

Article

Nonlinear Complementary Filter for Attitude Estimation by Fusing Inertial Sensors and A Camera

Lingxiao Zheng , Xingqun Zhan *  and Xin Zhang 

School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai 200240, China;
lx_zheng@sjtu.edu.cn (L.Z.); xin.zhang@sjtu.edu.cn (X.Z.)

* Correspondence: xqzhan@sjtu.edu.cn

Received: 27 October 2020; Accepted: 19 November 2020; Published: 26 November 2020



Abstract: Using a standalone camera for pose estimation has been quite a standard task. However, the point correspondence-based algorithms require at least four feature points in the field of view. This paper considers the situation that there are only two feature points. Focusing on the attitude estimation, we propose to fuse a camera with low-cost inertial sensors based on a nonlinear complementary filter design. An implicit geometry measurement model is derived using two feature points in an image. This geometry measurement is fused with the angle rate measurement and vector measurement from inertial sensors using the proposed nonlinear complementary filter with only two parameters to be adjusted. The proposed nonlinear complementary filter is posed directly on the special orthogonal group $SO(3)$. Based on the theory of nonlinear system stability analysis, the proposed filter ensures locally asymptotic stability. A quaternion-based discrete implementation of the filter is also given in this paper for computational efficiency. The proposed algorithm is validated using a smartphone with built-in inertial sensors and a rear camera. The experimental results indicate that the proposed algorithm outperforms all the compared counterparts in estimated accuracy and provides competitive computational complexity.

Keywords: consumer electronics; attitude estimation; inertial sensors; camera; nonlinear complementary filter

1. Introduction

Attitude estimation using low-cost sensors plays an important role in many consumer electronic applications and has attracted much research attention. For example, in rehabilitation and biomedical engineering, the attitude information is applied for elderly fall detection [1]. In indoor pedestrian dead reckoning application, the attitude of the measurement unit is used for step detection and heading estimation [2].

MARG (magnetic, angular rate, and gravity) sensors and monocular cameras are two kinds of low-cost sensors that are widely used in consumer electronic applications to provide attitude information [3–5]. However, using MARG sensors or camera standalone for attitude estimation has some limitations.

On the one hand, the MARG sensors contain a magnetometer which is used to correct heading drift of the attitude estimation. Generally speaking, the magnetometer is factory calibrated to compensate for any error sources that are internal to the device. However, for the errors that are introduced externally by mounting structures or adjacent devices, an additional calibration process is essential [6]. In order to calibrate the magnetometer, the device needs to be moved in all possible directions to collect data. This is not user-friendly. Moreover, when it works in an environment with abnormal magnetic fields, the attitude estimation performance will deteriorate significantly.

On the other hand, when there are artificial vision fiducials arranged in the environment, the attitude and position of a camera can be recovered from one image by solving the perspective-and-point (PnP) problem [7,8]. The point correspondence-based algorithms require at least four feature points in the field of view. However, in a dynamic and possibly cluttered environment, the number of feature points may be less than four.

The goal of this work is to propose a low-cost fusion method to achieve absolute attitude estimation in an environment with only two pre-calibrated artificial vision fiducials. The sensors to be fused include gyroscope, accelerometer, and camera. The advantages of such a sensor combination are twofold:

- Compared with the MARG sensor combination, our method can work in an abnormal magnetic field environment, especially in an indoor environment.
- Compared with the combination of gyroscope and camera, our method still converges when there are only two feature points in the image.

The second one is important for the resource-constrained artificial fiducial system, such as the AprilTag system [9] and the visible light communication reference system [10,11].

The main contributions of our work are summarised as:

- We derive an implicit geometry measurement for camera-based attitude estimation. This measurement is associated with attitude and is independent of the position of the camera.
- A nonlinear complementary filter is proposed to fuse angle rate measurement, vector measurement, and geometry measurement. There are only two parameters to be adjusted.

The remainder of this paper is organized as follows. Section 2 explores the literature of attitude estimation algorithms based on MARG sensors, camera standalone, and visual-inertial fusion, respectively. Section 3 presents the sensor models including angle rate measurement, vector measurement, and the proposed geometry measurement model. Section 4 presents the nonlinear complementary filter fusing inertial sensors and a camera. The stability analysis of the proposed filter is in this section. An attitude initial alignment method is proposed to provide the initial value of the filter. The discrete implementation of the filter on quaternion is also given in this section. The algorithms are validated using data collected by a smartphone with built-in inertial sensors and a rear camera. Three other representative methods of attitude estimation algorithms are also implemented on the collected data. The results are shown in Section 5. Finally, concluding remarks and future work are presented in Section 6.

2. Related Works

Various solutions have been proposed for attitude estimation using low-cost sensors, including (1) MARG (magnetic, angular rate, and gravity) sensors-based methods, (2) monocular camera-based methods.

For MARG sensors-based methods, the gyroscope provides angle rate measurement. The accelerometer and magnetometer provide attitude associated vector measurements. When the initial attitude is known, attitude can be computed by integrating the angle rate measurement [12]. Meanwhile, the attitude can be constructed directly from the vector measurements [13]. To estimate the attitude from vector measurements is to solve a least square problem, the Wahba's problem. A unique closed-form solution can be provided by the QUEST (quaternion estimator) algorithm in [14] and singular value decomposition (SVD)-based method in [15].

The attitude estimation accuracy obtained by numerical integration of the angle rate measurement is good in a short time. However, the bias and noise of the gyroscope make the estimated value deviate more and more from the true value over time. On the other hand, the attitude recovered from vector measurements hosts long-term stability. However, the instantaneous linear acceleration and magnetic field anomalies will decrease the estimation accuracy.

To achieve good bandwidth and long-term stability, many MARG sensors-based attitude estimation algorithms resort to fuse the angle rate measurement and the vector measurements. The classic fusion algorithms are based on the extended Kalman filter [16,17]. These stochastic approaches involve the update of the error covariance matrix and gain matrix which will lead to a large computational burden.

The Mahony's nonlinear complementary filter formulates the fusion problem as deterministic nonlinear observer kinematics on the special orthogonal group [18]. The observer kinematics include a prediction term based on the angle rate measurement and a correction term derived from the estimation residual. To calculate the correction term, the direct and passive versions of Mahony's complementary filter rely on the algebraic reconstruction of attitude from vector observations, while the explicit version explicitly uses the cross product of the reference vectors and the observed vectors. Mahony's complementary filter has only two adjustable parameters and ensures almost global asymptotic stability. Two passive nonlinear complementary filters algorithms that are implemented on quaternion are proposed in [19,20]. The algebraic reconstruction of attitude from vector observation is based on Levenberg–Marquardt optimization algorithm in [19] and singular value decomposition (SVD) in [20], respectively.

Different from the nonlinear complementary filter, the linear complementary filter linearly combines the attitude quaternion integrated from the angular velocity with the one reconstructed from the vector observations. Therefore, the linear complementary filter has a frequency domain interpretation. The Madgwick's linear complementary filter in [21] applies the gradient descent algorithm to solve the quaternion version of Wahba's problem. The optimization algorithm only computes one iteration per time sample, provided that the convergence rate of the estimated attitude is equal to or greater than the change rate of physical orientation. In [22], a fast complementary filter is proposed by deriving a quaternion increment that is free of iterations. An improved gradient descent based attitude complementary filter in [23] provides fast error convergence and robustness by decoupling the magnetic field variance from roll and pitch. Gain-scheduled or adaptive complementary filters are more robust to strong accelerations and magnetic field disturbances than gain-fixed complementary filters [4,24].

For attitude estimation, in addition to MARG sensors, the monocular camera is also an attractive low-cost sensor. Using a camera standalone for pose estimation is quite a standard task. When there are artificial vision fiducials arranged in the environment, the attitude and position of the camera can be recovered from one image by solving the PnP problem [7,8]. In an environment without artificial fiducials, the relative rotation and scaled translation can be restored from two images with natural features. This problem has been extensively researched, and a large number of algorithms have been developed. The most well-known ones are the 8-point algorithm [25] and the 5-point minimal algorithm [26].

