# Quaternion-Based Iterative Extended Kalman Filter for Sensor Fusion of Vision Sensor and IMU in 6-DOF Displacement Monitoring

Haemin Jeon, Jiyoung Min, Hyuntae Bang, and Wonkeun Youn

Abstract—As civil structures are exposed to various external loads, their periodic evaluation is paramount to ensure their safety. By estimating the 6-degree-of-freedom (DOF) displacement of structures, structural behavior can be monitored directly. Therefore, this study aims to develop a translational and rotational displacement estimation method by fusing a vision sensor and inertial measurement unit (IMU) using a quaternion-based iterative extended Kalman filter (QIEKF). The QIEKF algorithm was applied to reduce the nonlinear influence on the measurement model. The 6-DOF displacement is predicted using the integral of the gyroscope output and updated via a combination of an accelerometer and a magnetometer through a vector matching process in the Kalman filter framework. Subsequently, the 6-DOF displacement estimation result is updated through a vision sensor using a 2-D planar marker and homography transformation in the Kalman filter framework. The performance of the proposed sensor fusion method was verified with experiments using a motorized motion stage, and the results show that the displacements can be estimated with high accuracy regardless of measurement noise and slowly varying signal drift.

Index Terms—6-degree-of-freedom (DOF) displacement, inertial measurement unit (IMU), quaternion-based iterative extended Kalman filter (QIEKF), sensor fusion, vision sensor.
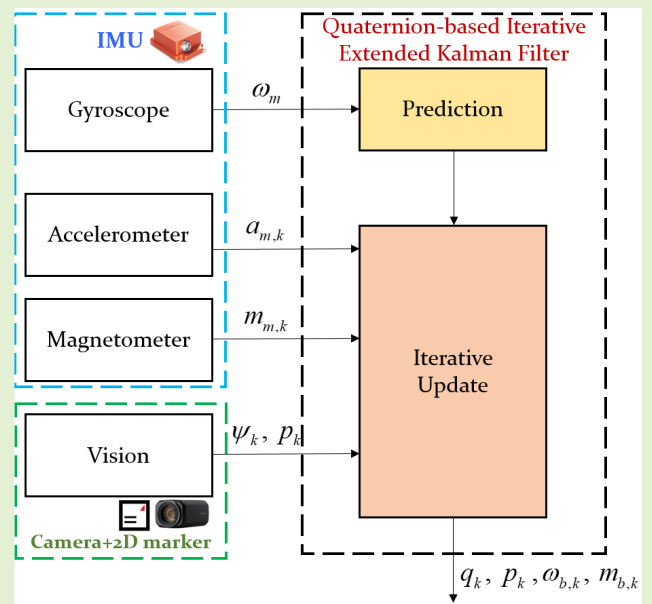
I. INTRODUCTION

AS STRUCTURES age, the demand for effective and reliable monitoring methods has increased. Periodic

Haemin Jeon is with the Department of Civil and Environmental Engineering, Hanbat National University, Yuseong-gu, Daejeon 34158, Republic of Korea (e-mail: hjeon@hanbat.ac.kr).
Jiyoung Min is with the Korea Institute of Civil Engineering and Building Technology, Gyeonggi-do 10233, Republic of Korea (e-mail: atome83@kict.re.kr).
Hyuntae Bang and Wonkeun Youn are with the Department of Autonomous Vehicle System Engineering, Chungnam National University, Yuseong-gu, Daejeon 34134, Republic of Korea (e-mail: htbang@g.cnu.ac.kr; wkyoun@cnu.ac.kr).

evaluation of civil structures exposed to various excessive dynamic loads, such as earthquakes or strong winds, is paramount to ensure their safety. To monitor structural displacements that can directly estimate structural behavior, various sensors, such as accelerometers, global positioning system (GPS), linear variable differential transformers (LVDTs), and laser Doppler vibrometers (LDVs), have been widely studied. However, many of the aforementioned sensors measure the displacement indirectly, are difficult to install, and are expensive [1]. An accelerometer, one of the most widely used sensors for civil engineering applications, has difficulty accurately measuring using low-frequency dynamics [2]. Some studies combine accelerometers and displacement measurement units to achieve reliable structural behavior data, but in these cases, prior knowledge such as the natural frequency of the structure is needed [3], [4]. GPS sensors measure displacement with high accuracy but the accuracy is degraded in shadowed areas [5]. Additionally, LVDTs and LDVs require a fixed reference point for displacement measurements.

To solve the aforementioned problems, vision sensors that provide high accuracy at a low cost have been studied [6],

[7], [8], [9], [10], [11], [12]. With the development of hardware and digital image processing technology, various studies related to structural displacement measurement using vision sensors have been conducted. Image processing techniques, such as feature point extraction at subpixel accuracy, digital image correlation, and optical-flow-based tracking algorithms, have been used to estimate structural displacements [6], [7], [8], [9]. By calculating the positions of artificial or natural feature points on structure, the movement and translational displacement of the structure in the vertical or horizontal direction can be estimated in most cases. To estimate 3-degree-of-freedom (DOF) translational and rotational displacements, a homography-based displacement measurement method was proposed [2], [10]. By calculating the positions of more than four feature points in an obtained image, the relative 6-DOF displacement between the camera and a target point can be estimated. Vision sensor-based displacement measurement methods are computationally efficient and provide high accuracy. However, these methods have the following drawbacks: 1) the displacement is properly measured only if the sensors are installed in the line of sight, i.e., it is easily affected by changes in environmental conditions such as illumination or weather [6], [10] and 2) the displacement resolution and sampling speed of the vision sensor are relatively low [4], [11], [12].

