INDOOR ATTITUDE ESTIMATION USING EQUIPPED GYROSCOPES AND DEPTH SENSORS

Qin Shi[1,2,*], Zhan Song[1], Zhenzhong Xiao[2], Shoubin Chen[2], Fei Wang[3]

[1] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[2] Orbbec Inc., Shenzhen 518062, China
[3] Oxin Inc., Shenzhen 518000, China

Commission IV, WG IV/5

KEY WORDS: Attitude estimation, Indoor localization, Sensor fusion, Depth sensor, Error-state Kalman filter, Manhattan world.

ABSTRACT:

Attitude estimation is central to a wide range of applications such as robotics, virtual reality and mobile smart devices. With the development of sensor technologies, these application devices are often equipped with gyroscopes and depth sensors. In this paper, we propose a novel method to fuse gyroscope and depth information for drift-free and robust attitude estimation in structured indoor applications. Our method relies on the depth information and the Manhattan world assumption to estimate the absolute orientation, which is then fused to correct the accumulated error of the gyroscope-determined attitude. We first utilize the mean shift algorithm on the unit sphere to align the surface normals from the depth measurements with the orthogonal planar structures of the Manhattan world. Therefore, the orientation estimates are drift-free and absolute with respect to the Manhattan world. We then fuse the orientation estimates with the gyroscope measurements in an error-state Kalman filter manner to further improve the attitude estimation accuracy and robustness. We validate the performance of our method on public datasets, demonstrating the robustness and accuracy of the method for attitude estimation.

1. INTRODUCTION

Attitude estimation is the problem of determining the orientation of a rigid body with respect to a reference frame. It has been widely investigated with the guidance, navigation and control communities for several decades (Lefferts et al., 1982). It is a fundamental building block for applications in aerospace, robotics, virtual and augmented reality (VR/AR) and autonomous vehicles. In general, attitude estimation systems require two main types of sensors: high-rate angular rate sensors, namely gyroscopes that measure the instant angular velocity of the body and reference vector sensors that measure a set of known direction vectors in the reference frame. Integrating angular velocities to estimate attitude leads to drift due to measurement noises. Reducing drift is a matter of fusing this information with absolute attitude readings from reference vector sensors.

With advancements in micro-electro-mechanical system (MEMS) technologies, the MEMS-based inertial measurement unit (IMU) has become the most critical sensor for attitude estimation due to its low-cost, small size and low power consumption. An IMU typically consists of tri-axial gyroscopes, tri-axial accelerometers and tri-axial magnetometers. Significant literature has introduced attitude heading reference systems (AHRS) based on IMUs. In AHRS, the reference vectors, namely Earth's gravity, and the Earth's magnetic field are derived from accelerometer and magnetometer measurements. The vectors are then fused to correct the attitude errors accumulated from integration of noisy gyroscope measurements, where sensor fusion methods such as complementary filters (Mahony et al., 2008, Fourati et al., 2010, Wu et al., 2016), Kalman filters (Marins et al., 2001, Crassidis et al., 2007, Del Rosario et al., 2018) or recent Bingham filter (Gilitschenski et al., 2015, Wang and Adamczyk, 2019) are usually employed. Most works assume that the body's acceleration is negligible and there is little electromagnetic interference that perturbs the measurable geomagnetic field (Del Rosario et al., 2018). However, the acceleration assumption is not always possible for a highly dynamic body. Furthermore, indoor environments which may compose of ferromagnetic materials and significant electromagnetic disturbances will lead to divergence of attitude error (De Vries et al., 2009).

Parallel to the work in AHRS there is a lot of work on attitude determination for robotics and consumer electronics applications based on 3D depth sensors. Since depth sensors such as depth cameras and Light Detection and Ranging (LiDAR) sensors have also become lightweight and low-cost. For example, Apple released iPads and iPhones equipped with a LiDAR sensor recently (Dehghan et al., 2021). Most indoor environments consist of orthogonal and parallel planar structures, such as corridors and bedrooms, which exhibit Manhattan World (MW) characteristics (Coughlan and Yuille, 1999). MW is defined by three orthogonal vectors, which form a Manhattan Frame (MF). Under the MW assumption, some recent works perform attitude estimation using the surface normal vectors calculated from depth measurements. The attitude of the body with respect to the surrounding MW is estimated by exploiting the relationship between the surface normal vectors and the dominant direction vectors of the MF. In this way, the depth sensor acts like a reference vector sensor that measures the orientation of the MF. A "structure compass" is introduced in (Straub et al., 2015), where a maximum a posteriori (MAP) inference is used to estimate the orientation of a reference MF. Furthermore, the authors extended the MF model to a mixture of MFs (MMF) and proposed a manifold-aware Gibbs sampling algorithm with Metropolis-Hastings split/merge proposals for adaptive and robust MMF inference (Straub et al., 2017). To

[*] Corresponding author (E-mail address: sqn175@gmail.com).

guarantee a globally optimal solution, (Joo et al., 2016) introduced a branch-and-bound framework to estimate the MF orientation. Instead of MAP inference, (Zhou et al., 2016) proposed a manifold-constrained mean-shift tracking scheme, which is simpler and more computationally efficient. However, the above methods strongly rely on the MW assumption, which will fail when only one or no direction vector of the MF can be found. Furthermore, the depth sensors can only provide a limited field of view (FoV), such as 64.6×50.8 in horizontal FoV and vertical FoV respectively (Orbbec, 2022). This will significantly reduce the chance to find more direction vectors of the MF. Therefore, the depth sensor can only provide intermittent attitude estimation and is less robust.

In this paper, considering that the MF orientation estimation is drift-free with respect to the reference frame and inspired by the AHRS methods, we choose to fuse the depth measurements with the gyroscope measurements to guarantee robustness and accuracy for attitude estimation. The depth sensor acts like a reference vector sensor that measures the absolute orientation of the MF and corrects the accumulated errors from gyroscope measurements. Closest to our spirit is the work in (Straub et al., 2015), which utilized a standard extended Kalman filter (EKF) to fuse gyroscope measurements with the inferred MF orientations. However, the method used a MAP inference to estimate the MF orientation, which requires significant computation cost. And the EKF-based attitude estimation may cause over-parametrization issues (Sola, 2017). In our method, we first develop the MF orientation estimation based on the mean shift algorithm for efficiency (Zhou et al., 2016). A quaternion-based error-state Kalman filter (ESKF) (Sola, 2017) is then used to jointly estimate the attitude and the gyroscope bias. In this way, the noisy gyroscope measurements are used for continuous filter prediction, the MF orientation estimates are used for periodic filter updates when available. The overview of our proposed method is shown in Fig. **1**.

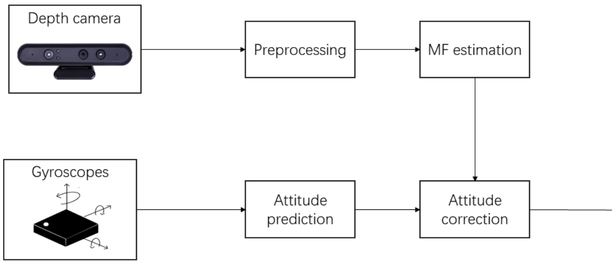

Figure 1. The overview of our proposed indoor attitude estimation method using gyroscopes and depth sensors.

In summary, the key contributions of this work are as follows:

1. A tightly coupled attitude estimation method using the noisy gyroscope measurements and depth measurements in indoor structured environments.

2. Extensive evaluations on public datasets, which demonstrate better performance of our method in terms of accuracy and robustness when compared to the state-of-the-art methods.

2. PROBLAM FORMULATION

In this section, we briefly state the attitude estimation problem in indoor environments and introduce the preliminaries and background theories for the problem.

2.1 Problem Statement

We first define some notations that are employed throughout this paper for clarification. Rotation matrix $\mathbf{R} \in \mathrm{SO}(3)$ and Hamilton quaternion $\mathbf{q}$ are both used to represent rotation, with quaternions used in state vector and rotation matrices used in MF estimation, respectively. The i-th standard basis vector of $\mathbb{R}^n$ is denoted as e_i, i.e., the i-th column of $\mathbf{I}_{n\times n}$. $\mathbf{I}_{n\times n}$ denotes the identity matrix of size n. The quaternion product is denoted by $\otimes$. The n-dimensional unit sphere is denoted by $\mathbb{S}^n = \{\mathbf{x} \in \mathbf{R}^{n+1} |\ \mathbf{x}^T\mathbf{x} = 1\}$. The skew-symmetric matrix of vector $\mathbf{v}$ is denoted as $[\mathbf{v}]_\times$. We denote $(\cdot)^w$ as the world frame, which is defined the same as the MF. $(\cdot)^b$ is the body frame, which coincides with the IMU frame. $(\cdot)^c$ is the depth camera frame. The rotation from frame a to frame b is denoted as $\mathbf{R}_a^b$ or $\mathbf{q}_a^b$.

This paper studies the problem of estimating the attitude $\mathbf{q}_b^w$ in indoor environments utilizing gyroscopes and depth sensors. To be specific, we thoroughly investigate the depth and gyroscope sensor models that relate the measurements with the attitude state to formulate an information fusion framework for accurate and robust attitude estimation.

2.2 Rigid Body kinematics

In this paper, we model mobile devices and robotic systems as rigid bodies. Using the quaternion to represent the attitude (or orientation in reference frame), the kinematics of the rigid-body attitude are given by:

$$\dot{\mathbf{q}} = \frac{1}{2}\mathbf{q} \otimes \boldsymbol{\omega} \tag{1}$$

where $\boldsymbol{\omega}$ are the the angular rates defined locally with respect to the true quaternion. Therefore, the gyroscope measurements can be directly used for integration, as they provide body-referenced angular rates (Sola, 2017).

The rigid body is equipped with an IMU which consists of a 3-axis gyroscope. We note that we use only gyroscope measurements rather than all gyroscope, acceleration and magnetic measurements due to the reasons we have introduced in Sec 1. However, our method in this paper can be easily extended to combine with the acceleration and magnetic measurements like AHRS does.

The gyroscope measurements from IMU are given by

$$\boldsymbol{\omega}_m = \boldsymbol{\omega} + \mathbf{b} + \mathbf{n}_\omega \tag{2}$$

where $\mathbf{b}$ is the gyroscope bias and $\mathbf{n}$ the additive noise. We assume that the noise is Gaussian white noise, $\mathbf{n}_\omega \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}_\omega^2)$. The bias is modeled as the random-walk noise, which derivative is Gaussian white noise,

$$\dot{\mathbf{b}} = \mathbf{n}_b,\ \mathbf{n}_b \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}_b^2) \tag{3}$$

Directly integrating the the gyroscope measurements can obtain the attitude. However, the bias is slowly time-varying and will cause significant estimation error in long-term application. Therefore, the bias shoule be simultaneously estimated and compensated. The rigid body kinematics equipped with gyroscopes are then given as

$$\dot{\mathbf{q}} = \frac{1}{2}\mathbf{q} \otimes (\boldsymbol{\omega}_m - \mathbf{b} + \mathbf{n}_\omega) \tag{4}$$

$$\dot{\mathbf{b}} = \mathbf{n}_b \tag{5}$$

2.3 Manhattan World

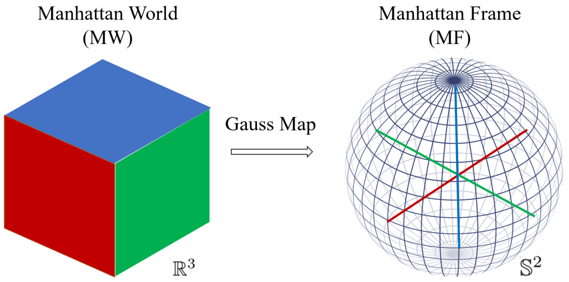

Figure 2. The MW can map to a unit sphere in the surface normal space.

Following the MW assumption, the buildings and objects in structured man-made environments always compose of orthogonal and parallel planes. The MF is then the interpretation of a 3D MW structure using the Gauss Mapping as shown in Fig. **2**. In other words, the MF describes the notion of the MW in the space of surface normals (Straub et al., 2017). For perfect and noise-free MW, the surface normals align with the six orthogonal directions as the columns in

$$\mathbf{E}_w = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \tag{6}$$

In the depth camera frame, the six orthogonal directions are rotated by $\mathbf{R}_w^c$ and given as

$$\mathbf{E}_c = \mathbf{R}_w^c \mathbf{E}_w \tag{7}$$

The attitude represented by $\mathbf{R}_w^c$ is unknown and to be estimated. For surface normal vectors extracted from noisy depth measurements, they should appear some distribution that is more likely to locate around the columns of $\mathbf{E}_c$. In other words, $\mathbf{E}_c$ can be measured by depth measurements. The process of estimating $\mathbf{R}_w^c$ is therefore equivalent to recovering the columns of $\mathbf{E}_c$.

3. ATTITUDE ESTIMATION

3.1 Method Overview

The method overview is shown in Fig. **1**. We start by processing the depth measurements from the depth camera. In the preprocessing module, a bilateral filter is applied to the depth image to filter out outliers. The point cloud is computed from the depth image using the intrinsic matrix of the depth camera and then fed into the MF estimation module for MF orientation estimation. At the same time, the gyroscope measurements are integrated in a high frequency during the ESKF prediction stage. Once the MF orientation estimation is available, the ESKF correction stage is utilized. In this way, the attitude estimation is less prone to gyroscope integration error and MF estimation failure, thus leading to more accuracy and robustness. The attitude estimates are finally streamed out in the frequency of IMU or depth camera.

The system initialization process is done once the MF is successfully estimated for the first time. When the absolute MF is found, the initial attitude can be set and the gyroscope integration process can be continuously performed.

3.2 MF Orientation Estimation

To recover the orthogonal directions of the MW in the depth camera frame, we extract the surface normals from the depth image. After Gauss mapping to unit sphere $\mathbb{S}^2$, the surface normals $\mathbf{n}_i$ always appear some kind of distribution around the orthogonal directions on the unit sphere $\mathbb{S}^2$. Finding the local maxima, i.e., the modes, in this distribution can reveal the orthogonal directions, which is a typical mode-seeking problem. A fast and robust method for solving mode-seeking problems is the popular mean shift algorithm (Carreira-Perpinán, 2015).

We use the SO(3)-manifold constrained mean shift algorithm to align the surface normals from the depth measurements with the orthogonal directions of the MW (Zhou et al., 2016). The details are given in Algorithm 1. The method is composed of two procedures, mode seeking and orientation adjusting. The mode-seeking procedure utilizes the mean shift method for each planar mode. It starts by collecting all the surface normals that are within a neighborhood of the previous orthogonal direction $\mathbf{r}_j$ (line 5). For every surface normal vector in the neighborhood, it is rotated by $\mathbf{Q}$ such that the z-coordinate is along the direction of $\mathbf{r}_j$ (line 8). The normal vector is then lifted to the tangent space $\mathbb{R}^2$ using a Riemann exponential map for convenient distance calculation (line 9). The mean shift can then be easily computed in the tangential space with a Gaussian kernel, where c is the band with of the kernel (line 11). The mean shift is then retracted back to the unit sphere $\mathbb{S}^2$ using the Riemann logarithmic map to update $\mathbf{r}_j$. The updated $\mathbf{r}_j$ is finally rotated back to the depth camera frame using $\mathbf{Q}$ (line 12). In practical implementation, the columns in $\mathbf{E}_c$ can form 24 possible representations of the same one MF. We put together the negative and positive direction of the same orthogonal directions into one mode (line 5). Therefore, we perform mode seeking for three modes $\hat{\mathbf{r}}_1$, $\hat{\mathbf{r}}_2$, $\hat{\mathbf{r}}_3$, which effectively form the MF basis axes.

After mode seeking, we can form the columns of $\mathbf{E}_c$, which is simply $\mathbf{R}_w^c$. Since $\mathbf{E}_w$ is the world reference frame, i.e., the identity matrix $\mathbf{I}$. To satisfy SO(3) orthogonality constraint on rotation matrix $\mathbf{R}_w^c$, we proceed with the orientation adjusting procedure. A singular value decomposition (SVD) method is employed against the combined modes, where λ_i is a weighting factor indicating how certain the observation of the direction is (line 16). The factors are determined by a non-parametric variance approximation using the local Kernel Density Estimation (KDE).

In this way, the drift-free rotation of the rigid body with respect to the reference frame of the MW, i.e., the MF orientation, can be estimated in every depth camera frame. We note that our MF orientation estimation method differs from the method of (Zhou et al., 2016) in that our method utilizes only one mean shift iteration to save computation time. This may lose some accuracy for MF estimation results. However, it is an implementation trade-off since we have an ESKF filtering mechanism to ensure that final attitude estimation results satisfy the accuracy requirements.

Algorithm 1 MF orientation estimation algorithm

```
procedure MODE SEEKING
    for $j \leftarrow 1$ to 3 do
        let $\mathbf{n}_i$ be the $i$th normal vector.
        let $\mathbf{r}_j$ be the $j$th mode.
        Find the collection of normal vectors $\mathbf{N}_j$ satisfying that:
            $\|\mathbf{n}_{i,j} \times \mathbf{r}_j\| < \sin(\theta_{th}), \mathbf{n}_{i,j} \in \mathbf{N}_j$    (8)
        Compute $\mathbf{Q}$ to rotate the normal vectors along the direction of $\mathbf{r}_j$:
            $\mathbf{Q} = \mathbf{I} + [\mathbf{v}]_\times + [\mathbf{v}]_\times^2 \frac{1-c}{s^2}$    (9)
        where $\mathbf{v} = \mathbf{r}_j \times [0,0,1]^T$, $s = \|\mathbf{v}\|$, $c = \mathbf{r}_j^T \cdot [0,0,1]^T$.
        for $\mathbf{n}_{i,j} \in \mathbf{N}_j$ do
            Rotate the normal vectors: $\mathbf{n}_{i,j} \leftarrow \mathbf{Q}\mathbf{n}_{i,j}$
            Compute the rescaled coordinates of $\mathbf{n}_{i,j}$ in the tangential plane of $\mathbf{r}_j$:
                $\mathbf{m}_{i,j} = \frac{\sin^{-1}(\lambda) sign(z_{n_{i,j}})}{\lambda} \begin{pmatrix} x_{n_{i,j}} \\ y_{n_{i,j}} \end{pmatrix}$    (10)
            where $\lambda = \sqrt{x_{n_{i,j}}^2 + y_{n_{i,j}}^2}$.
        end for
        Compute the mean shift in the tangential plane:
            $\mathbf{s}_j = \frac{\sum_{\mathbf{n}_{i,j} \in \mathbf{N}_j} e^{-c\|\mathbf{m}_{i,j}\|^2} \cdot m_{i,j}}{\sum_{\mathbf{n}_{i,j} \in \mathbf{N}_j} e^{-c\|\mathbf{m}_{i,j}\|^2}}$    (11)
        Update the mode:
            $\mathbf{r}_j \leftarrow \mathbf{Q}^T \frac{\hat{\mathbf{r}_j}^T}{\|\mathbf{r}_j\|}$    (12)
        where $\hat{\mathbf{r}_j} = \left[\frac{\tan(\|\mathbf{s}_j\|)}{\|\mathbf{s}_j\|} \mathbf{s}_j^T \; 1\right]$
    end for
end procedure
procedure ORIENTATION ADJUSTING
    Reassemble the estimate $\mathbf{R}_{w,m}^c$ of MF orientation:
        $\mathbf{R}_{w,m}^c = \mathbf{U}\mathbf{V}^T$    (13)
        $[\mathbf{U}, \mathbf{D}, \mathbf{V}] = SVD([\lambda_1 \hat{\mathbf{r}}_1, \lambda_2 \hat{\mathbf{r}}_2, \lambda_3 \hat{\mathbf{r}}_3])$    (14)
end procedure
```

3.3 Attitude Estimation

At the arrival of MF orientation estimation results, the gyroscope biases are rendered observable and thus can be correctly estimated. In this work, we choose to use the ESKF to fuse the information. Considering that the nominal state $\mathbf{x}$ is integrated by the high-frequency gyroscope measurements $\boldsymbol{\omega}_m$, the noise terms $\mathbf{n}_\omega$ are not considered. The integration errors thus will grow. The errors can be collected into the error state $\delta\mathbf{x}$ and estimated in the ESKF. As a consequence, the magnitudes of the error states are very small, and its evolution function can be defined by a linear dynamic system associated with the values of the nominal state. Since the attitude error state can be parametrized by a minimal three degrees of freedom, the attitude over-parametrization issues can be avoided. When integrating the nominal state, the ESKF predicts the error state under the Gaussian assumption. At the arrival of the MF orientation estimation, the ESKF correction is performed. To this end, the error state is observable and can be estimated. Finally, ESKF injects the error state into the nominal state and updates the covariance matrix of the error state.

For our attitude estimation method, we define the nominal state as follows:

$$\mathbf{x} = [\mathbf{q}_b^w, \mathbf{b}] \tag{15}$$

For notation brevity, the superscript and subscript of $\mathbf{q}_b^w$ are omitted. The corresponding error-state is given as:

$$\delta\mathbf{x} = [\delta\boldsymbol{\theta}, \delta\mathbf{b}] \tag{16}$$

Using the first-order integration, the nominal state kinematics in discrete time are written as:

$$\mathbf{q} \leftarrow \mathbf{q} \otimes \mathbf{q}\{(\boldsymbol{\omega}_m - \mathbf{b})\Delta t\} \tag{17}$$

$$\mathbf{b} \leftarrow \mathbf{b} \tag{18}$$

where $\mathbf{q}\{\mathbf{v}\}$ is the associated quaternion to the rotation vector $\mathbf{v}$:

$$\mathbf{q}\{\mathbf{v}\} = \exp(\mathbf{v}) = \exp(\phi\mathbf{u}) = \begin{bmatrix} \cos(\phi/2) \\ \mathbf{u}\sin(\phi/2) \end{bmatrix} \tag{19}$$

where $\mathbf{v} = \phi\mathbf{u}$ is the rotation vector with ϕ the rotation angle and $\mathbf{u}$ the rotation axis.

The error-state kinematics in discrete time are then derived as:

$$\delta\boldsymbol{\theta} \leftarrow \mathbf{R}^T\{(\boldsymbol{\omega}_m - \mathbf{b})\Delta t\}\delta\boldsymbol{\theta} - \delta\mathbf{b}\Delta t + \boldsymbol{\theta}_\mathbf{i} \tag{20}$$

$$\delta\mathbf{b} \leftarrow \delta\mathbf{b} + \mathbf{b}_\mathbf{i} \tag{21}$$

where $\boldsymbol{\theta}_\mathbf{i}$ and $\mathbf{b}_\mathbf{i}$ are the random zero-mean Gaussian impulses with covariance matrices as $\boldsymbol{\Theta}_\mathbf{i} = \boldsymbol{\sigma}_\omega^2 \Delta t^2$, $\mathbf{B}_\mathbf{i} = \boldsymbol{\sigma}_b^2 \Delta t$, respectively. $\mathbf{R}\{\mathbf{v}\}$ is the associated rotation matrix to the rotation vector $\mathbf{v}$.

The error-state system transition model is written as

$$\delta\mathbf{x} \leftarrow f(\mathbf{x}, \delta\mathbf{x}, \boldsymbol{\omega}_m, \mathbf{i}) = \mathbf{F}(\mathbf{x}, \boldsymbol{\omega}_m) \cdot \delta\mathbf{x} + \mathbf{i} \tag{22}$$

where

$$\mathbf{i} = \begin{bmatrix} \boldsymbol{\theta}_\mathbf{i} \\ \mathbf{b}_\mathbf{i} \end{bmatrix} \tag{23}$$

The ESKF state and covariance prediction equations are given as

$$\hat{\delta\mathbf{x}} \leftarrow \mathbf{F}(\mathbf{x}, \boldsymbol{\omega}_m) \cdot \hat{\delta\mathbf{x}} \tag{24}$$

$$\mathbf{P} \leftarrow \mathbf{F}\mathbf{P}\mathbf{F}^T + \mathbf{Q}_\mathbf{i} \tag{25}$$

such that $\delta\mathbf{x} \sim \mathcal{N}(\hat{\delta\mathbf{x}}, \mathbf{P})$. $\mathbf{Q}_\mathbf{i}$ is the covariance matrix of $\mathbf{i}$. $\mathbf{F}$ is the Jacobian of $f()$ with respect to $\delta\mathbf{x}$ and is written as

$$\mathbf{F} = \frac{\partial f}{\partial \delta\mathbf{x}}\Big|_{\mathbf{x},\boldsymbol{\omega}_m} = \begin{bmatrix} \mathbf{R}^T\{(\boldsymbol{\omega}_m - \mathbf{b})\Delta t\} & -\mathbf{I}\Delta t \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{26}$$

At the arrival of the MF orientation estimation, the system observation model can be written as

$$\mathbf{y} = h(\mathbf{x}) + \mathbf{v} \tag{27}$$

where $y = \mathbf{q}\{\mathbf{R}^w_{c,m}\mathbf{R}^c_b\} = \begin{bmatrix} q_w & q_x & q_y & q_z \end{bmatrix}$ is the associated quaternion of MF orientation estimate in the body frame, $h()$ is equal to $\mathbf{I}$ and $\mathbf{v}$ is a zero-mean white Gaussian noise with covariance $\mathbf{V}$. we therefore proceed with the ESKF correction equations:

$$\mathbf{K} = \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{V})^{-1} \tag{28}$$

$$\hat{\delta \mathbf{x}} \leftarrow \mathbf{K}(\mathbf{y} - h(\hat{\delta \mathbf{x}})) \tag{29}$$

$$\mathbf{P} \leftarrow (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P} \tag{30}$$

where $\mathbf{H}$ is the Jacobian matrix of $h()$ with respect to the error state $\delta\mathbf{x}$:

$$\mathbf{H} = \frac{\partial h}{\partial \delta \mathbf{x}}\Big|_{\mathbf{x}} = \frac{1}{2}\begin{bmatrix} -q_x & -q_y & -q_z & 0 & 0 & 0 \\ q_w & -q_z & q_y & 0 & 0 & 0 \\ q_z & q_w & -q_x & 0 & 0 & 0 \\ -q_y & q_x & q_w & 0 & 0 & 0 \end{bmatrix} \tag{31}$$

We note that $\mathbf{V}$ is set as a fixed value in this work. Readers are encouraged to derive an appropriate value of $\mathbf{V}$ by taking into consideration the uncertainties of the estimate of MF orientation.

After this correction, the observed error state is used to update the nominal state:

$$\mathbf{q} \leftarrow \mathbf{q} \otimes \mathbf{q}\{\hat{\delta \boldsymbol{\theta}}\} \tag{32}$$

$$\mathbf{b} \leftarrow \mathbf{b} + \hat{\delta \mathbf{b}} \tag{33}$$

To this end, the information from depth measurements and gyroscope measurements are tightly fused and the optimal attitude estimate can therefore be continuously output.

4. EXPERIMENTAL EVALUATIONS

In this section, the experimental evaluations are performed to show the performance of our proposed method. We start by giving a detailed implementation and parameter settings of our method. We then evaluate the method on a real-world dataset to show the robustness and accuracy of our method. The evaluation results are also compared to the methods which use only gyroscope measurements or depth measurements.

4.1 Detailed Implementation

For computation efficiency, we calculate the surface normals on every 10 × 10 depth pixel block using the Area Weighted method (Klasing et al., 2009). The parameter θ_{th} is set to be a small value of 10°, since we can keep a smooth track of the orientation of the MF. During implementation, we also require that the normals within θ_{th} should have a minimum number of 100 for more robust mode seeking. The factor c is set to be 15. The depth camera intrinsic matrix and the extrinsic rotation matrix between depth camera and gyroscope are all given by sensors suits settings.

4.2 Experimental Results

We perform the experimental evaluations on the open-source OpenLORIS-Scene dataset (Shi et al., 2020). The dataset provides visual and inertial data recorded with real-world wheeled robots in real indoor scenes. The equipped sensor is primarily a RealSense D435i camera, which is mounted at a fixed height of 1m. The camera records the depth images with a resolution of 848 × 480 in 30Hz and provides gyroscope data with a frequency of 400Hz. The gyroscopes are hardware synchronized to the image sensor. The ground-truth robot trajectories are derived from offline LiDAR Simultaneous Localization and Mapping (SLAM) based on the Hokuyo laser scans.

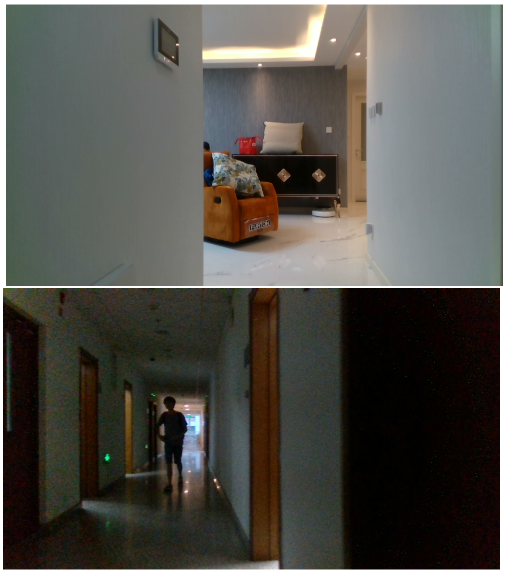

Figure 3. The structured indoor scenes from OpenLORIS-Scene dataset. (top) *home1* scene. (bottom) *corridor1* scene.

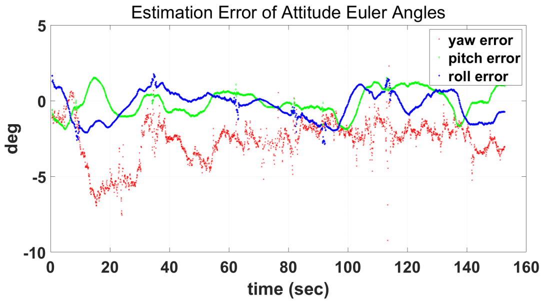

Figure 4. The attitude estimation error for the *home1* consequence.

We evaluate our method on the dataset sequences of *home1* and *corridor1*, since they are collected on a structured indoor environments. The duration of *home1* sequence is 153 seconds, *corridor1* 115 seconds. The snapshots of the scenes are given in Fig. **3**.

In our experimental implementation, the gyroscope noise characteristics are derived from the camera datasheet, the covariance matrix $\mathbf{V}$ is set to be $\mathbf{I}_{4\times 4} \cdot 1e^{-5}$. We first present the effectiveness of our method based on the *home1* consequence. The attitude estimation errors are given in Fig. **4**. The estimated attitudes along with the ground-truth are shown in Fig. **5**-**7**. The attitudes are expressed in the form of Euler angles with *yaw, pitch, roll* sequence. From the result, we can see that combining the depth measurements and gyroscope data can produce comparable results with the ground-truth attitude, thus proving the effectiveness of our method.

We then show the performance of our method in terms of accuracy and robustness by comparing against MWO (Zhou et al., 2016), which is the state-of-the-art attitude estimation method for structured indoor environments using only depth measure-

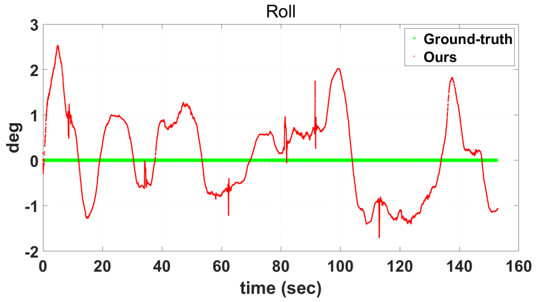

Figure 5. Comparison of our estimated roll angle with the ground-truth for the *home1* consequence.

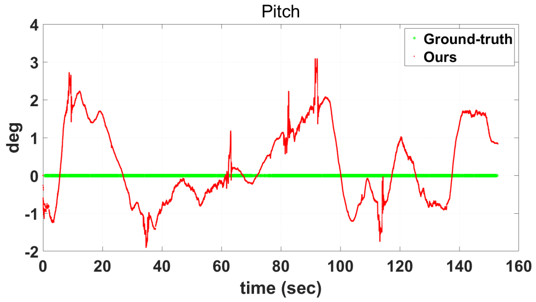

Figure 6. Comparison of our estimated pitch angle with the ground-truth for the *home1* consequence.

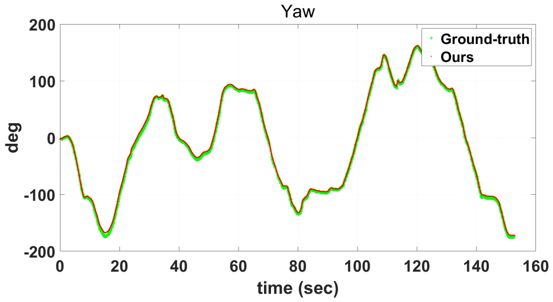

Figure 7. Comparison of our estimated yaw angle with the ground-truth for the *home1* consequence.

Scene	Ours		DR		MWO	
	RMS	Max	RMS	Max	RMS	Max
home1	3.3	9.4	32.8	54.2	1.9*	3.9*
corridor1	3.6	10.7	20.7	38.3	3.5*	13.1*

*: MWO failed to run the entire sequence.

Table 1. Performance Comparison (unit: degrees).

ments. We also implement the integration method using only gyroscope measurements, namely Dead Reckoning (DR), for comparison. The results against the datasets are summaries in Table. **1**. It can be easily shown that our method achieves better accuracy than DR in terms of root mean square (RMS) and maximum angular error. We also note that MWO failed when it runs to 19% of the sequence *home1* and 30% of *corridor1*. The performance of MWO is relatively better than ours since the performance statistic are collected in a much shorter range of data. However, MWO is not guaranteed to run successfully for entire sequences. This clearly shows the robustness of our method when compared with the state-of-the-art method.

It is worth mentioning that our method cannot process the entire sequence *corridor1*. This is because some part of the scene does not follow the MW assumption, which has multiple vertical walls that are not orthogonal to each other. This is typically an Atlanta World (AW), which is outside the scope of this paper. The extension of our work from MW to AW is left as future work. For a fair comparison, all methods are performed on the same segment of *corridor1*.

5. CONCLUSIONS

In this paper, we proposed a novel attitude estimation method for structured indoor environments that integrates depth and gyroscope measurements. The depth information is used to provide absolute orientation by modeling the structured man-made environments as a MW. The depth sensor is then treated as a reference vector sensor, which renders the gyroscope biases observable. In this way, this paper tightly fused the depth and gyroscope measurements using a quaternion-based ESKF. To realize the information fusion, we made several technical contributions, including the one-iteration mean shift algorithm for MF orientation estimation, and the underlying dynamic system modeling. Extensive evaluations on real-world datasets demonstrate the effectiveness, accuracy and robustness of our proposed method.