Considering that the frame rate of the low-cost camera is relatively low, it is expected to fuse inertial sensors and a camera to get a higher data rate. The monocular visual-inertial system (VINS) based on extended Kalman filter [27] or bundle adjustment formulation [28] can provide relative attitude estimation. However, VINS or standalone camera pose estimation simultaneously calculates the attitude and position of the camera. When the attitude is the only one to be interested, position and attitude should be decoupled to avoid unnecessary calculations.

Recently, a generalized linear complementary filter for attitude estimation from multi-sensor measurements is proposed in [29]. The point-correspondence constraints of the camera are modeled as vector measurements. This model allows the camera to be fused with a gyroscope in the same way as an accelerometer and a magnetometer. However, to satisfy the premise of vector measurement, the position of the camera must be close enough to the origin of the reference coordinate system.

An implicit measurement model proposed in [30] enables a strict decoupling of attitude and position. The measurements of the camera are fused with angle rate measurement using a nonlinear observer. However, this implicit measurement model is based on line-correspondences instead of

point-correspondence. Compared with the point feature, tracking the line feature is computationally more intensive.

In this paper, we use two feature points to derive an implicit geometry measurement that has the same expression as the implicit measurement in [30], but without tracking the line-correspondences. The new geometry measurement is fused with the vector measurement from accelerometers and the angle-rate measurement from gyroscopes. This combination of sensors makes it possible to determine the absolute attitude with only two feature points in the field of view. Meanwhile, the fusion of these three kinds of measurements removes the restrictions on the position of the camera in [29].

3. Sensor Models

This section presents the sensor measurement models for attitude estimation. The angle-rate measurement and vector measurement from gyroscopes and accelerometers are briefly described. The camera geometry measurement that is associated with the attitude but is independent of the position of the camera is derived in detail.

3.1. Inertial Sensors

In this paper, the body-fixed frame of reference $\{b\}$ is a right-forward-up coordinate system. The navigation frame of reference $\{n\}$ is an east-north-up coordinate system. The direction cosine matrix C_b^n denotes the relative orientation of $\{b\}$ with respect to $\{n\}$. To avoid the repeated occurrence of superscript and subscript, we use R to represent C_b^n . R and C_b^n belong to special orthogonal group denoted by $SO(3)$.

The measurements available from inertial sensors are 3 axis gyroscopes and 3 axis accelerometers.

Gyroscopes measure the angle rate of $\{b\}$ relative to $\{n\}$ expressed in $\{b\}$. We assume that the initial bias of the gyroscope has been calibrated in the initial static stage and is subtracted from the gyroscope measurement. Therefore, the angle-rate measurement model for a low-cost gyroscope is

$$\tilde{\omega} = \omega + \mu_{\omega} \quad (1)$$

where ω denotes the true angle rate, μ_{ω} denotes the additive error. It should be noticed that the earth rotation angle rate is submerged in error μ_{ω} for low-cost gyroscopes [31].

The kinematics of the true system that describe the relationship between attitude R and angle rate ω is

$$\dot{R} = R\omega \times \quad (2)$$

where $(\cdot) \times$ donates the skew-symmetric matrix form of the preceding vector

$$\omega \times = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} \quad (3)$$

Accelerometers provide the measurement of “up” acceleration against the earth’s gravity. The measurement model of the corresponding “vector measurement” [13,29] is

$$\tilde{a} = R^T e_3 + \mu_a \quad (4)$$

where $e_3 = [0 \ 0 \ 1]^T$ is the reference vector of “up” in the navigation frame. $\tilde{a}$ is the normalized measurement vector in the body frame $\{b\}$ from the accelerometer. μ_a is the error term caused by the measurement noise and the potential linear acceleration in which the latter one is small and fast, varying about zero. That is to say, the gravity dominates the value of $\tilde{a}$ for sufficiently low

frequency response. $\mathbf{R}^T = \mathbf{C}_n^b$ is the coordinate transformation matrix that can be used to transform the components of a vector from $\{n\}$ into $\{b\}$.

3.2. Monocular Camera

We assume that the reference frame of the camera $\{c\}$ is aligned with the body frame so that $\mathbf{C}_c^n = \mathbf{C}_b^n = \mathbf{R}$. The z-axis of $\{c\}$ is the optical axis shown in Figure 1. The geometric relationship between the origins of $\{c\}$ and $\{n\}$ and any point P in the 3D world satisfies vector addition $\overrightarrow{O_c P} = \overrightarrow{O_c O_n} + \overrightarrow{O_n P}$. Using this geometric relationship, an observation of the attitude and position of the camera can be achieved as

$$\mathbf{P}^{\{c\}} = \mathbf{t} + \mathbf{R}^T \mathbf{P}^{\{n\}} \quad (5)$$

where $\mathbf{P}^{\{c\}}$ and $\mathbf{P}^{\{n\}}$ are the 3D coordinates of point P expressed in the camera frame and navigation frame. $\mathbf{t}$ is a translation vector. The components of $\mathbf{t}$ are equal to the coordinates of O_n in camera frame $\{c\}$.

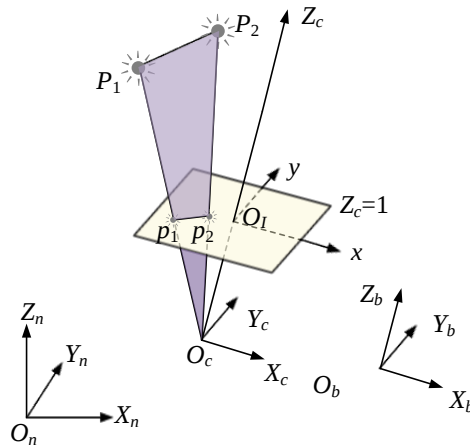


Figure 1. The frames of reference and camera model.

As seen from (5), if $\mathbf{t}$ is a zero vector, the normalized $\mathbf{P}^{\{c\}}$ and $\mathbf{P}^{\{n\}}$ can be used as a vector measurement for attitude estimation. However, this assumption is not always feasible in practice. To avoid restrictions on camera translation, we use two point-correspondences to derive a geometry measurement. No translation vector $\mathbf{t}$ is in the new measurement model.

As shown in Figure 1, P_1 and P_2 are two artificial feature points that can be captured by a camera. Considering that the camera is modeled by a perspective camera using the “frontal projection model” [32], p_1 and p_2 are the projections on normalized image plane $Z_c = 1$.

Let $\mathbf{P}_1^{\{c\}}$ and $\mathbf{P}_2^{\{c\}}$ denote the coordinates of feature points in frame $\{c\}$. Let p_1 and p_2 denote the normalized image coordinate values obtained from the image. A homogeneous model describing the relationship between $\mathbf{p}_i$ and $\mathbf{P}_i^{\{c\}}$ is

$$\mathbf{p}_i = \begin{bmatrix} \frac{P_{i,x}^{\{c\}}}{P_{i,z}^{\{c\}}} & \frac{P_{i,y}^{\{c\}}}{P_{i,z}^{\{c\}}} & 1 \end{bmatrix}^T + \boldsymbol{\mu}_{pi}, \quad i = 1, 2 \quad (6)$$

where $\boldsymbol{\mu}_{pi}$ is the error of the projection model. $P_{i,z}^{\{c\}}$ is the so-called “depth” of a feature point in the camera frame.

The plane formed by the camera center point and two feature points is defined as the “feature plane” in this paper, as shown with light purple color in Figure 1. Let $\tilde{\mathbf{y}}$ denote the computed unit normal vector of the feature plane expressed in $\{c\}$. $\tilde{\mathbf{y}}$ is calculated by

$$\tilde{\mathbf{y}} = \frac{p_1 \times p_2}{\text{norm}(p_1 \times p_2)} \quad (7)$$

Let $\mathbf{d}$ denote the true unit normal vector of the feature plane expressed in $\{n\}$. Let $\mathbf{r}$ denote the unit direction vector of $\overrightarrow{P_1 P_2}$ expressed in $\{n\}$.

The explicit error model of the proposed geometry measurement is

$$\tilde{\mathbf{y}} = \mathbf{R}^T \mathbf{d} + \mu_y \quad (8)$$

and the implicit model of geometry measurement is

$$\tilde{\mathbf{y}}^T \mathbf{R}^T \mathbf{r} = 0 + \mu'_y \quad (9)$$

where μ_y and μ'_y are the measurement errors that come from the errors in p_1 and p_2 .