To overcome the aforementioned limitations, data fusion with acceleration measurement has been conducted [13], [14], [15]. In [13], complementary filters and a time synchronization method were proposed to integrate two different sources of estimated displacements with different sampling frequencies. Wu et al. [14] developed a displacement monitoring method combined with a consumer-grade camera and accelerometers. To improve the accuracy of the displacement measurement, the camera vibrations were calculated using the reconstructed displacements from an accelerometer attached on the same side, and then eliminated via subtraction. Park et al. [15] proposed an adaptive multirate Kalman filter for fusing low vision sampling rate and high acceleration sampling rate measurements. These techniques estimate pseudostatic translational displacement from the vision sensor and dynamic translational displacement from the accelerometer. To improve the accuracy of displacements and measure both translational and rotational displacements, this article proposes a quaternion-based iterative extended Kalman filter (QIEKF) for fusing an inertial measurement unit (IMU) and a vision sensor that can estimate structural displacement with high accuracy. To solve the gimbal lock issue in the measurement of rotational displacement, the rotation is represented using quaternions.

The 6-DOF displacement between two points is predicted using the vision sensor, and the translational and rotational displacements are updated using the IMU. Through sensor fusion, the IMU drift error and the loss of accuracy when the line of sight of the vision sensor is not guaranteed are calculated. Euler angles provide us with an angle of rotation around each axis and are generally much more user-friendly and have more intuitive and predictable values but have disadvantages such as gimbal locking. As a way to solve gimbal locking [16], quaternions are arguably an appropriate choice to represent object rotation. They are simple and efficient for interpolating and unambiguously representing rotation. Among the various estimation filter algorithms, an extended Kalman filter (EKF) is the most widely used approach for processing nonlinear system state estimation [17]. To approximate the mean and covariance of a nonlinear system, an EKF uses a first-order Taylor extension to linearize a nonlinear model. If the model has large nonlinearities and large initial state estimation errors, there can be large errors in the true posterior mean and covariance, which can lead to suboptimal performance and sometimes filter divergence [18]. To overcome the aforementioned issues, we implemented an iterative EKF to improve the linearization performance of the EKF by iteratively correcting the linearization point of the first-order Taylor extension, thereby reducing the linearization error [19], [20]. The proposed system combines vision sensor-based displacement and an IMU capable of low- and high-frequency displacement measurement, respectively, using a QIEKF.

To verify the performance of the proposed QIEKF-based sensor fusion method, an experimental test was performed using a motorized motion stage capable of artificial translational and rotational motions. Regarding the behavior of increasingly deteriorating port structures, the vertical settlement and rotational movement in the yaw direction were simulated [21]. The results show that robust and high-precision 6-DOF displacement estimation can be achieved for outliers caused by feature point position errors in a captured image or temporary loss of the communication link in the vision system. Additionally, drift errors in the IMU were mitigated as a result of sensor fusion. The remainder of this article is organized as follows. In Section II, a vision-based displacement measurement method and sensitivity analysis of a proposed planar marker are described. The quaternion-based displacement estimation using an IMU and the QIEKF-based displacement estimation algorithm are described in Section III. To validate the performance of the proposed algorithm, experimental tests were conducted, and the results are discussed in Section IV. Conclusions and further research directions are discussed in Section V.

II. HOMOGRAPHY-BASED DISPLACEMENT ESTIMATION
A. 6-DOF Displacement Estimation Using a Vision Sensor

The relative displacement between the vision sensor and the target structure can be estimated using four or more feature points and the intrinsic parameters composed of the optical properties of the camera lens [21], [22], [23]. To find the translational and rotational displacements, the geometric relationship between two images of the same planar surface, called a homography matrix, needs to be estimated. To compute the planar homography, camera calibration and undistortion of the captured images are performed using the following equations. In (1), r_{ij} ($i = 1, 2, 3; j = 1, 2, 3$) and t_k ($k = x, y, z$) are components of the rotational displacement matrix and the translational displacement vector, respectively. $\mathbf{Q}^i = [X, Y, Z, 1]^T$ and $\mathbf{q}^i = [u, v, 1]^T$ are object points in the global frame represented in the three-dimensional coordinate system and correspondence points captured on the 2-D
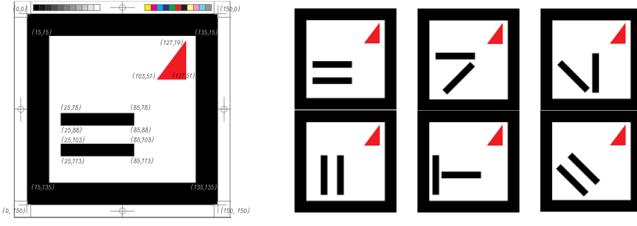
Fig. 1. Dimensions of planar markers and examples of internal rectangular arrangement (unit: mm).

image frame, respectively. The image point coordinates after undistortion are represented in (2) and (3). In (3), as shown at the bottom of the next page, $r^2 = x_c^2 + y_c^2$ and K_i are radial and tangential distortion coefficients, respectively. In (4), f_u and f_v are the focal lengths; c_u and c_v are the principle points; and u and v are the corresponding coordinates of the feature point on the image plane

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = \begin{bmatrix} X_c/Z_c \\ Y_c/Z_c \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix}. \tag{4}$$

After the calculation of the camera intrinsic parameters using checker boards, a homography matrix, **H**, that maps the points between two images of the same planar surface taken from different camera locations and orientations can be estimated. The relationship between the corresponding pair of points, $\mathbf{q}_1^i$ and $\mathbf{q}_2^i$, projected on the image plane of I_1 and I_2 is as follows:

$$\mathbf{q}_1^{i\,T} \mathbf{H} \mathbf{q}_2^i = 0. \tag{5}$$

With the 3×3 homography matrix, $\mathbf{H}_{3\times3}$, more than four corresponding pairs of points ($i \geq 4$) are required to create a matrix with eight unknowns.

B. Design of Planar Marker and Sensitivity Analysis of 6-DOF Displacement Estimation

As described in Section II-A, for homography-based 6-DOF displacement measurement, at least four feature points are required to construct homography matrix, **H**. In this study, a 2-D planar marker was designed for 6-DOF displacement estimation with two rectangles on the lower left and a triangle on the upper right, as shown in Fig. 1. The inner rectangles are arranged vertically, diagonally, or horizontally so that multiple markers can be generated and applied to various positions on the structures. The 6-DOF displacements are measured based on a total of 16 feature points, i.e., the corner points of two pairs of rectangular, inner and outer squares. For stable displacement measurement, only the corner points of the black shapes with a high contrast ratio as shown in Fig. 2(b)
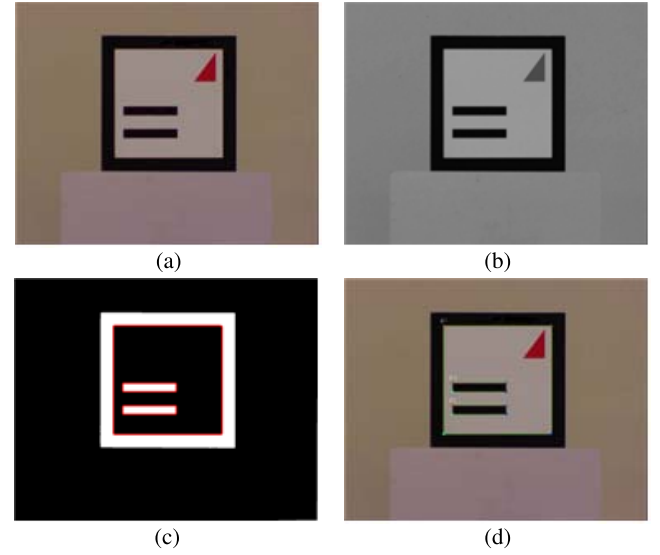

Fig. 2. Image processing process for 6-DOF displacement estimation. (a) Undistorted image. (b) Red channel image. (c) Contour detection. (d) Corner detection.

were used. To avoid copyright issues, the marker has been redesigned in consideration of possible application to actual port structures.

The marker images are captured with a distant camera, and the process for feature point detection using various image processing techniques is shown in Fig. 2. The camera lens distortion is corrected via the previously calculated intrinsic parameters using checker boards. From the undistorted image, the corner points of the rectangles are detected using various image processing techniques, such as RGB channel separation, image binarization, and contour detection. By calculating more than four feature points on a planar marker, the 6-DOF displacement between the camera and the marker can be estimated using homography transformation equations.

To numerically verify the accuracy of the homography-based displacement measurement system using the designed planar marker, sensitivity analysis with Monte Carlo simulations was performed as follows [10]: 1) 6-DOF displacements, $\mathbf{x} = [x, y, z, \theta, \phi, \psi]$, were established considering the translational and rotational movement; 2) the measurement data, the positions of the corners $\mathbf{q}^i$, ($i = 1, \ldots, n_c$), where n_c is the number of corner points, were calculated with $\mathbf{x}$ and the intrinsic camera parameters given a priori; 3) measurement noises were added to $\mathbf{q}^i$ considering measurement errors during image processing; and 4) the displacements were calculated using (1)–(4) [10]. The simulation for sensitivity analysis of the homography-based displacement measurement method is described in the pseudocode of Algorithm 1. To reflect various errors that occur during image processing, uniform random noise of ± 0.2 pixels considering measurement variance with the sensor used in this study was added.

The simulation was performed using 100 cases with a distance between the marker and camera of 10 m. For the estimation, the Newton–Raphson (NR) method was used to solve the nonlinear problem as described in the algorithm. The number of NR iterations was set to 20, and the intrin-

---

Algorithm 1 Pseudocode for the Simulation of Homography-Based Displacement Measurements

---

for $i = 1, \ldots, n_c$ **do** $\triangleright$ Generation of measurement equation: $\mathbf{q_p}^i$

$$^gT_c = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & c\theta & -s\theta & 0 \\ 0 & s\theta & c\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} c\phi & 0 & s\phi & 0 \\ 0 & 1 & 0 & 0 \\ -s\phi & 0 & c\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} c\psi & -s\psi & 0 & 0 \\ s\psi & c\psi & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & x \\ 0 & 1 & 0 & y \\ 0 & 0 & 1 & z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

        $\triangleright$ $c\theta$ and $s\theta$ are $\cos\theta$ and $\sin\theta$, respectively

$\mathbf{Q_c}^i = {}^gT_c \cdot \mathbf{Q}^i$ $\triangleright$ 3-D global coord. to 3-D camera coord. See eq. (1)

$\mathbf{q}^i = \begin{bmatrix} x_c \\ y_c \end{bmatrix} = \begin{bmatrix} \mathbf{Q_c}^i(1)/\mathbf{Q_c}^i(3) \\ \mathbf{Q_c}^i(2)/\mathbf{Q_c}^i(3) \end{bmatrix}$ $\triangleright$ 3-D camera coord. to 2-D image plane coord. See eq. (2)

$r = \sqrt{x_c^2 + y_c^2}$

$\begin{bmatrix} x_d \\ y_d \end{bmatrix} = \begin{bmatrix} x_c(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2K_3 x_c y_c + K_4(r^2 + 2x_c^2) \\ y_c(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + K_3(r^2 + 2y_c^2) + 2K_4 x_c y_c \end{bmatrix}$

$\begin{bmatrix} u^i \\ v^i \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix}, \ \mathbf{q_p}^i = \begin{bmatrix} u^i \\ v^i \end{bmatrix}$ $\triangleright$ 2-D image plane coord. to 2-D pixel coord. See eqs. (3)-(4)

end for

$J_q = \begin{bmatrix} \frac{\partial q_p}{\partial x}, & \frac{\partial q_p}{\partial y}, & \frac{\partial q_p}{\partial z}, & \frac{\partial q_p}{\partial \theta}, & \frac{\partial q_p}{\partial \phi}, & \frac{\partial q_p}{\partial \psi} \end{bmatrix}$ $\triangleright$ Generation of the Jacobian matrix of q_p: J_q

for $t = 1, .., N_s$ **do** $\triangleright$ Generation of 6-DOF disp. with translational movement and estimated using NR

$\bar{\mathbf{x}} = \begin{bmatrix} 5 & 5\sin(\frac{2\pi}{N_s})t & H & 1 & 5 & -1 \end{bmatrix}$; $\triangleright$ N_s: number of samples for MCS, units are mm and degrees

$\hat{\mathbf{x}}^0 = \begin{bmatrix} 0 & 0 & H & 0 & 0 & 0 \end{bmatrix}$;

$\bar{\mathbf{q}} = \mathbf{q_p}|_{\mathbf{x} = \hat{x}^i} + (2e_p \times rand(size(\bar{q}, 1), 1) - e_p)$ $\triangleright$ e_p: measurement noise at the pixel level

    for $j = 1 : N_{iter}$ **do** $\triangleright$ N_{iter}: number of iterations

        $\mathbf{J} = \mathbf{J_q}|_{\mathbf{x} = \hat{x}^j}$

        $\bar{\mathbf{q}} = \mathbf{q_p}|_{\mathbf{x} = \hat{x}^j}$ $\triangleright$ $\bar{q}, \hat{q}$: absolute and estimated value of measurement q, respectively

        $\hat{\mathbf{x}}_{j|j} = \hat{\mathbf{x}}_{j|j-1} + \lambda J^{-1}(\bar{\mathbf{q}} - \hat{\mathbf{q}})$; $\triangleright$ $\bar{x}, \hat{x}$: absolute and estimated value of the 6-DOF displacement D, respectively

    end for

    $\hat{\mathbf{x}}^i = \hat{\mathbf{x}}_{j|j}$

end for

---

sic parameters were set as follows: $[f_x, f_y] = [2634.64, 2624.02]$, $[c_x, c_y] = [1023.62, 380.04]$, and $[k_1, k_2, k_3, k_4] = [-0.4629, 0.3087, 0.0059, 0.0008]$ [21]. By considering the variance of the measurements, the measurement noise at the pixel level was set to $e_p = 0.2$. Fig. 3 shows a box-plot of the translational and rotational errors estimated from the 2-D marker using the homography transformation. The true displacement values for the translational and rotational movement are $\bar{\mathbf{x}} = [5, 5\sin((2\pi/100)t), 10\,000, 1, 5, -1]$ and $\bar{\mathbf{x}} = [5, 5, 10\,000, 1, 5\sin((2\pi/100)t), -1]$, respectively, and all units are in mm or degrees. The results obtained by applying sinusoidal movements with uniform random measurement noise to the translational displacement along the Y-axis and the rotational displacement around the yaw axis are shown in Fig. 3(a) and (b), respectively. As shown in the figure, the translational displacement along the Z-axis and rotational displacements around the pitch and yaw axes are more sensitive to measurement noise since the distance between the marker and the camera is much larger than the other components [10]. Since vision-based displacement measurement results do not guarantee the same level of high

accuracy in all axes, a quaternion-based sensor fusion method with an IMU is proposed.

III. QUATERNION-BASED ITERATIVE EXTENDED KALMAN FILTERING FOR SENSOR FUSION

To improve the accuracy of the displacement measurement, an IMU was added for sensor fusion. Due to the nonlinearity of the kinematics, EKF measurement updates were iterated (see Algorithm 2). The iterations were stopped when there were no significant changes in consecutive iterations or the maximum number of iterations was met. In this study, a QIEKF is introduced to address the gimbal locking in the Euler angle-based pose estimation problem.

A. Preliminary Definitions

The attitude can be expressed using the unit quaternion as follows [24]:

$$\mathbf{q} = q_w + q_x \mathbf{i} + q_y \mathbf{j} + q_z \mathbf{k} = \begin{bmatrix} q_w \\ \mathbf{q}_v \end{bmatrix} = \begin{bmatrix} \cos(\frac{\gamma}{2}) \\ \mathbf{e}_x \sin(\frac{\gamma}{2}) \\ \mathbf{e}_y \sin(\frac{\gamma}{2}) \\ \mathbf{e}_z \sin(\frac{\gamma}{2}) \end{bmatrix} \quad (6)$$

$$\begin{bmatrix} x_d \\ y_d \end{bmatrix} = \begin{bmatrix} x_c \left(1 + K_1 r^2 + k_2 r^4 + k_3 r^6\right) + 2K_3 x_c y_c + K_4 \left(r^2 + 2x_c^2\right) \\ y_c \left(1 + K_1 r^2 + k_2 r^4 + k_3 r^6\right) + K_3 \left(r^2 + 2y_c^2\right) + 2K_4 x_c y_c \end{bmatrix} \quad (3)$$
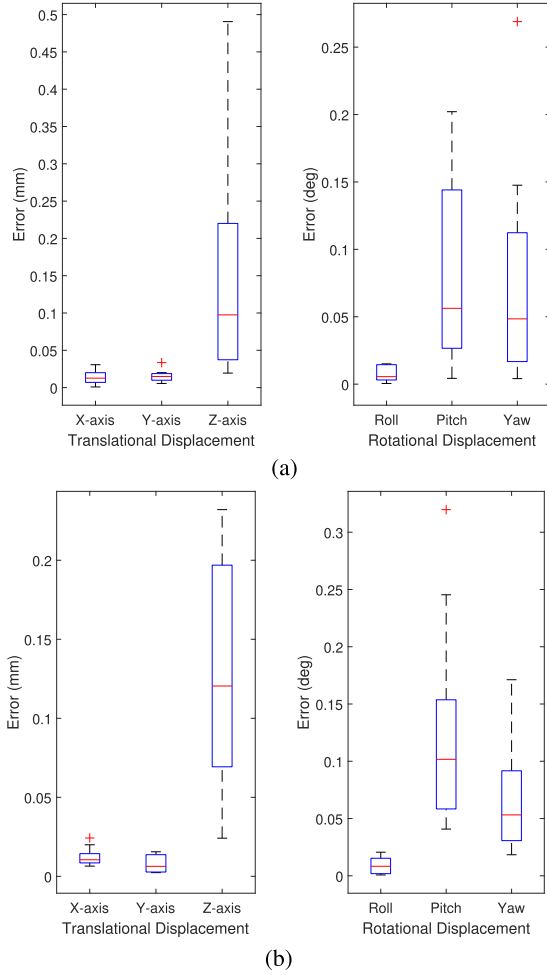
Fig. 3. Error analysis of 2-D marker with applying (a) translational movement in *Y*-axis and (b) rotational movement around the yaw axis.

where q_w and $\mathbf{q}_v = q_x\mathbf{i} + q_y\mathbf{j} + q_z\mathbf{k}$ represent the real and vector parts of the quaternion, respectively. γ and $\bar{\mathbf{e}} = [\mathbf{e}_x,\ \mathbf{e}_y,\ \mathbf{e}_z]^T$ represent the angle of rotation and the axis of rotation, respectively.

The quaternion multiplication $\otimes$ is defined by

$$\mathbf{p} \otimes \mathbf{q} = \begin{bmatrix} p_w q_w - \mathbf{p}_v{}^T \mathbf{q}_v \\ p_w \mathbf{q}_v + q_w \mathbf{p}_v + \mathbf{p}_v \times \mathbf{q}_v \end{bmatrix}. \quad (7)$$

In this article, an arbitrary 3-D vector in the body-fixed frame, $\mathbf{v}^b$, can be transformed into the corresponding vector in the navigation frame, $\mathbf{v}^n$, by the rotation matrix, $\mathbf{C}_b^n(\mathbf{q})$, as follows:

$$\mathbf{v}^n = \mathbf{C}_b^n(\mathbf{q})\,\mathbf{v}^b \quad (8)$$

where superscripts $\{n\}$ and $\{b\}$ refer to the navigation frame and body-fixed frame, respectively.

The rotation matrix $\mathbf{C}_b^n(\mathbf{q})$ represents the rotation matrix from the body frame to the navigation frame as follows (9), as shown at the bottom of the next page.

A unit quaternion can be expressed by its Euler angles $[\phi,\ \theta,\ \psi]^T$, where ϕ, θ, and ψ represent roll, pitch, and yaw,

respectively,

$$\mathbf{q} = \begin{bmatrix} \cos\left(\frac{\phi}{2}\right)\cos\left(\frac{\theta}{2}\right)\cos\left(\frac{\psi}{2}\right) + \sin\left(\frac{\phi}{2}\right)\sin\left(\frac{\theta}{2}\right)\sin\left(\frac{\psi}{2}\right) \\ \sin\left(\frac{\phi}{2}\right)\cos\left(\frac{\theta}{2}\right)\cos\left(\frac{\psi}{2}\right) - \cos\left(\frac{\phi}{2}\right)\sin\left(\frac{\theta}{2}\right)\sin\left(\frac{\psi}{2}\right) \\ \cos\left(\frac{\phi}{2}\right)\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\psi}{2}\right) + \sin\left(\frac{\phi}{2}\right)\cos\left(\frac{\theta}{2}\right)\sin\left(\frac{\psi}{2}\right) \\ \cos\left(\frac{\phi}{2}\right)\cos\left(\frac{\theta}{2}\right)\sin\left(\frac{\psi}{2}\right) - \sin\left(\frac{\phi}{2}\right)\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\psi}{2}\right) \end{bmatrix}. \quad (10)$$

Conversely, the Euler angles can be calculated from their unit quaternion through the following relation [16]:

$$\begin{bmatrix} \phi \\ \theta \\ \psi \end{bmatrix} = \begin{bmatrix} \mathrm{atan2}\left(2\left(q_w q_x + q_y q_z\right)/1 - 2\left(q_x^2 + q_y^2\right)\right) \\ \mathrm{asin}\left(2\left(q_w q_y - q_z q_x\right)\right) \\ \mathrm{atan2}\left(2\left(q_w q_z + q_x q_y\right)/1 - 2\left(q_y^2 + q_z^2\right)\right) \end{bmatrix}. \quad (11)$$

B. Sensors

An IMU measures the acceleration in three axes, the angular velocity about three axes, and the magnetic field in three axes relative to a body-fixed frame, and the measurements are corrupted by white Gaussian noise and slowly changing bias terms. The following mathematical model represents the relationship between the IMU-measured signal and the true value [24], [25]:

$$\boldsymbol{\omega}_m = \mathbf{K}^g \boldsymbol{\omega}_t + \boldsymbol{\omega}_b + \boldsymbol{\omega}_n \quad (12)$$

$$\mathbf{a}_m = \mathbf{K}^a \mathbf{C}_n^b(\mathbf{q})\left(\mathbf{a}_t - \mathbf{g}\right) + \mathbf{a}_b + \mathbf{a}_n \quad (13)$$

$$\mathbf{m}_m = \mathbf{K}^m \mathbf{C}_n^b(\mathbf{q}) \cdot \mathbf{M}^n + \mathbf{m}_b + \mathbf{m}_n \quad (14)$$

$$\mathbf{M}^n = \begin{bmatrix} |\mathbf{M}| \cos(\alpha)\cos(\beta) \\ |\mathbf{M}| \cos(\alpha)\sin(\beta) \\ |\mathbf{M}| \sin(\alpha) \end{bmatrix} \quad (15)$$

where the subscripts $\{m\}$, $\{t\}$, $\{b\}$, and $\{n\}$ denote measured, true, bias, and nonzero, respectively. $\mathbf{K}^g$, $\mathbf{K}^a$, and $\mathbf{K}^m$ are the scale matrices, which are ideally equal to the 3×3 identity matrix $\mathbf{I}_{3\times3}$. $\boldsymbol{\omega}_m = \{p_m, q_m, r_m\} \in \mathbb{R}^3$ are the measured angular rate signals, and $\boldsymbol{\omega}_t$, $\boldsymbol{\omega}_b$, and $\boldsymbol{\omega}_n \in \mathbb{R}^3$ are the true angular rate signal, slowly varying bias, and zero-mean Gaussian noise, respectively. $\mathbf{a}_m = \{a_{mx}, a_{my}, a_{mz}\} \in \mathbb{R}^3$ are the measured accelerations, and $\mathbf{g} \in \mathbb{R}^3$ is the gravitational acceleration in the inertial frame, and $\mathbf{a}_t$, $\mathbf{a}_b$, and $\mathbf{a}_n \in \mathbb{R}^3$ are the true acceleration, slowly varying bias, and zero-mean Gaussian noise, respectively. $\mathbf{m}_m \in \mathbb{R}^3$ is the measured magnetic field, $\mathbf{m}_b$ is a slowly varying bias term, and $\mathbf{m}_n$ is zero-mean Gaussian noise [26]. In (15), $\mathbf{M}^n$ is the magnetic field vector of the Earth, $|\mathbf{M}|$ is the magnitude of the magnetic flux density, α is the inclination angle of the Earth's magnetic field, and β is the declination angle. $|\mathbf{M}|$, α, and β depend on the geodetic location and time and can be obtained from the World Magnetic Model database. The effects of misalignment and scaling factor errors on the IMU were ignored in this study, as the effects of these parameters could be minimized with appropriate corrections [27].

C. Problem Formulation

In this study, the following discrete-time nonlinear state space system was investigated [28]:

$$\mathbf{x}_k = \mathbf{f}\left(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{w}_k\right) \tag{16}$$

$$\mathbf{z}_k = \mathbf{h}\left(\mathbf{x}_k, \mathbf{v}_k\right) \tag{17}$$

where $\mathbf{x}_k \in \mathbb{R}^n$, $\mathbf{u}_k \in \mathbb{R}^l$, and $\mathbf{z}_k \in \mathbb{R}^m$ are the state, input, and measurement at time k, respectively, and $\mathbf{f}(\cdot)$ and $\mathbf{h}(\cdot)$ represent the equations of state and the measurements, respectively. The process noise $\mathbf{w}_k$ and measurement noise $\mathbf{v}_k$ were assumed to be uncorrelated, white, zero-mean Gaussian noise as follows [29]:

$$\mathbf{w}_k \sim \mathcal{N}\left(0, \mathbf{Q}_k\right)$$
$$\mathbf{v}_k \sim \mathcal{N}\left(0, \mathbf{R}_k\right)$$
$$\mathrm{E}\left[\mathbf{w}_k \mathbf{v}_k^T\right] = 0 \tag{18}$$

where $\mathcal{N}(\boldsymbol{\mu}_k, \sum_k)$ is a Gaussian distribution with mean $\boldsymbol{\mu}_k$ and covariance $\sum_k$.

In the QIEKF prediction step in the continuous time domain, the true state, biases of the position, gyroscope, and magnetometer can be represented as follows:

$$\dot{\mathbf{x}} = \begin{bmatrix} \dot{\mathbf{q}} \\ \dot{\mathbf{p}} \\ \dot{\boldsymbol{\omega}}_b \\ \dot{\mathbf{m}}_b \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\mathbf{q} \otimes (\boldsymbol{\omega}_m - \boldsymbol{\omega}_b - \boldsymbol{\omega}_n) \\ \mathbf{p}_\omega \\ \boldsymbol{\omega}_\omega \\ \mathbf{m}_\omega \end{bmatrix} \tag{19}$$

where $\mathbf{p}_\omega$, $\boldsymbol{\omega}_\omega$, and $\mathbf{m}_\omega$ are the driving noises for the position, gyroscope, and magnetometer biases, respectively, which are assumed to follow a random walk (RW) process. The discrete-time expression of (19) is defined by

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{q}_k \\ \mathbf{p}_k \\ \boldsymbol{\omega}_{b,k} \\ \mathbf{m}_{b,k} \end{bmatrix} = \begin{bmatrix} \mathbf{q}_{k-1} \otimes \mathbf{q}\left\{(\boldsymbol{\omega}_{m,k} - \boldsymbol{\omega}_{b,k})\,\Delta t\right\} \\ \mathbf{p}_{k-1} \\ \boldsymbol{\omega}_{b,k-1} \\ \mathbf{m}_{b,k-1} \end{bmatrix}$$
$$+ \begin{bmatrix} \mathbf{w}_{q,k-1} \\ \mathbf{w}_{p,k-1} \\ \mathbf{w}_{g,k-1} \\ \mathbf{w}_{m,k-1} \end{bmatrix} \tag{20}$$

where Δt is the sampling time interval of the IMU, which is almost constant

$$\mathbf{q}\left\{(\boldsymbol{\omega}_{m,k} - \boldsymbol{\omega}_{b,k})\,\Delta t\right\}$$
$$= \begin{bmatrix} \cos\left((\boldsymbol{\omega}_{m,k} - \boldsymbol{\omega}_{b,k})\,\Delta t/2\right) \\ \frac{(\boldsymbol{\omega}_{m,k} - \boldsymbol{\omega}_{b,k})}{\|(\boldsymbol{\omega}_{m,k} - \boldsymbol{\omega}_{b,k})\|}\sin\left((\boldsymbol{\omega}_{m,k} - \boldsymbol{\omega}_{b,k})\,\Delta t/2\right) \end{bmatrix}. \tag{21}$$

During the sampling time interval Δt, linear evolution of ω is assumed. Thus, the first-order quaternion integrator is implemented as follows:

$$\mathbf{q}_k = \mathbf{q}_{k-1} \otimes \left(\mathbf{q}\left(\bar{\boldsymbol{\omega}}_k\,\Delta t\right) + \frac{\Delta t^2}{24}\begin{bmatrix} 0 \\ \boldsymbol{\omega}_{k-1} \times \boldsymbol{\omega}_k \end{bmatrix}\right) \tag{22}$$

where $\bar{\boldsymbol{\omega}}_k = 0.5(\boldsymbol{\omega}_k + \boldsymbol{\omega}_{k-1})$ is the average angular rate during time steps $k-1$ and k.

At (13), the initial accelerometer bias $\mathbf{a}_b$ is usually removed through calibration in the lab for various temperatures. In our approach, the measured acceleration magnitudes were pretested for significant deviations from gravity. Under stationary or low-dynamics conditions, the measured acceleration from (13) can be approximated as follows:

$$\mathbf{a}_m = \begin{bmatrix} a_{mx} \\ a_{my} \\ a_{mz} \end{bmatrix} \approx -\mathbf{C}_n^b(\mathbf{q})\,\mathbf{g} + \mathbf{a}_n$$
$$= \begin{bmatrix} 2\left(q_x \cdot q_z - q_w \cdot q_y\right) \\ 2\left(q_y \cdot q_z + q_w \cdot q_x\right) \\ q_w^2 - q_x^2 - q_y^2 - q_z^2 \end{bmatrix} + \mathbf{a}_n. \tag{23}$$

The measurement model is defined by stacking the accelerometers $\mathbf{a}_{m,k}$, magnetometers $\mathbf{m}_{m,k}$, vision-based yaws ψ_k, and vision-based position measurement vectors $\mathbf{p}_k$ as follows:

$$\mathbf{z}_k = \begin{bmatrix} \mathbf{a}_{m,k} \\ \mathbf{m}_{m,k} \\ \psi_k \\ \mathbf{p}_k \end{bmatrix}$$
$$= \begin{bmatrix} -\mathbf{C}_n^b(\mathbf{q}_k)\,\mathbf{g} + \mathbf{v}_{a,k} \\ \mathbf{C}_n^b(\mathbf{q}_k)\,\mathbf{B}^n + \boldsymbol{\omega}_{b,k} + \mathbf{v}_{m,k} \\ \mathrm{atan2}\left(2\left(q_w q_z + q_x q_y\right)/1 - 2\left(q_y^2 + q_z^2\right)\right) + \mathbf{v}_{\psi,k} \\ \mathbf{p}_k + \mathbf{v}_{p,k} \end{bmatrix}. \tag{24}$$

The accelerometer, magnetometer, vision-based yaw, and vision-based position measurement noises $\mathbf{v}_{a,k}$, $\mathbf{v}_{m,k}$, $\mathbf{v}_{\psi,k}$, and $\mathbf{v}_{p,k}$ are considered as uncorrelated zero-mean white Gaussian noise. Their measurement covariance matrices are $\mathbf{R}_k^a = \sigma_a^2 \cdot \mathbf{I}_{3\times3}$, $\mathbf{R}_k^m = \sigma_m^2 \cdot \mathbf{I}_{3\times3}$, $\mathbf{R}_k^\psi = \sigma_\psi^2 \cdot \mathbf{I}_{1\times1}$, and $\mathbf{R}_k^p = \sigma_p^2 \cdot \mathbf{I}_{3\times3}$, where $\mathbf{I}_{3\times3}$ and $\mathbf{I}_{1\times1}$ denote the 3×3 and 1×1 identity matrices, respectively. Therefore, the measurement covariance matrix $\mathbf{R}_k$ is defined as

$$\mathbf{R}_k = \begin{bmatrix} \mathbf{R}_k^a & 0 & 0 & 0 \\ 0 & \mathbf{R}_k^m & 0 & 0 \\ 0 & 0 & \mathbf{R}_k^\psi & 0 \\ 0 & 0 & 0 & \mathbf{R}_k^p \end{bmatrix}. \tag{25}$$

To compute the predicted state $\hat{\mathbf{x}}_{k|k-1}$ and the covariance $\mathbf{P}_{k|k-1}$, a standard EKF is implemented as follows:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{f}\left(\hat{\mathbf{x}}_{k-1|k-1}\right)$$
$$\mathbf{P}_{k|k-1} = \mathbf{F}_{k-1}\,\mathbf{P}_{k-1|k-1}\,\mathbf{F}_{k-1}^T + \Gamma_{k-1}\mathbf{Q}_{k-1}\Gamma_{k-1}^T \tag{26}$$

$$\mathbf{C}_b^n(\mathbf{q}) = \begin{bmatrix} q_w^2 + q_x^2 - q_y^2 - q_z^2 & 2\left(q_x \cdot q_y - q_w \cdot q_z\right) & 2\left(q_x \cdot q_z - q_w \cdot q_y\right) \\ 2\left(q_x \cdot q_y + q_w \cdot q_z\right) & q_w^2 - q_x^2 + q_y^2 - q_z^2 & 2\left(q_y \cdot q_z - q_w \cdot q_x\right) \\ 2\left(q_x \cdot q_z - q_w \cdot q_y\right) & 2\left(q_y \cdot q_z + q_w \cdot q_x\right) & q_w^2 - q_x^2 - q_y^2 - q_z^2 \end{bmatrix} \tag{9}$$

where the Jacobian matrix for the system dynamic equation, $\mathbf{F}_{k-1}$, is defined as

$$\mathbf{F}_{k-1} = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_{k|k-1}} = \mathbf{I}_{13\times13} + \mathbf{A}_N \Delta t \qquad (27)$$

where $\mathbf{A}_N$ is defined as

$$\mathbf{A}_N = \partial \mathbf{f}(\mathbf{x},\ \cdot)/\partial \mathbf{x} = \begin{bmatrix} \mathbf{W} & \mathbf{0}_{3\times3} & \mathbf{E} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \end{bmatrix} \qquad (28)$$

where

$$\mathbf{W} = \frac{\partial \frac{1}{2}\mathbf{q} \otimes (\boldsymbol{\omega}_m - \boldsymbol{\omega}_b)}{\partial \mathbf{q}} \qquad (29)$$

$$\mathbf{E} = \frac{\partial \frac{1}{2}\mathbf{q} \otimes (\boldsymbol{\omega}_m - \boldsymbol{\omega}_b)}{\partial \boldsymbol{\omega}_b} \qquad (30)$$

and Γ_{k-1} is defined by

$$\Gamma_{k-1} = \partial \mathbf{f}(\mathbf{x},\ \cdot)/\partial \mathbf{w} = \begin{bmatrix} \mathbf{Z} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} \end{bmatrix} \qquad (31)$$

$\mathbf{0}_{i\times j}$ refers to a matrix with rows i and columns j where all values are zero, and $\mathbf{Z}$ is defined by

$$\mathbf{Z} = \frac{\partial \frac{1}{2}\mathbf{q} \otimes (\boldsymbol{\omega}_m - \boldsymbol{\omega}_b - \boldsymbol{\omega}_n)}{\partial \boldsymbol{\omega}_n}. \qquad (32)$$

The updated state $\hat{\mathbf{x}}_{k|k}$ and the corresponding error covariance $\mathbf{P}_{k|k}$ are computed as

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \boldsymbol{v}_k$$
$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T \qquad (33)$$

where the Kalman gain $\mathbf{K}_k$ and the innovation $\boldsymbol{v}_k$ are defined as

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T \mathbf{S}_k^{-1} \qquad (34)$$

$$\boldsymbol{v}_k = \mathbf{z}_k - \mathbf{h}\left(\hat{\mathbf{x}}_{k|k-1}\right) \qquad (35)$$

and $\mathbf{S}_k$ can be expressed as

$$\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k \qquad (36)$$

where $\mathbf{H}_k$ is defined as

$$\mathbf{H}_k = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_{k|k-1}}.$$

An iterated EKF (IEKF) is designed to improve estimates since there is significant nonlinearity in the accelerometer $\mathbf{a}_{m,k}$, magnetometer $\mathbf{m}_{m,k}$, and vision-based yaw measurement models ψ_k. These improvements are achieved through local iterations of the EKF measurement update [30], [31]. IEKF iterations are usually stopped when there is no significant change in the estimate of the measurement update iterations or when other criteria are met, such as the maximum number of iterations, as described in Algorithm 2 [32]. The IEKF recursively corrects the center point of the Taylor extension to improve the linearization of the EKF while reducing the

linearization error, but it has a limitation in that the computation time increases. Since the vision-based position $\mathbf{x}_k$ is a measurement in the direct state, the measurement model is linear. