_I &= \mathbf{A}\mathbf{m}_I + \mathbf{b}_m + \boldsymbol{\eta}_m \\ \text{with } \boldsymbol{\eta}_m &\sim \mathscr{N}(\mathbf{0}, \sigma_m^2 \mathbf{I}_3) \end{aligned} \tag{9}$$

where $\mathbf{A}$ is a scale-misalignment matrix and $\mathbf{b}_m$ is the magnetometer bias.

Experiments on our dataset (see Section V-A) show that $\mathbf{A}$ does not really change over long time spans while the bias $\mathbf{b}_m$ must be re-estimated regularly. A random walk appears to be a suitable representation of its evolution over time. Thus, $\mathbf{A}$ is calibrated once for all [22], [16], [17] and $\mathbf{b}_m$ is added in the MSCKF vector state (see Section IV-C).

C. State vector of the magneto-visual-inertial MSCKF

The state $\mathbf{X}$ of the proposed MSCKF is composed of two parts: one that corresponds to the current state of the system, $\mathbf{x}^C$, and the second to the sliding window of the W last body states, $\mathbf{x}^W$ representing the W last images captured.

$$\begin{aligned} \mathbf{X} &= (\mathbf{x}^C, \mathbf{x}^W), \\ \mathbf{x}^C &= (\mathrm{R}_{GI}, \mathbf{t}_{GI}, \mathbf{v}_{GI}, \mathbf{b}_g, \mathbf{b}_a, \mathbf{b}_m), \\ \mathbf{x}^W &= (\mathrm{R}^1_{GI}, \mathbf{t}^1_{GI}, ..., \mathrm{R}^W_{GI}, \mathbf{t}^W_{GI}), \end{aligned} \tag{10}$$

with $\mathbf{v}_{GI} \in \mathbb{R}^3$ the velocity of frame I in frame G, $\mathbf{b}_g$, $\mathbf{b}_a$ and $\mathbf{b}_m$ respectively the bias of the gyroscope, the accelerometer and the magnetometer. The state definition is the same as the classical visual-inertial MSCKF [21] with addition of $\mathbf{b}_m$. The covariance of the state $\mathbf{X}$ is written as P.

D. Propagation of the magneto-visual-inertial MSCKF

The accelerometer and the gyroscope provide measurements of the acceleration and angular velocity that are integrated to predict the position and orientation and their covariance. The IMU propagation equations are detailed in [21]. With the magnetic data, a propagation equation on the magnetometer bias is added in the MSCKF:

$$\begin{gathered} \dot{\mathbf{b}}_m = \boldsymbol{\eta}_{wm} \\ \text{with } \boldsymbol{\eta}_{wm} \sim \mathcal{N}(\mathbf{0}, \sigma^2_{wm}\mathbf{I}_3). \end{gathered} \tag{11}$$

The propagation of $\mathbf{b}_m$ during Δt is done by updating its covariance:

$$\mathrm{P}_{\mathbf{b}_m} \leftarrow \mathrm{P}_{\mathbf{b}_m} + \Delta t \sigma^2_{wm}\mathbf{I}_3, \tag{12}$$

where $\mathrm{P}_{\mathbf{b}_m}$ is the 3×3 sub-matrix of P corresponding to the covariance of the magnetic bias.

E. Measurements of the magneto-visual-inertial MSCKF

To apply the classical EKF update of the MSCKF, the measurements function $\mathbf{h}$, the Jacobian H and covariance Σ of the sensors data must be defined. For the vision sensors, the measurement function $\mathbf{h}_{vision}$ is the same as the original MSCKF. It is the 2D reprojection of the 3D points on the current pose and on the W last poses. As in [21] the dependency of $\mathbf{h}_{vision}$ on the 3D points is marginalized with the left-nullspace transformation.

For the magnetometer, $\mathbf{h}_{mag}$ is the prediction of the magnetometer values from the magnetic map at the current position $\mathbf{t}_{GI}$:

$$\begin{gathered} \mathbf{h}_{mag}(\mathbf{x}^C) = \mathbf{A}\mathrm{R}^\top_{GI}\mathbf{m}_G(\mathbf{t}_{GI}) + \mathbf{b}_m \\ \text{with } \mathbf{m}_G(\mathbf{t}_{GI}) = \mathbb{E}[B(\mathbf{t}_{GI})] = \nabla\Phi(\mathbf{t}_{GI})\boldsymbol{\mu}_l. \end{gathered} \tag{13}$$