The positions of artificial feature points in frame $\{n\}$ can be pre-calibrated in the offline stage. Therefore, $\mathbf{r}$ is a known reference vector, and the implicit geometry measurement in (9) can be used for attitude estimation. The explicit geometry measurement model in (8) is not useless in this paper. It appears in the stability analysis of the attitude observer in Section 4.2.

4. Attitude Estimation Algorithm

In this section, a nonlinear complementary filter on the special orthogonal group is introduced. This filter fuses the angle-rate measurement, vector measurement, and implicit geometry measurement to estimate the attitude in continuous dynamics. The attitude estimation algorithm for real-world signals is also considered in this section. Specifically, it includes the initial alignment and the discrete realization of the filter based on the unit quaternion.

4.1. Nonlinear Complementary Filter on SO(3)

The proposed attitude estimation algorithm is based on a nonlinear observer. The goal of the attitude estimation observer is to provide a set of dynamics for an estimated attitude to drive the estimation error converges.

The observer kinematics include a prediction term and a correction term. In this paper, the prediction term is based on the angle rate measurement. The correction term is added to the measured angle rate as it does in classic attitude observer design [18,30].

Let $\hat{\mathbf{R}}$ denote the estimated direction cosine matrix $\mathbf{C}_b^{n'}$. The proposed attitude observer is

$$\dot{\hat{\mathbf{R}}} = \hat{\mathbf{R}}(\tilde{\omega} + \Delta\omega) \times, \quad \hat{\mathbf{R}}(0) = \hat{\mathbf{R}}_0 \quad (10)$$

where the correction term $\Delta\omega$ is

$$\begin{aligned} \Delta\omega &= k_a \Delta\omega_a + k_c \Delta\omega_c \\ \Delta\omega_a &= \tilde{\mathbf{a}} \times \hat{\mathbf{R}}^T \mathbf{e}_3 \\ \Delta\omega_c &= -\tilde{\mathbf{y}}^T \hat{\mathbf{R}}^T \mathbf{r} \cdot \tilde{\mathbf{y}} \times \hat{\mathbf{R}}^T \mathbf{r} \end{aligned} \quad (11)$$

$\Delta\omega_a$ is the correction term caused by the vector measurement. Referring to the explicit version of Mahony's complementary filter [18], the cross product of the estimated vectors $\hat{\mathbf{R}}^T \mathbf{e}_3$ and the observed vectors $\tilde{\mathbf{a}}$ is used to construct the correction term.

$\Delta\omega_c$ is the correction term caused by the geometry measurement. This correction term is referred from the observer design in [30] that fuses angle-rate measurement and geometry measurement.

The rationale for the $\Delta\omega_c$ is the following. From the implicit geometry measurement Equation (9), the ideal $\hat{R}$ satisfies $\hat{R}^T r \in \ker \tilde{y}$. If it is not satisfied, then a corrective angle rate should be applied. The angle rate needed for this is directed along $-\tilde{y} \times \hat{R}^T r$. The magnitude of the correction is simply the $\tilde{y}^T \hat{R}^T r$.

In the above attitude observer, $k_a > 0$ and $k_c > 0$ are two fixed parameters. The stability of the new attitude observer is analyzed in Section 4.2.

As [18] does, we also term the observer as a nonlinear complementary filter here.

4.2. Stability Analysis

The estimation error $\tilde{R}$ is defined as the relative rotation from the true navigation frame $\{n\}$ to the estimated navigation frame $\{n'\}$, that is

$$\tilde{R} := R\hat{R}^T, \quad \tilde{R} = C_{n'}^n \quad (12)$$

Differentiate both sides of estimation error definition Equation (12) and it is straightforward to verify that the error system is

$$\begin{aligned} \dot{\tilde{R}} &= \dot{R}\hat{R}^T + R\dot{\hat{R}}^T \\ &= R\omega \times \hat{R}^T + R(\tilde{\omega} + \Delta\omega) \times^T \hat{R}^T \\ &= R\omega \times \hat{R}^T - R(\tilde{\omega} + \Delta\omega) \times \hat{R}^T \end{aligned} \quad (13)$$

Substitute the angle rate measurement $\tilde{\omega}$ using its true values ω . Using $R\Delta\omega \times = (R\Delta\omega) \times R$, we obtain:

$$\begin{aligned} \dot{\tilde{R}} &= -R\Delta\omega \times \hat{R}^T \\ &= -(R\Delta\omega) \times R\hat{R}^T \\ &= -(R\Delta\omega) \times \tilde{R} \end{aligned} \quad (14)$$

Based on the correction term in (11), the error system is given by

$$\begin{aligned} \dot{\tilde{R}} &= -k_a(R\Delta\omega_a) \times \tilde{R} - k_c(R\Delta\omega_c) \times \tilde{R} \\ &= -k_a(R\tilde{a} \times \hat{R}^T e_3) \times \tilde{R} \\ &\quad + k_c \tilde{y}^T \hat{R}^T r \cdot (R\tilde{y} \times \hat{R}^T r) \times \tilde{R} \end{aligned} \quad (15)$$

Substitute the vector measurement $\tilde{a}$ and geometry measurement $\tilde{y}$ using their true values $R^T e_3$ and $R^T d$. Here, the explicit geometry measurement model is used. It is straightforward that

$$\begin{aligned} \dot{\tilde{R}} &= -k_a \left[R(R^T e_3) \times \hat{R}^T e_3 \right] \times \tilde{R} \\ &\quad + k_c d^T R\hat{R}^T r \cdot \left[R(R^T d) \times \hat{R}^T r \right] \times \tilde{R} \end{aligned} \quad (16)$$

For any vector v , there is $(R^T v) \times = R^T v \times R$. Using this relationship, R is eliminated from the error dynamics. The error system is

$$\begin{aligned} \dot{\tilde{R}} &= -k_a(e_3 \times \tilde{R}e_3) \times \tilde{R} \\ &\quad + k_c d^T \tilde{R}r \cdot (d \times \tilde{R}r) \times \tilde{R} \end{aligned} \quad (17)$$

The error system in (17) is a nonlinear time-varying system because the feature plane's unit normal vector d is varying with the position of the camera.

It is easily verified that the identity matrix I is an equilibrium point of the error system. According to the definition of the small perturbing rotation [32], the attitude error around the equilibrium point $\tilde{R}_{eq}$ is approximated by

$$\tilde{R} \approx (I + \phi \times) \tilde{R}_{eq} \quad (18)$$

where vector ϕ is the angle-axis form of the small perturbing rotation.

The linearization of the error dynamics is computed to analyze the local stability of the equilibrium point [33]. Substitute (18) into (17) with $\tilde{R}_{eq} = I$, we get

$$\begin{aligned} \dot{\phi} \times &\approx -k_a (e_3 \times \phi \times e_3) \times \\ &+ k_c d^T \phi \times r (d \times r) \times \end{aligned} \quad (19)$$

Using the properties of linear operations, including the skew-symmetric matrix transformation and the vector cross product, the state equation describing the evolving of the ϕ is obtained as follows:

$$\begin{aligned} \dot{\phi} &= k_a (e_3 \times) (e_3 \times) \phi \\ &- k_c (d \times r) (d \times r)^T \phi \end{aligned} \quad (20)$$

For $k_a > 0$, $k_c > 0$, the linearized system is asymptotically stable as long as the third component of $d \times r$ is not 0. Otherwise, the error of the yaw angle will not converge to zero. According to the geometric relationship, $d \times r$ is the vector located in the feature plane and perpendicular to the line of two feature points. This geometric structure is determined by the positions of features and camera. Since the linearized error system is asymptotically stable, the nonlinear error system ensures locally asymptotic stability around the equilibrium point I .

The nonlinear complementary filter proposed in this paper is based on observer design. The local asymptotic stability of the filter ensures that the initial attitude estimation error will converge to zero when the gains are any constant greater than zero. Generally speaking, the larger the gain, the faster the error convergence. However, the observer design method is a deterministic method in which the measurement errors are assumed to be zero. The measurement noise in practice will lead to a big attitude estimation variance when the gains are tuned too large inappropriately. Therefore, gains tuning is a compromise process. Moreover, a proper initial attitude guess is important for the filter with local asymptotic stability.

4.3. Initial Alignment

The initial alignment is the problem of attitude determination in the initial static stage using accelerometers and a camera. It is important since the attitude estimation filter in (10) requires a proper initial value.

According to the chain rule, the direction cosine matrix's transpose C_n^b can be written as the multiplication of two rotation matrixes

$$C_n^b = C_h^b C_n^h \quad (21)$$

where $\{h\}$ is an intermediate frame system, referred to as the horizontal frame, whose z-axis coincides with the z-axis of $\{n\}$. Under the definition of the navigation frame and the body frame in this paper, C_h^b and C_n^h can be written in the following form:

$$C_h^b = \begin{bmatrix} \cos \varphi & \sin \theta \sin \varphi & -\cos \theta \sin \varphi \\ 0 & \cos \theta & \sin \theta \\ \sin \varphi & -\sin \theta \cos \varphi & \cos \theta \cos \varphi \end{bmatrix} \quad (22)$$

$$C_n^h = \begin{bmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (23)$$

The $(\theta \ \varphi \ \psi)$ is a set of Euler angles. θ is the pitch angle with $\theta \in [-90^\circ, 90^\circ)$. φ and ψ are the roll angle and yaw angle, where $\varphi, \psi \in [-180^\circ, 180^\circ)$. The yaw angle ψ equals to the angle formed by the y-axis of $\{h\}$ and the y-axis of $\{n\}$.

The initial alignment method proposed in this paper includes horizontal alignment and azimuth alignment and solution confirmation.

First, use normalized accelerometer vector measurement to calculate the horizontal attitude. It is easily verified that the third column of C_h^b is the projection of the normalized “up” vector in the body frame $\{b\}$. Therefore, C_h^b can be recovered roughly from the acceleration measurement by following expressions

$$\begin{aligned} \sin \theta &= \tilde{a}_y, \quad \cos \theta = \sqrt{1 - \tilde{a}_y^2} \\ \sin \varphi &= \begin{cases} \frac{-\tilde{a}_x}{\sqrt{1 - \tilde{a}_y^2}}, & \tilde{a}_y^2 \neq 1 \\ 0, & \tilde{a}_y^2 = 1 \end{cases} \\ \cos \varphi &= \begin{cases} \frac{\tilde{a}_z}{\sqrt{1 - \tilde{a}_y^2}}, & \tilde{a}_y^2 \neq 1 \\ 1, & \tilde{a}_y^2 = 1 \end{cases} \end{aligned}$$

Second, determine the azimuth. After calculating of the C_h^b , use camera implicit geometry constraint

$$\tilde{\mathbf{y}}^T C_h^b C_n^h \mathbf{r} = 0 \quad (24)$$

to construct one linear equation about $\sin \psi$ and $\cos \psi$. With the constraint $\sin^2 \psi + \cos^2 \psi = 1$, two sets of solutions can be found, one of which can be confirmed by the constraints of the camera’s field of view.

Third, confirm the true direction cosine matrix. According to the camera model in (5) and (6), it is straightforward that

$$C_n^b(P_1^{\{n\}} - P_2^{\{n\}}) = z_1 \mathbf{p}_1 - z_2 \mathbf{p}_2 \quad (25)$$

In (25), z_1 and z_2 are the depth of feature points P1 and P2 in the camera frame. Using (25), three linear equations about z_1 and z_2 can be obtained. The C_n^b that leads to positive z_1 and z_2 will be accepted as the initial direction cosine matrix.

4.4. Discrete Implementation on Quaternion

The filter in (10) is a continuous system. In practical implementation, sensor data will be sampled and the filter needs to be integrated in discrete time. The unit quaternion representation of the rotations is commonly used for the realization of algorithms on SO(3) since it offers considerable efficiency in code implementation [18]. The proposed attitude observer in quaternion representation is

$$\dot{\hat{\mathbf{q}}} = \frac{1}{2} \mathbf{\Omega}(\tilde{\boldsymbol{\omega}} + \Delta \boldsymbol{\omega}) \hat{\mathbf{q}}, \quad \hat{\mathbf{q}}(0) = \hat{\mathbf{q}}_0 \quad (26)$$

where

$$\mathbf{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} 0 & -\omega_1 & -\omega_2 & -\omega_3 \\ \omega_1 & 0 & \omega_3 & -\omega_2 \\ \omega_2 & -\omega_3 & 0 & \omega_1 \\ \omega_3 & \omega_2 & -\omega_1 & 0 \end{bmatrix} \quad (27)$$

Since the term $\tilde{\omega} + \Delta\omega$ can be seen as the corrected angle rate in the frame of $\{b\}$, according to the quaternion-based attitude update algorithm [12], the discrete implementation of filter (26) is

$$\hat{q}_k = \hat{q}_{k-1} \circ \Delta q_{\tilde{\omega} + \Delta\omega_k} \quad (28)$$

◦ in (28) is the quaternion multiplication operator. $p \circ q$ is defined by

$$p \circ q = \begin{bmatrix} p_1q_1 - p_2q_2 - p_3q_3 - p_4q_4 \\ p_1q_2 + p_2q_1 + p_3q_4 - p_4q_3 \\ p_1q_3 + p_3q_1 + p_4q_2 - p_2q_4 \\ p_1q_4 + p_4q_1 + p_2q_3 - p_3q_2 \end{bmatrix} \quad (29)$$

$\Delta q_{\tilde{\omega} + \Delta\omega_k}$ in (28) is the quaternion increment from time t_{k-1} to time t_k and is calculated by

$$\Delta q_{\tilde{\omega} + \Delta\omega_k} = \begin{bmatrix} \cos \frac{\Delta\theta_k}{2} \\ \frac{\Delta\theta_k}{\Delta\theta_k} \sin \frac{\Delta\theta_k}{2} \end{bmatrix} \quad (30)$$

$\Delta\theta_k$ in (30) is the angle increment vector, $\Delta\theta_k = |\Delta\theta_k|$ and

$$\Delta\theta_k = (\tilde{\omega}_k + \Delta\omega_k)\Delta t \quad (31)$$

where Δt is the time interval between t_{k-1} and t_k . $\tilde{\omega}_k$ is calculated by the average value of the gyroscope measurements in t_{k-1} and t_k .

$\Delta\omega_k$ in angle increment vector (31) is the angle rate correction term constructed by the estimation in time t_{k-1} and measurements in time t_k as

$$\begin{aligned} \Delta\omega_k &= k_a\Delta\omega_{a,k} + k_c\Delta\omega_{c,k} \\ \Delta\omega_{a,k} &= \tilde{a}_k \times C(\hat{q}_{k-1})e_3 \\ \Delta\omega_{c,k} &= -\tilde{y}_k^T C(\hat{q}_{k-1})r \cdot \tilde{y}_k \times C(\hat{q}_{k-1})r \end{aligned} \quad (32)$$

$C(q)$ is the coordinate transformation matrix and equals to

$$\begin{bmatrix} q_1^2 + q_2^2 - q_3^2 - q_4^2 & 2(q_2q_3 + q_1q_4) & 2(q_2q_4 - q_1q_3) \\ 2(q_2q_3 - q_1q_4) & q_1^2 - q_2^2 + q_3^2 - q_4^2 & 2(q_3q_4 + q_1q_2) \\ 2(q_2q_4 + q_1q_3) & 2(q_3q_4 - q_1q_2) & q_1^2 - q_2^2 - q_3^2 + q_4^2 \end{bmatrix}$$

To get a high-bandwidth system, the sample interval of the gyroscope can be selected as the discretization time interval of the filter. Considering that the measurements from the camera and accelerometer play a role in providing long-term stability, they can be updated at a low frequency. If vector or geometry measurement is not available in current sample time t_k , the corresponding correction term $\Delta\omega_k$ is set to zero, where x is a or c .

This structure of the proposed filter is the so-called “explicit version” of the nonlinear attitude observer. As we discussed in the section “Related works”, Mahony provides three versions of nonlinear attitude observer in his work: direct, passive, and explicit. The direct and passive versions depend on the algebraic reconstruction of the attitude. For MARG sensors, the sampling frequencies of accelerometer and magnetometer are the same. This will not cause any problems. However, for inertial sensors/camera combination, the difference in sampling frequency makes the algebraic reconstruction have to align with the sensor sampled in low-frequency. The consequence is that a lot of information in the high-frequency sensor is lost. On the contrary, the angle rate correction term of the explicit complementary filter in (32) is aligned with the sensor sampled in high-frequency. This is the advantage of the explicit version of the nonlinear attitude observer in handling sensors with different sample frequencies.

5. Evaluation

5.1. Experiment Setup

To evaluate the performance of the proposed attitude estimation algorithm, an experiment system is constructed as shown in Figure 2. A smartphone with built-in inertial sensors and a rear camera is used as the measurement equipment. Two artificial fiducials from AprilTag family Tag36H11 are placed on the ground. The 3DM-GX3-25 Attitude and Heading Reference System (AHRS) attached to the smartphone is used as the ground truth provider.

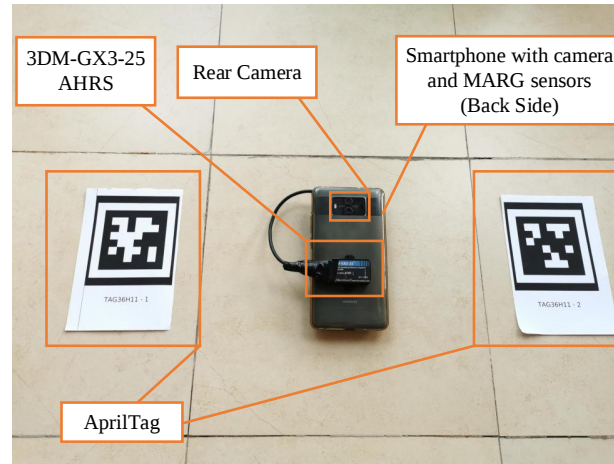


Figure 2. Experiment system for algorithm evaluation.

An Android application is developed to capture the data from inertial sensors and images from the camera. The sampling frequency of inertial sensors and camera are 100 Hz and 5 Hz respectively. All the data are time-stamped and stored in the smartphone's SD card. The recorded data are processed offline on the laptop so that the different algorithms and control variables can be evaluated on the same recorded data.

To keep the AHRS from magnetic disturbance, the experiment is implemented in an outdoor environment and the magnetometer of the 3DM AHRS is re-calibrated after installation. The y-axis of the navigation frame is chosen as the magnetic north.

The intrinsic parameters and distortion parameters of the camera are pre-calibrated using the geometric method proposed in [34]. The process of extracting the normalized image coordinates of the feature points from the image is as follows. Undistort the image according to the distortion parameters. Detect the AprilTag features using the Python package `pupil_apriltags`. Recover the normalized image coordinates of features from the corresponding index coordinates using camera intrinsic parameters.

In Sections 2 and 3, we assume that the axis direction of the camera frame $\{c\}$ is the same as that of the body frame $\{b\}$. However, in our smartphone experiment platform, the optical axis of the camera is opposite to the z-axis of the body frame. Moreover, in data collection software, the image orientation is "Landscape" relative to the smartphone. The actual direction relationship between the body frame $X_b-Y_b-Z_b$, image pixel frame $u-v$, and camera frame $X_c-Y_c-Z_c$ is as shown in Figure 3. Therefore, the normalized image coordinate values should be transformed to adapt to the complementary filter algorithm. The coordinate transformation matrix M is

$$M = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Under the definition of the reference frames in Figure 3, the relationship between the normalized image coordinates of the feature points and their camera coordinates is as follows.

$$\mathbf{p}_i = \begin{bmatrix} \frac{P_{ix}^{(c)}}{-P_{iz}^{(c)}} & \frac{P_{iy}^{(c)}}{-P_{iz}^{(c)}} & \frac{P_{iz}^{(c)}}{-P_{iz}^{(c)}} \end{bmatrix}^T + \boldsymbol{\mu}_{pi}, \quad i = 1, 2 \quad (33)$$

This makes the third step of the initial alignment a little bit different from Section 4.3. Specifically, the “depth equations” in (25) should be replaced by (34). The attitude solutions that lead to negative z_1 and z_2 will be accepted as the initial attitude.

$$C_n^b(\mathbf{P}_1^{\{n\}} - \mathbf{P}_2^{\{n\}}) = -z_1 \mathbf{p}_1 + z_2 \mathbf{p}_2 \quad (34)$$

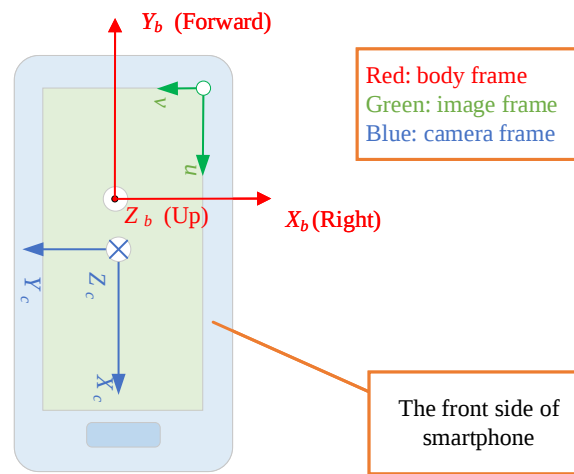


Figure 3. Direction relationship between the body frame image pixel frame and camera frame in experiment system.

Two sets of measurement data are collected to evaluate the static and dynamic performance of the proposed attitude estimation algorithm.

- (1) In the static case, the smartphone keeps static in each fixed attitude about 15 s. Moreover, the smartphone in this case only makes the rotational motion between each fixed attitude without displacement relative to the navigation frame.
- (2) In the dynamic case, the smartphone makes arbitrary rotation and translation movements while ensuring that the two visual labels are always within the camera’s field of view.

Representative algorithms are implemented to process the collected data for performance comparison.

- (1) SINS: Attitude update algorithm of strap-down inertial navigation in [12].
- (2) VIN-EKF: This is a vision-aided inertial navigation algorithm based on extended Kalman filter (EKF). The system state vector of the filter includes unit quaternion, velocity, position, gyroscope and accelerometer measurement bias. The measurement residual is computed by the measured normalized image coordinates of the feature points with the predicted normalized image coordinates in the filter. The linearized system model for the IMU error-state and the linearized measurement model about the estimates for the attitude and position of the camera are as described in [27]. The standard deviations of sensor noise are $\sigma_{gyro} = 0.04^\circ/\text{s}$, $\sigma_{acc} = 1 \text{ mg}$, and $\sigma_{camera} = 7/f_c$, where f_c is the camera focal length comes from the calibrated intrinsic parameters. Fusing camera and inertial sensors in a tightly coupled scheme is the core idea of [10]

so that the vision-based visible light communication positioning system is able to continually provide location service when the number of the feature points is less than four.

- (3) Proposed CF: This algorithm fuses inertial sensors and a camera using the quaternion-based discrete implementation in (28) and (32). Two gain parameters of the proposed nonlinear complementary filter are $k_a = 0.6$ and $k_c = 0.8$.
- (4) CF-1: This algorithm also fuses inertial sensors and a camera using nonlinear complementary filter. The main difference between CF-1 and the proposed CF is that in CF-1, the measurement models of accelerometer and camera are all the vector measurements. The reference vectors are the normalized gravity vector and the normalized image coordinates of feature points in the initial frame [29]. Therefore, this is a relative attitude estimation method. To get the full attitude, the initial attitude should be known. This filter is implemented as a quaternion filter in (28), but the correction terms are all constructed using cross product of the estimated vectors and the observed vectors. There are three parameters for CF-1 when two feature points are captured in one image, that is, $k_a = 0.6$, $k_{c1} = 0.8$ and $k_{c2} = 0.8$.
- (5) CF-2: This algorithm fuses the gyroscope with a camera using nonlinear complementary filter. Same as the proposed CF, CF-2 also uses the camera measurements as the implicit geometry measurement. However, there is no gravity measurement in CF-2 [30]. This filter is implemented as a quaternion filter in (28), but the accelerometer correction term is zero. There is only one parameter for complementary filter when two feature points are captured in one image, that is, $k_c = 0.8$.

The gains of the proposed CF are tuned to achieve relatively good attitude estimation results in our experimental conditions. To be fair, the gains of the other two complementary filter algorithms are set to the same constant as the proposed CF. The parameters of VINS-EKF are chosen according to the sensor measurement noise characteristics.

5.2. Result and Discussion

Figure 4 shows the attitude ground truth from 3DM AHRS, the attitude updated from the gyroscope, and the attitude estimation results from four estimation algorithms, in the static case. The attitude errors with respect to reference angles from AHRS are shown in Figure 5. To further verify the performances, Table 1 gives the root-mean-squared errors (RMSEs) of various estimation algorithms.

As can be seen from Figure 5 and Table 1, in the static case, the accuracy of the proposed algorithm is as good as that of VIN-EKF. For the CF-1, the estimation error of pitch and roll is close to that of the proposed filter, but the yaw angle error is the maximum among all the algorithms. This is due to the fact that the CF-1 algorithm corrects the gyroscope prediction error resort to vector measurement and the reference vectors of two feature points are near the gravity vector during the experiment. The result of CF-2 is just the opposite. The yaw angle error is close to the proposed algorithm, but the pitch and roll errors are almost equal to the errors of the SINS algorithm. The reason is that artificial vision fiducials are arranged on the ground, which leads to a horizontal reference vector of r in the implicit geometry measurement. Since no gravity vector or other reference vectors with vertical components are fused in the filter, the pitch and roll errors cannot converge.

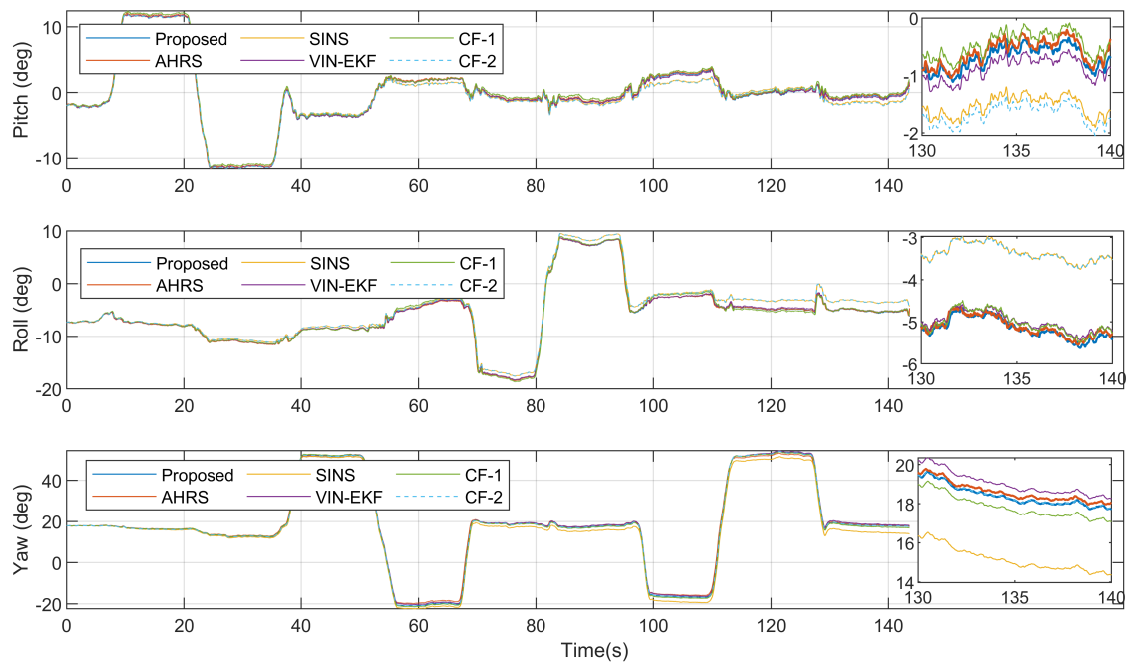


Figure 4. Attitude estimation results in the static case.

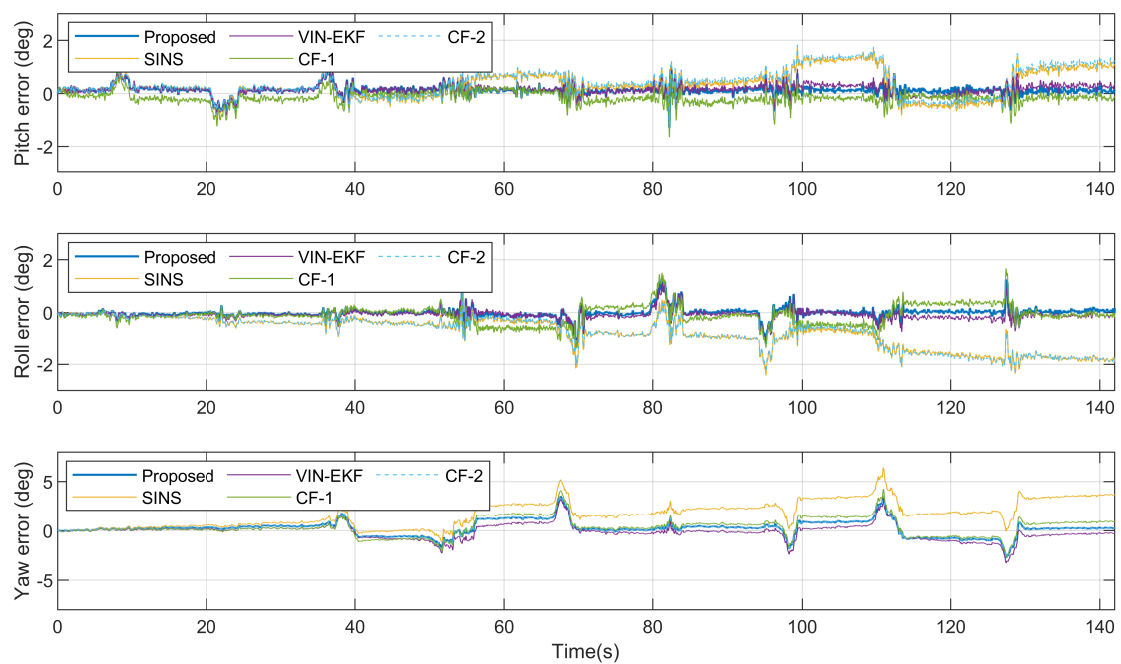


Figure 5. Attitude estimation error in the static case.

Table 1. RMSE of attitude angles in the static case.

Algorithm	Pitch	Roll	Yaw
CF-1	0.2701°	0.3575°	0.9912°
CF-2	0.6538°	0.9769°	0.7974°
Proposed CF	0.2195°	0.2008°	0.7977°
VIN-EKF	0.2556°	0.2160°	0.7805°

Figure 6 shows the attitude ground truth from 3DM AHRS, the attitude updated from the gyroscope, and the attitude estimation results from four estimation algorithms, in the dynamic case. The attitude errors with respect to reference angles from AHRS are shown in Figure 7. Table 2 gives the root-mean-squared errors (RMSEs) of various estimation algorithms.

Table 2. RMSE of attitude angles in the dynamic case.

Algorithm	Pitch	Roll	Yaw
CF-1	1.3112°	1.0108°	2.1121°
CF-2	0.6727°	1.2696°	1.7517°
Proposed CF	0.2906°	0.3071°	1.6495°
VIN-EKF	0.5077°	0.5211°	2.1093°

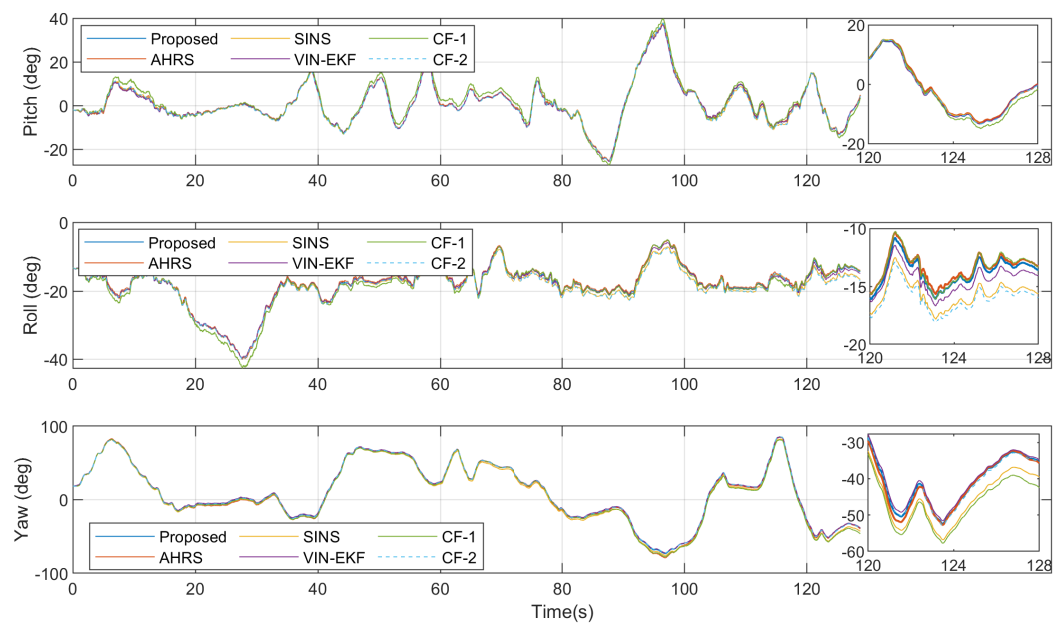


Figure 6. Attitude estimation results in the dynamic case.

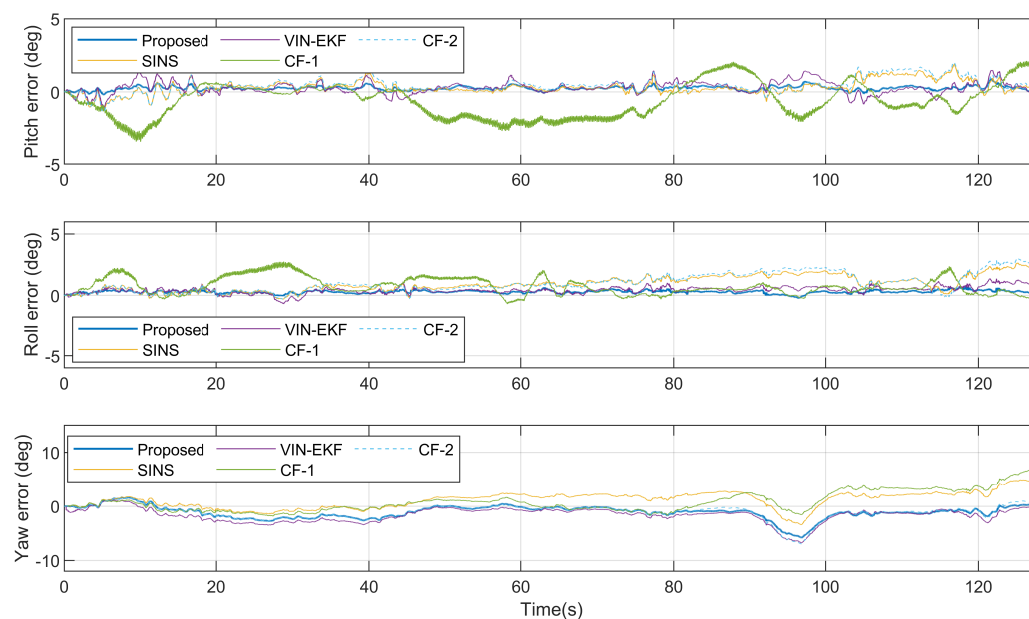


Figure 7. Attitude estimation error in the dynamic case.

As can be seen from Figure 7 and Table 2, in the dynamic case, due to the translation motion of the smartphone, the accuracy of the VIN-EKF and CF-1 is significantly reduced. The translation motion even increases the attitude error of SINS in the early stage. In the dynamic case, the proposed algorithm offers the best performance in estimation accuracy.

For the CF-1 algorithm, the translation motion increases the error of the camera vector measurement model. The estimation accuracy of pitch and roll angle is obviously affected since the reference vector is near the vertical line. Different from the CF-1, the proposed nonlinear complementary filter fuses the implicit geometry measurement so that the position of the camera is decoupled with the attitude estimation.

When it comes to the VIN-EKF, another reason for the performance degradation that must be mentioned is the update rate of the filter. The filter propagates the state in 100 Hz and performs update in 5 Hz. Although the gravity vector constraint is implicit in the velocity equation of the system model, the innovation of this constraint becomes available only when a new image comes and the filter is updated. The estimation error of pitch and roll angle using our method is less than VIN-EKF since the gravity vector constraint is an explicit measurement and the corresponded error is corrected in 100 Hz.

Figure 8 shows the instantaneous magnitude of the accelerometer output during the static case and dynamic case.

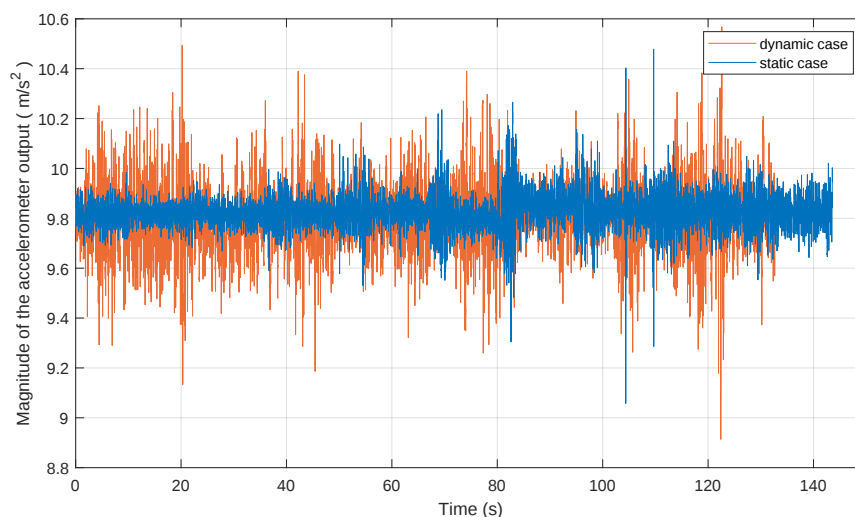


Figure 8. Magnitude of the accelerometer output.

The mean time consumption and related standard deviation of different algorithms are presented in Table 3. Here, time consumption refers to the execution time between two consecutive camera sample updates. The mean values and the standard deviation are calculated after 700 image sample updates. It can be seen that three nonlinear complementary filters show comparable time consumption. The execution time of CF-2 is the smallest since there is no accelerometer correction. The execution time of CF-2 is less than the proposed CF which means the update of vector measurement is simpler than the update of implicit geometry measurement. The execution time of VIN-EKF is the largest due to the calculation of the gain matrix and covariance matrix.

Table 3. Mean and standard deviation of time consumption of various algorithms.

Algorithm	Mean Time	STD
CF-2	1.1162×10^{-4} s	3.7175×10^{-5} s
CF-1	1.4818×10^{-4} s	2.8018×10^{-5} s
Proposed CF	1.7010×10^{-4} s	4.3637×10^{-5} s
VIN-EKF	5.4172×10^{-4} s	8.3306×10^{-5} s

6. Conclusions

In this paper, we consider the attitude estimation problem of fusing camera and inertial sensors in an environment with only two pre-calibrated artificial vision fiducials. The main contributions of this paper are twofold. First, we derive an implicit geometry measurement for camera-based attitude estimation. Second, a nonlinear complementary filter with only two parameters to be adjusted is proposed to fuse angle rate measurement, vector measurement, and geometry measurement.

Compared to MARG sensor-based attitude estimation, the method in this paper doesn't rely on a magnetometer. It provides an attractive solution for the environment with complex magnetic field distribution. Compared to camera- and gyroscope-based attitude estimation, our method still converges when there are only two feature points in the field of view. The experimental results show that our algorithm outperforms all the compared counterparts in estimated accuracy and provides competitive computational complexity.

We believe that the proposed method can potentially benefit related navigation applications. The main drawback of our method is that the designed nonlinear complementary filter only ensures locally asymptotic stability. Future work will focus on system improvements to achieve unrestricted local stability and even global stability.

Author Contributions: Conceptualization and methodology, L.Z. and X.Z. (Xingqun Zhan); software and validation, L.Z. and X.Z. (Xin Zhang); writing—original draft preparation, L.Z.; writing—review and editing, L.Z., X.Z. (Xingqun Zhan) and X.Z. (Xin Zhang). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Key R&D Program Projects in Jiangxi Province under Grant 20181ACE50027 and 20193ABC03A006.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Pierleoni, P.; Belli, A.; Palma, L.; Pellegrini, M.; Pernini, L.; Valenti, S. A High Reliability Wearable Device for Elderly Fall Detection. *IEEE Sens. J.* **2015**, *15*, 4544–4553. [\[CrossRef\]](#)
- Zheng, L.; Zhan, X.; Zhang, X.; Wang, S.; Yuan, W. Heading Estimation for Multimode Pedestrian Dead Reckoning. *IEEE Sens. J.* **2020**, *20*, 8731–8739. [\[CrossRef\]](#)
- Tian, Y.; Wei, H.; Tan, J. An Adaptive-Gain Complementary Filter for Real-Time Human Motion Tracking with MARG Sensors in Free-Living Environments. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2013**, *21*, 254–264. [\[CrossRef\]](#) [\[PubMed\]](#)
- Marantos, P.; Koveos, Y.; Kyriakopoulos, K.J. UAV State Estimation Using Adaptive Complementary Filters. *IEEE Trans. Control Syst. Technol.* **2016**, *24*, 1214–1226. [\[CrossRef\]](#)
- Königseder, F.; Kemmetmüller, W.; Kugi, A. Attitude Estimation Using Redundant Inertial Measurement Units for the Control of a Camera Stabilization Platform. *IEEE Trans. Control Syst. Technol.* **2016**, *24*, 1837–1844. [\[CrossRef\]](#)
- Zhongguo, S.; Jinsheng, Z.; Xuehui, Z.; Xiaoli, X. A Calibration Method of Three-Axis Magnetometer with Noise Suppression. *IEEE Trans. Magn.* **2014**, *50*, 1–4. [\[CrossRef\]](#)
- Wu, Y.; Hu, Z. PnP Problem Revisited. *J. Math. Imaging Vision* **2006**, *24*, 131–141. [\[CrossRef\]](#)
- Lepetit, V.; Morenonoguer, F.; Fua, P. EPnP: An Accurate O(n) Solution to the PnP Problem. *Int. J. Comput. Vision* **2009**, *81*, 155–166. [\[CrossRef\]](#)
- Wang, J.; Olson, E. AprilTag 2: Efficient and robust fiducial detection. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016; pp. 4193–4198. [\[CrossRef\]](#)
- Qin, C.; Zhan, X.Q. VLIP: Tightly Coupled Visible-Light/Inertial Positioning System to Cope With Intermittent Outage. *IEEE Photonics Technol. Lett.* **2019**, *31*, 129–132. [\[CrossRef\]](#)
- Chow, C.W.; Chen, C.Y.; Chen, S.H. Enhancement of Signal Performance in LED Visible Light Communications Using Mobile Phone Camera. *IEEE Photonics J.* **2015**, *7*, 1–7. [\[CrossRef\]](#)
- Savage, P.G. Strapdown inertial navigation integration algorithm design part 1: Attitude algorithms. *J. Guidance Control Dyn. Syst.* **1998**, *21*, 19–28. [\[CrossRef\]](#)

13. Yun, X.; Bachmann, E.R.; McGhee, R.B. A Simplified Quaternion-Based Algorithm for Orientation Estimation From Earth Gravity and Magnetic Field Measurements. *IEEE Trans. Instrum. Meas.* **2008**, *57*, 638–650. [\[CrossRef\]](#)
14. Shuster, M.D.; Oh, S.D. Three-axis attitude determination from vector observations. *J. Guidance Control Dyn. Syst.* **1981**, *4*, 70–77. [\[CrossRef\]](#)
15. Markley, F.L. Attitude determination using vector observations and the singular value decomposition. *J. Astronaut. Sci.* **1988**, *36*, 245–258.
16. Sabatini, A.M. Quaternion-based extended Kalman filter for determining orientation by inertial and magnetic sensing. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 1346–1356. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Marins, J.L.; Xiaoping, Y.; Bachmann, E.R.; McGhee, R.B.; Zyda, M.J. An extended Kalman filter for quaternion-based orientation estimation using MARG sensors. In Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems, Maui, HI, USA, 29 October–3 November 2001; pp. 2003–2011. [\[CrossRef\]](#)
18. Mahony, R.; Hamel, T.; Pflimlin, J.M. Nonlinear complementary filters on the special orthogonal group. *IEEE Trans. Autom. Control* **2008**, *53*, 1203–1218. [\[CrossRef\]](#)
19. Fourati, H.; Manamanni, N.; Afilal, L.; Handrich, Y. A Nonlinear Filtering Approach for the Attitude and Dynamic Body Acceleration Estimation Based on Inertial and Magnetic Sensors: Bio-Logging Application. *IEEE Sens. J.* **2011**, *11*, 233–244. [\[CrossRef\]](#)
20. Guerrero-Castellanos, J.F.; Madrigal-Sastre, H.; Durand, S.; Torres, L.; Muñoz-Hernández, G.A. A robust nonlinear observer for real-time attitude estimation using low-cost MEMS inertial sensors. *Sensors* **2013**, *13*, 15138–15158. [\[CrossRef\]](#)
21. Madgwick, S.O.H.; Harrison, A.J.L.; Vaidyanathan, R. Estimation of IMU and MARG orientation using a gradient descent algorithm. In Proceedings of the 2011 IEEE International Conference on Rehabilitation Robotics (Icorr), Zurich, Switzerland, 29 June–1 July 2011; pp. 1–7.
22. Wu, J.; Zhou, Z.B.; Chen, J.J.; Fourati, H.; Li, R. Fast Complementary Filter for Attitude Estimation Using Low-Cost MARG Sensors. *IEEE Sens. J.* **2016**, *16*, 6997–7007. [\[CrossRef\]](#)
23. Wilson, S.; Eberle, H.; Hayashi, Y.; Madgwick, S.O.H.; McGregor, A.; Jing, X.; Vaidyanathan, R. Formulation of a new gradient descent MARG orientation algorithm: Case study on robot teleoperation. *Mech. Syst. Sig. Process.* **2019**, *130*, 183–200. [\[CrossRef\]](#)
24. Yoo, T.S.; Hong, S.K.; Yoon, H.M.; Park, S. Gain-scheduled complementary filter design for a MEMS based attitude and heading reference system. *Sensors* **2011**, *11*, 3816–3830. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Hartley, R.I. In defense of the eight-point algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 580–593. [\[CrossRef\]](#)
26. Nister, D. An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 756–770. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Mourikis, A.I.; Roumeliotis, S.I. A multi-state constraint Kalman filter for vision-aided inertial navigation. In Proceedings of the 2007 IEEE International Conference on Robotics and Automation, Roma, Italy, 10–14 April 2007; p. 3565. [\[CrossRef\]](#)
28. Qin, T.; Li, P.; Shen, S. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. *IEEE Trans. Rob.* **2018**, *34*, 1004–1020. [\[CrossRef\]](#)
29. Wu, J.; Zhou, Z.B.; Fourati, H.; Li, R.; Liu, M. Generalized Linear Quaternion Complementary Filter for Attitude Estimation From Multisensor Observations: An Optimization Approach. *IEEE Trans. Autom. Sci. Eng.* **2019**, *16*, 1330–1343. [\[CrossRef\]](#)
30. Rehbindler, H.; Ghosh, B.K. Pose estimation using line-based dynamic vision and inertial sensors. *IEEE Trans. Autom. Control* **2003**, *48*, 186–199. [\[CrossRef\]](#)
31. Zhang, P.; Zhan, X.; Zhang, X.; Zheng, L. Error characteristics analysis and calibration testing for MEMS IMU gyroscope. *Aerosp. Syst.* **2019**, *2*, 97–104. [\[CrossRef\]](#)
32. Barfoot, T.D. *State Estimation for Robotics: A Matrix Lie Group Approach*; Cambridge University Press: Cambridge, UK, 2016; pp. 199, 242.

33. Khalil, H.K. *Nonlinear Systems*; Prentice Hall: Upper Saddle River, NJ, USA, 2002; pp. 156–162.
34. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).