ACKNOWLEDGEMENTS

This work has been funded through the National Natural Science Foundation of China (42101445) and the China Postdoctoral Science Foundation (2021M702229).

REFERENCES

Carreira-Perpinán, M. A., 2015. A review of mean-shift algorithms for clustering. *arXiv preprint arXiv:1503.00687*.

Coughlan, J. M., Yuille, A. L., 1999. Manhattan world: Compass direction from a single image by bayesian inference. *Proceedings of the seventh IEEE international conference on computer vision*, 2, IEEE, 941–947.

Crassidis, J. L., Markley, F. L., Cheng, Y., 2007. Survey of nonlinear attitude estimation methods. *Journal of guidance, control, and dynamics*, 30(1), 12–28.

De Vries, W., Veeger, H., Baten, C., Van Der Helm, F., 2009. Magnetic distortion in motion labs, implications for validating inertial magnetic sensors. *Gait & posture*, 29(4), 535–541.

Dehghan, A., Baruch, G., Chen, Z., Feigin, Y., Fu, P., Gebauer, T., Kurz, D., Dimry, T., Joffe, B., Schwartz, A. et al., 2021. ARKitScenes-A Diverse Real-World Dataset for 3D Indoor Scene Understanding Using Mobile RGB-D Data.

Del Rosario, M. B., Khamis, H., Ngo, P., Lovell, N. H., Redmond, S. J., 2018. Computationally efficient adaptive error-state Kalman filter for attitude estimation. *IEEE Sensors Journal*, 18(22), 9332–9342.

Fourati, H., Manamanni, N., Afilal, L., Handrich, Y., 2010. A nonlinear filtering approach for the attitude and dynamic body acceleration estimation based on inertial and magnetic sensors: Bio-logging application. *IEEE Sensors Journal*, 11(1), 233–244.

Gilitschenski, I., Kurz, G., Julier, S. J., Hanebeck, U. D., 2015. Unscented orientation estimation based on the Bingham distribution. *IEEE Transactions on Automatic Control*, 61(1), 172–177.

Joo, K., Oh, T.-H., Kim, J., Kweon, I. S., 2016. Globally optimal manhattan frame estimation in real-time. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1763–1771.

Klasing, K., Althoff, D., Wollherr, D., Buss, M., 2009. Comparison of surface normal estimation methods for range sensing applications. *2009 IEEE international conference on robotics and automation*, IEEE, 3206–3211.

Lefferts, E. J., Markley, F. L., Shuster, M. D., 1982. Kalman filtering for spacecraft attitude estimation. *Journal of Guidance, Control, and Dynamics*, 5(5), 417–429.

Mahony, R., Hamel, T., Pflimlin, J.-M., 2008. Nonlinear complementary filters on the special orthogonal group. *IEEE Transactions on automatic control*, 53(5), 1203–1218.

Marins, J. L., Yun, X., Bachmann, E. R., McGhee, R. B., Zyda, M. J., 2001. An extended kalman filter for quaternion-based orientation estimation using marg sensors. *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180)*, 4, IEEE, 2003–2011.

Orbbec, 2022. Orbbec femto iToF 3D sensor.

Shi, X., Li, D., Zhao, P., Tian, Q., Tian, Y., Long, Q., Zhu, C., Song, J., Qiao, F., Song, L., Guo, Y., Wang, Z., Zhang, Y., Qin, B., Yang, W., Wang, F., Chan, R. H. M., She, Q., 2020. Are we ready for service robots? the OpenLORIS-Scene datasets for lifelong SLAM. *2020 International Conference on Robotics and Automation (ICRA)*, 3139–3145.

Sola, J., 2017. Quaternion kinematics for the error-state Kalman filter. *arXiv preprint arXiv:1711.02508*.

Straub, J., Bhandari, N., Leonard, J. J., Fisher, J. W., 2015. Real-time manhattan world rotation estimation in 3d. *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 1913–1920.

Straub, J., Freifeld, O., Rosman, G., Leonard, J. J., Fisher, J. W., 2017. The Manhattan frame model—Manhattan world inference in the space of surface normals. *IEEE transactions on pattern analysis and machine intelligence*, 40(1), 235–249.

Wang, W., Adamczyk, P. G., 2019. Comparison of bingham filter and extended Kalman filter in IMU attitude estimation. *IEEE Sensors Journal*, 19(19), 8845–8854.

Wu, Z., Sun, Z., Zhang, W., Chen, Q., 2016. A novel approach for attitude estimation based on MEMS inertial sensors using nonlinear complementary filters. *IEEE Sensors Journal*, 16(10), 3856–3864.

Zhou, Y., Kneip, L., Rodriguez, C., Li, H., 2016. Divide and conquer: Efficient density-based tracking of 3d sensors in manhattan worlds. *Asian Conference on Computer Vision*, Springer, 3–19.