The Jacobian H_{mag} of $\mathbf{h}_{mag}$ with regards to $\mathbf{x}^C$ is decomposed in three parts: H_t the derivative with respect to translation, H_R to rotation and H_b to bias (obviously equal to $\mathbf{I}_3$),

$$\mathrm{H}_{mag} = \begin{bmatrix} \mathrm{H}_R & \mathrm{H}_t & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathrm{H}_b \end{bmatrix}. \tag{14}$$

For H_t, the derivative is straightforward since only the derivative of $\nabla\Phi$ is needed:

$$\mathrm{H}_t = \mathbf{A}\mathrm{R}^\top_{GI}\nabla^2\Phi(\mathbf{t}_{GI})\boldsymbol{\mu}_l, \tag{15}$$

while for H_R, the derivative is conducted with regards to $SO(3)$ and the approximation $\exp_{SO(3)}(\delta\theta) \simeq (I_3 + \lfloor\delta\theta\rfloor_\times)$:

$$\begin{aligned} \mathbf{h}_{mag}\big(\exp_{SO(3)}(\delta\theta)\mathrm{R}_{GI}) - \mathbf{h}_{mag}(\mathrm{R}_{GI})\big) &\simeq \mathbf{A}\big(\lfloor\delta\theta\rfloor_\times\mathrm{R}_{GI}\big)^\top\mathbf{m}_G \\ &\simeq -\mathbf{A}\mathrm{R}^\top_{GI}\lfloor\delta\theta\rfloor_\times\mathbf{m}_G \\ &\simeq \mathbf{A}\mathrm{R}^\top_{GI}\lfloor\mathbf{m}_G\rfloor_\times\delta\theta. \end{aligned}$$

Thus, H_R is given by:

$$\mathrm{H}_R = \mathbf{A}\mathrm{R}^\top_{GI}\lfloor\mathbf{m}_G\rfloor_\times. \tag{16}$$

Finally, the Σ_{mag} covariance is composed of the covariance of the magnetic sensor and the covariance of the GP prediction to take into account the uncertainties of the magnetic map. The uncertainty related to $\boldsymbol{\mu}_l$ is therefore added to the magnetometer noise:

$$\Sigma_{mag} = \sigma^2_m\mathrm{I}_3 + \mathbf{A}\mathrm{R}^\top_{GI}\mathbb{V}[B(\mathbf{t}_{GI})]\mathrm{R}_{GI}\mathbf{A}^\top \tag{17}$$

with $\mathbb{V}[B(\mathbf{t}_{GI})] = \nabla\Phi(\mathbf{t}_{GI})\Sigma_l\nabla\Phi(\mathbf{t}_{GI})^\top$.

V. EXPERIMENTS

In this section, the proposed magneto-visual-inertial MSCKF is evaluated in terms of accuracy, robustness over time and computation time. A comparison with a state-of-the-art VISLAM localization algorithm [1] that reuses visual map is also presented.

A. Experimental setup

To the best of our knowledge, there does not exist any data set with magnetic, visual and inertial data acquired over long periods of time (one year or more). Hence, we needed to create one. Our acquisition platform, illustrated in Figure 2, is composed of a helmet with 4 FLIR Blackfly S cameras and an SBG-Ellipse-N sensor that contains an IMU, a magnetometer (whose axes are aligned with those of the IMU) and a GNSS. All the sensors are rigidly mounted and their data are synchronized. The 4 cameras and the IMU were calibrated using the software Kalibr [23]. The cameras are disposed in two pairs: one stereo pair in front and another one in the rear of the helmet. Several acquisitions of magnetic, visual and inertial data have been captured between May 2020 and August 2021. Online localization algorithms (i.e. the proposed magneto-visual-inertial MSCKF and those used for comparison ([1], [21]) operate on one camera of the helmet (the left front one). The four cameras streams are only used to obtain ground truth positions. They are achieved by a key-frame based VISLAM algorithm with loop closure followed by a global bundle adjustment. It is run on the 4 camera streams, IMU data and GNSS data. The GNSS is used to georeference the ground truth (the first data are acquired outdoor). Ground truth positions are required for magnetic map building (see Section V-B) and for evaluating the localization accuracy reached online by our magneto-visual-inertial MSCKF (Sections V-C and V-E).

Three criteria are used to evaluate the online localization accuracy: the Absolute Translation Error (ATE) and two rotational errors, on the azimuth and the leveling. The ATE is defined as $e = \sqrt{\frac{1}{n}\sum_{j=1}^{n}||\mathbf{t}^j_{GI} - \hat{\mathbf{t}}^j_{GI}||^2}$

with n the total number of data. The rotational errors are $e_{azimuth} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}|(R^j_{GI} \ominus \hat{R}^j_{GI})_z|^2}$ and $e_{leveling} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}||(R^j_{GI} \ominus \hat{R}^j_{GI})_{xy}||^2}$, with $\ominus : SO(3) \times SO(3) \rightarrow \mathbb{R}^3$, $R_1 \ominus R_2 = \log_{SO(3)}(R_1 R_2^{\top})$ which results in a 3-dimensional vector representing the rotation difference towards each axis. $\mathbf{t}_{GI}$ and $\mathbf{R}_{GI}$ are the ground truth positions and rotations and $\hat{t}_{GI}$ and $\hat{R}_{GI}$ their associated online estimation.

Fig. 2. The helmet used for the acquisitions. The IMU SBG-Ellipse-N sensor that includes the magnetometer, the IMU and the GNSS is on top of the helmet while the four cameras FLIR Blackfly S are on the sides.

B. Magnetic Map Building

The preliminary step to online localization by the magneto-visual-inertial MSCKF described in Section IV is the offline reconstruction of a magnetic map (see the upper part of Figure 1). It requires a collection of positions in space $\mathbf{p}_1$, ... $\mathbf{p}_n$ and their respective magnetic measurements $\mathbf{m}_1$, ... $\mathbf{m}_n$. Map accuracy depends on the quality of the input data. For accurate positions in space, we use the ground truth positions obtained with the protocol described in Section V-A. For the magnetic measurements, calibrating the magnetometer before building a magnetic map is crucial.

The magnetic calibration is done outdoors with a classical ellipsoid adjustment method on the whole helmet to estimate accurately both the hard and soft iron effects that generates the system [16]. The sources of soft iron distortions are the components inside the IMU, the cameras and the electric cables. For our acquisition system, their impact on the magnetic measures seems limited since during the calibration, the estimated ellipsoid is always close to a perfect sphere, with differences of dilatation of less than 5% between the three different axes. This is the one and only time the magnetometer is calibrated. During online localization, the bias of the magnetometer is initialized to the value used during the magnetic map building and then refined in the MSCKF.

Once accurate input data are collected, the first step of magnetic map building through Gaussian Process regression is to estimate the hyperparameters σ_{lin}, σ_{SE}, l_{SE} and σ_m. They are optimized through the log marginal likelihood function described in [14]. The following values are obtained $\sigma_{lin} = 25$ μT, $\sigma_{SE} = 15$ μT.m^{-1}, $l_{SE} = 1.3$ m and $\sigma_m = 1.4$ μT. The σ_m value is four times larger than the real sensor noise to compensate for inaccuracies of positions $\mathbf{p}_i$.

In practice, a single GP regression cannot be applied to all the data as the number N of eigenfunctions ϕ would be too high. Thus $5 \times 5 \times 2$ m parallelepipeds tiling the environment are created along the trajectory and a GP regression with $N = 512$ eigenfunctions is learned on each of them. Even though the GP regressions data are associated to $5 \times 5 \times 2$ m parallelepipeds, their eigenfunctions ϕ_j are computed on larger parallelepipeds, which side sizes are increased by $2l_{SE}$ meter. This reduces the effects of Dirichlet boundary conditions and smoothes the magnetic field transition at parallelepipeds boundaries (as explained in [4]). The final magnetic map is therefore the concatenation of the GP learned over all the parallelepipeds.

A magnetic map built from data acquired in May 2020 with the approach described above is shown in Figure 3. The length of the multi-storey learning trajectory used to create this map is 1150 meters. The memory size of the magnetic map is 116 MB (as a comparison the memory size of the visual map built with 4 cameras is 240 MB).

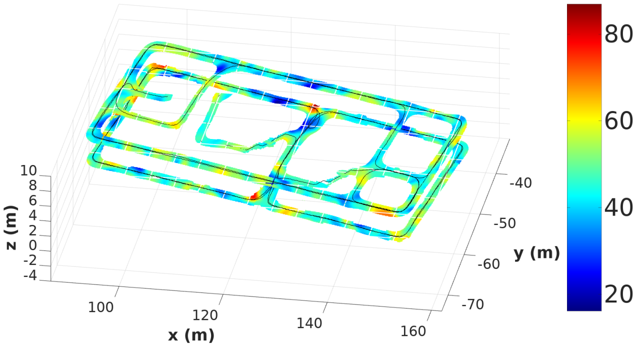

Fig. 3. Norm of the predicted magnetic field (in μT). Only predictions that are less than one meter away from the learning trajectory (in black) are represented.

C. Online localization with the magneto-visual-inertial MSCKF

The interest of adding magnetic data to the MSCKF is evaluated in this Section. The magneto-visual-inertial MSCKF proposed in Section IV is compared to the original visual-inertial MSCKF [21], on a 515 meters sequence acquired in July 2021 (*i.e* one year after the data acquisition to build the magnetic map). The sequence starts outside, GNSS data are then used in the MSCKF to georeference the localization in the coordinate frame of the magnetic map. Once inside the building, when the trajectory enters the domain of the magnetic map, the proposed MSCKF begins to process the magnetometer data. It results in a drift-free localization as illustrated in Figure 4, contrary to the localization of the MSCKF without magnetic data.

The magneto-visual-inertial MSCKF ATE on this sequence is 32.6 cm whereas for the original MSCKF the ATE error is 2.4 meters. In terms of rotation accuracy, the leveling error worsens but not significantly (0.54° against 0.38°) while the azimuth estimation is greatly improved by the addition of the magnetometer: the leveling error is only 0.19° with the proposed algorithm whereas for the classical MSCKF it is 3.51°. During the sequence, the trajectory goes through a corridor that does not belong to the magnetic map domain. In this corridor, the magnetic prediction (Eq. 13) is null and automatically ignored by the MSCKF: thus it runs as a visual-inertial MSCKF. Magnetic data are used again when the trajectory returns inside the magnetic map correcting localization inaccuracies accumulated in the meantime. However, staying for too long outside of areas covered by the magnetic map results in large drift that magnetic data cannot correct. In those cases, to use again the magnetic information, the drift must be corrected beforehand. One possibility is to use magnetic relocalization algorithms ([24], [25]). However, they require to accumulate several seconds of magnetic data to recognize a place without ambiguity. Another possibility is to use visual relocalization, storing jointly in memory a visual map with the magnetic one, but as demonstrated in this paper in Section V-E, it is not robust to appearance changes over time. Finally, visual-magnetic relocalization ([15]) seems promising for robust relocalization over time with less delay than purely magnetic-based solutions. However existing solutions are line-of-sight dependent. Dealing with exits and returns in magnetic maps is a research axis on its own that is outside the scope of the paper.

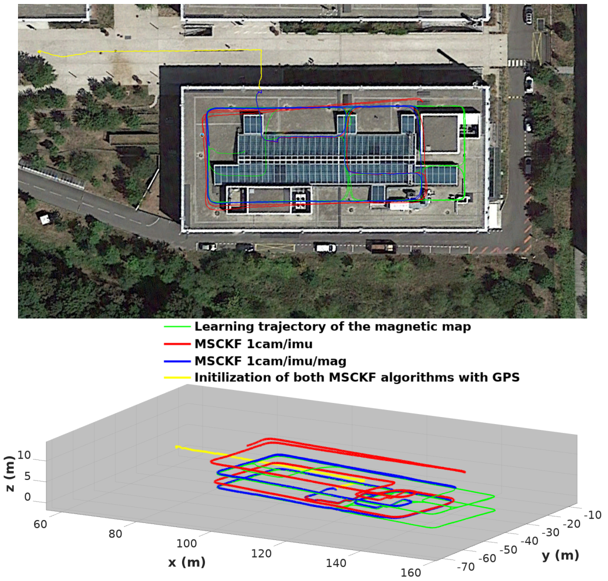

Fig. 4. Localization results of the magneto-visual-inertial MSCKF (Section IV) and the visual-inertial MSCKF [21] on a sequence acquired in July 2021. The magnetic map used in the proposed approach is more than one year old (May 29th, 2020) compared to the test data (July 2021). Top: Google Earth View. Bottom: 3D View.

D. Magnetic bias estimation

The amplitudes of the white Gaussian noise σ_m and the random walk σ_{wm} that are presented in Section IV-B are evaluated from a 50 hour-long magnetic acquisition at rest. On the Allan variance profile (explained in [26]) of the magnetic data, acquired at $f_m = 50$ Hz, we plotted two asymptotic lines, as close as possible to the obtained curves (see Figure 5). The white noise intensity corresponds to the value of the decreasing tangent with a slope of $\frac{-1}{2}$ at $1s$ multiplied by the square root of the sensor frequency: $\sigma_m = \sqrt{f_m} \times 0.047\ \mu\text{T.s}^{1/2} = 0.33\ \mu\text{T}$. The second part of the Allan variance, when it increases regularly, enables to evaluate the random walk law. σ_{wm} is obtained reading the value of the corresponding asymptotic line (of slope $\frac{+1}{2}$) at $3s$: $\sigma_{wm} = 1.93 \times 10^{-4}\ \mu\text{T.s}^{-1/2}$. The magnetic bias thus changes significantly over time (0.56 μT per day on average) and needs to be recalibrated or re-evaluated regularly.

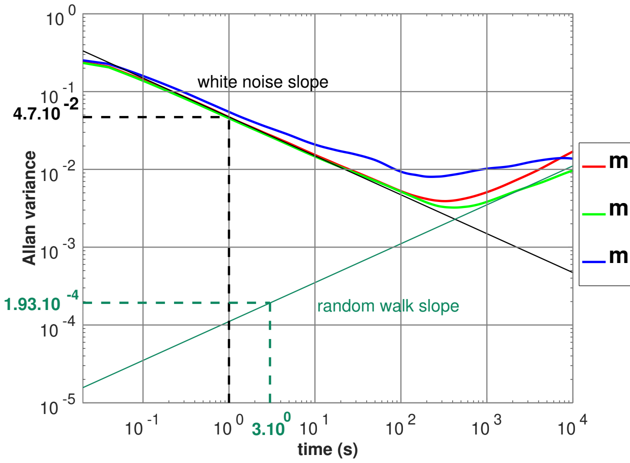

Fig. 5. Allan variance of the magnetometer. Results for a 50 hour-long acquisition.

The proposed MSCKF re-estimates the bias of the magnetometer. For the sequence of Figure 4, the bias initial value is set to the calibration value of May 2020. As shown in Figure 6, the magnetometer bias estimated by the magneto-visual-inertial MSCKF converges after a few seconds to a value close to the ground truth (obtained by recalibrating the magnetometer just before the acquisition with the method described in Section V-B). In one year, the magnetometer bias has changed drastically: from an initial bias of $(-9.07, -10.85, -24.17)\ \mu$T, the values estimated by the proposed MSCKF stabilizes around $(-2.99, 10.03, 17.30)\ \mu$T which is similar to the results of the outdoor calibration $(-2.57, 10.18, 17.39)\ \mu$T.

The initial covariance of the MSCKF must be carefully defined for the bias estimation to be successful. If σ^0_{yaw} and σ^0_{xy} are set to zero as in the visual-inertial MSCKF, the magnetic bias converges to a value that compensates yaw error. For the sensors of the acquisition system presented in Section V-A, $\sigma^0_{yaw} = 0.05$ rad, $\sigma^0_{xy} = 1$ m and $\sigma^0_{b_m} = 10\ \mu$T are relevant values.

The evolution of the magnetic predictions (Eq. 13) throughout the sequence of Figure 4 follows closely the variations of the raw magnetic data as shown in Figure 7. This demonstrates that the magnetic map built one year before, as well as the magnetometer model and the bias estimation method described in Section IV are all valid.

E. Comparison with VISLAM algorithm that reuses visual map

In this section, the proposed magneto-visual-inertial MSCKF is compared with a VISLAM algorithm that reuses a visual map [1]. For that, both algorithms are executed with increasingly older maps on sequences corresponding to different

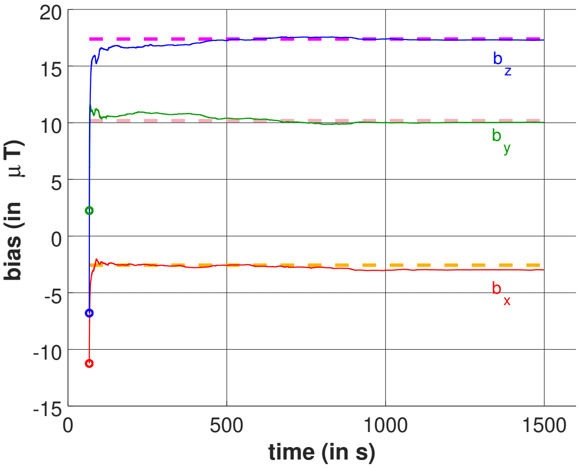

Fig. 6. Bias estimation on the sequence of Figure 4 (July 2021). The dotted lines are the ground truth values from an outdoor calibration performed just before the acquisition.

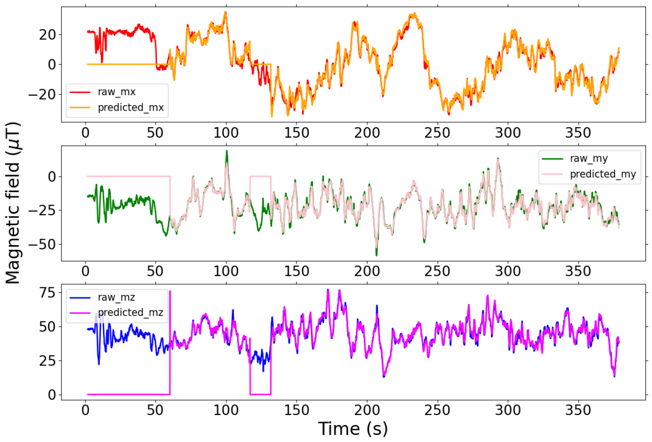

Fig. 7. Comparison of the predicted magnetic field and the raw magnetic data along axis x, y and z for the sequence of Figure 4 (acquired in July 2021). Outside of the magnetic map domain the prediction is null. The sequence starts outdoor for the first 60 seconds. Then between the 120th and 130th seconds the trajectory passes through a corridor which is outside the domain of the magnetic map.

trajectories in the building. The visual maps are reconstructed simultaneously (i.e. on the same sequences) with the magnetic maps from the ground truth poses (see Section V-A). The visual-inertial MSCKF is also run on these sequences. For fair comparison, all the algorithms are initialized at ground truth pose.

Results are reported in Table I. The visual-inertial MSCKF that does not reuse any map is the less accurate algorithm. VISLAM with prebuilt visual map performs slightly better than the magneto-visual-inertial algorithm when the map is recent with respect to the test sequence. Viewpoint recognition used in [1] performs well since the appearance of the environment had not changed much. Moreover visual maps are more locally discriminant than magnetic maps. On the other hand, the average loss of position accuracy of the proposed approach is only 10% over a few months and 14% after one year. For VISLAM with visual map reuse, the position accuracy degrades by an average factor of 2.6 (resp. 5.7) over a few months (resp. over one year). The leveling is well estimated by the classical MSCKF and only slightly deteriorated by the addition of the magnetometer (13% on average and for some sequences its estimation is slightly improved). On the other hand, reusing a visual map worsens the estimation of the leveling by 180% on average compared to the visual-inertial MSCKF. On the contrary, the classical MSCKF lacks of precision in the azimuth. Reusing a visual map in VISLAM greatly improves the quality of its estimation. However, the azimuth accuracy degrades when old maps are reused (4 times less accurate when a visual map that is more than one year old is reused). For the proposed magneto-visual-inertial MSCKF, the azimuth estimation remains accurate no matter when the magnetic map is created. The experimental results demonstrate that the accuracy of the proposed algorithm is almost the same over a year, unlike VISLAM methods reusing visual maps whose localization accuracy is degrading.

The processing time of the magneto-visual-inertial MSCKF and the MSCKF without magnetic data are quite close as illustrated in Table II. The time required by the Kalman step for the magnetometer is short since $\boldsymbol{\mu}_l$ and Σ_l, the matrices of Equation 7, have been pre-computed. Processing one magnetic data takes approximately 1.25 ms which is 15 times faster than the rate of magnetic data acquisition (20 ms per data). The computational overhead on an entire sequence with a frame rate of 20 Hz is about 10%. Furthermore, the magneto-visual-inertial MSCKF is more robust to image frequency reduction than the original MSCKF as illustrated in Table II. The proposed algorithm can process fewer images while remaining accurate.

VI. CONCLUSION

This paper presents a MSCKF-based algorithm that tightly couples magnetic, visual and inertial data for accurate long-term localization. The proposed algorithm is evaluated and compared on a data set with sequences acquired over a time period of more than one year. The experimental results demonstrate that the accuracy of the proposed algorithm is almost the same over a year, unlike VISLAM methods reusing visual maps whose localization accuracy is degrading.

The initialization of the magneto-visual-inertial MSCKF in the magnetic map is currently obtained either outdoor with GNSS data or indoor with a roughly accurate pose. Moreover, getting outside the magnetic map can currently be done only for a short period of time. As discussed in the paper, it seems relevant for future work to combine our approach with magnetic or visual-magnetic relocalization solutions (e.g. [24], [25], [15]) to initialize the MSCKF anywhere in the magnetic map and correct the accumulated drift when getting outside the magnetic map. We also plan to extend the use of magnetic data to other types of VISLAM algorithms such as those based on factor graph optimization [1].

Finally, the magnetic map accuracy may be improved by adding a covariance of the input positions $\mathbf{p}_i$ in the Gaussian Process regression.

Date of online localization		July 23rd, 2021 (2 min 23, 212 m)			August 30th, 2021 (3 min 8, 286 m)			August 31st, 2021 (1 min 57, 182 m)		
Date of magnetic or visual map building		07/23/21	05/18/21	05/29/20	08/30/21	05/18/21	05/29/20	08/31/21	05/18/21	05/29/20
MSCKF 1cam/IMU [21]	ATE	4.417 m	4.417 m	4.417 m	1.888 m	1.888 m	1.888 m	3.001 m	3.001 m	3.001 m
	Leveling err.	**0.38**°	**0.38**°	**0.38**°	0.40°	0.40°	0.40°	0.46°	0.46°	0.46°
	Azimuth err.	3.51°	3.51°	3.51°	2.28°	2.28°	2.28°	0.94°	0.94°	0.94°
MSCKF 1cam/IMU/mag with magnetic map	ATE	0.318 m	**0.293 m**	**0.321 m**	0.189 m	**0.294** m	**0.255 m**	**0.224 m**	**0.216 m**	**0.257 m**
	Leveling err.	0.62°	0.51°	0.54°	**0.37**°	0.37°	**0.37**°	**0.46**°	**0.45**°	**0.46**°
	Azimuth err.	0.29°	**0.18**°	**0.19**°	0.23°	**0.34**°	**0.35**°	**0.36**°	**0.30**°	**0.49**°
VISLAM 1cam/IMU with visual map [1]	ATE	**0.262 m**	0.788 m	0.771 m	**0.108 m**	0.582 m	2.317 m	0.435 m	0.775 m	1.576 m
	Leveling err.	1.04°	1.03°	1.75°	0.41°	**0.32**°	0.99°	1.68°	1.85°	1.43°
	Azimuth err.	**0.19**°	0.40°	0.62°	**0.13**°	0.42°	0.89°	1.09°	1.15°	1.32°

TABLE I

ATE, leveling and azimuth errors of MSCKF [21], MSCKF with Magnetic Map reuse (Section IV) and VISLAM with Visual Map reuse [1] algorithms on three different sequences. These sequences are processed with increasingly older maps (visual map for [1] and magnetic map for the proposed algorithm). For the MSCKF [21] that does not use any map, the same localization error is indicated on the Table for each date of map building.

	Execution Time (in s)	ATE	Leveling error	Azimuth error
MSCKF [21], f_{cam} 20 Hz	108.7	1.888 m	0.40°	2.28°
MSCKF [21], f_{cam} 5 Hz	19.5	diverges after 1 min 8 s		
MSCKF [21], f_{cam} 2 Hz	5.3	diverges after 50 s		
MSCKF mag, f_{cam} 20 Hz	120.0	0.255 m	0.37°	0.34°
MSCKF mag, f_{cam} 5 Hz	36.6	0.260 m	0.38°	0.53°
MSCKF mag, f_{cam} 2 Hz	18.5	0.351 m	0.40°	0.89°

TABLE II

Execution time and localization error of the MSCKF with and without magnetic data by reducing the number of images processed.

References

[1] C. Campos, R. Elvira, J. J. G. Rodriguez, J. M. M. Montiel, and J. D. Tardos, “ORB-SLAM3: An Accurate Open-Source Library for Visual, VisualInertial, and Multimap SLAM,” *IEEE Transactions on Robotics*, 2021, early access.

[2] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, “IMU Preintegration on Manifold for Efficient Visual-Inertial Maximum-a-Posteriori Estimation,” in *RSS*, July 2015.

[3] D. Galvez-Lopez and J. D. Tardos, “Bags of Binary Words for Fast Place Recognition in Image Sequences,” *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188–1197, May 2012.

[4] M. Kok and A. Solin, “Scalable Magnetic Field SLAM in 3D Using Gaussian Process Maps,” in *FUSION*, July 2018.

[5] H. Germain, G. Bourmaud, and V. Lepetit, “S2Dnet: Learning Image Features for Accurate Sparse-to-Dense Matching,” in *ECCV*, December 2020.

[6] M. Labbé and F. Michaud, “Multi-Session Visual SLAM for Illumination Invariant Localization in Indoor Environments,” in *ICRA (submitted)*, March 2021, early access.

[7] X. Zuo, W. Ye, Y. Yang, R. Zheng, T. Vidal-Calleja, G. Huang, and Y. Liu, “Multimodal localization: Stereo over LiDAR map,” *Journal of Field Robotics*, vol. 37, no. 6, pp. 1003–1026, January 2020.

[8] Z. Hashemifar, C. Adhivarahan, A. Balakrishnan, and K. Dantu, “Augmenting Visual SLAM with Wi-Fi Sensing For Indoor Applications,” *Autonomous Robots*, vol. 43, pp. 2245–2260, July 2019.

[9] T. Nguyen, S. Yuan, M. Cao, T. Nguyen, and L. Xie, “VIRAL SLAM: Tightly Coupled Camera-IMU-UWB-LIDAR SLAM,” May 2021, early access.

[10] B. Li, T. Gallagher, A. G. Dempster, and C. Rizos, “How feasible is the use of magnetic field alone for indoor positioning?” in *IPIN*, November 2012.

[11] P. Robertson, M. Frassl, M. Angermann, M. Doniec, B. J. Julian, M. G. Puyol, M. Khider, M. Lichtenstern, and L. Bruno, “Simultaneous Localization and Mapping for pedestrians using distortions of the local magnetic field intensity in large indoor environments,” in *IPIN*, October 2013.

[12] Z. Wu, M. Wen, G. Peng, X. Tang, and D. Wang, “Magnetic-Assisted Initialization for Infrastructure-free Mobile Robot Localization,” in *CIS and RAM*, November 2019.

[13] I. Vallivaara, J. Haverinen, A. Kemppainen, and J. Roning, “Magnetic field-based slam method for solving the localization problem in mobile robot floor-cleaning task,” in *ICAR*, June 2011.

[14] A. Solin, M. Kok, N. Wahlström, T. Schön, and S. Särkkä, “Modeling and Interpolation of the Ambient Magnetic Field by Gaussian Processes,” *IEEE Transactions on Robotics*, vol. 34, no. 4, p. 11121127, September 2015.

[15] Z. Liu, L. Zhang, Q. Liu, Y. Yin, L. Cheng, and R. Zimmermann, “Fusion of Magnetic and Visual Sensors for Indoor Localization: Infrastructure-Free and More Effective,” *IEEE Transactions on Multimedia*, April 2017.

[16] J. F. Vasconcelos, G. Elkaim, C. Silvestre, P. Oliveira, and B. Cardeira, “Geometric Approach to Strapdown Magnetometer Calibration in Sensor Frame,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 2, pp. 1293–1306, April 2011.

[17] F. Vitiello, F. Causa, R. Opromolla, and G. Fasano, “Onboard and External Magnetic Bias Estimation for UAS through CDGNSS/Visual Cooperative Navigation,” *Sensors*, vol. 21, no. 11, p. 3582, May 2021.

[18] K. Mohamadabadi and M. Hillion, “An Automated Indoor Scalar Calibration Method for Three-Axis Vector Magnetometers,” *IEEE Sensors Journal*, vol. 14, no. 9, pp. 3076–3083, September 2014.

[19] M. Kok and T. Schön, “Magnetometer Calibration Using Inertial Sensors,” *IEEE Sensors Journal*, vol. 16, no. 14, pp. 5679–5689, May 2016.

[20] D. Caruso, A. Eudes, M. Sanfourche, D. Vissire, and G. le Besnerais, “Robust indoor/outdoor navigation through magneto-visual-inertial optimization-based estimation,” in *IROS*, September 2017.

[21] A. I. Mourikis and S. I. Roumeliotis, “A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation,” in *ICRA*, vol. preprint, April 2007.

[22] G. Troni and L. Whitcomb, “Adaptive Estimation of Measurement Bias in Three-Dimensional Field Sensors with Angular Rate Sensors: Theory and Comparative Experimental Evaluation,” in *RSS*, June 2013.

[23] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, and R. Siegwart, “Extending *kalibr*: Calibrating the Extrinsics of Multiple IMUs and of Individual Axes,” May 2016.

[24] N. Lee, S. Ahn, and D. Han, “AMID: Accurate Magnetic Indoor Localization Using Deep Learning,” *Sensors*, vol. 18, no. 5, p. 1598, May 2018.

[25] R. Montoliu, J. Torres-Sospedra, and O. Belmonte, “Magnetic field based Indoor positioning using the Bag of Words paradigm,” in *IPIN*. IEEE, October 2016, pp. 1–7.

[26] “IEEE Standard Specification Format Guide and Test Procedure for Single-Axis Interferometric Fiber Optic Gyros,” *IEEE Std 952-1997*, pp. 1–84, 1998.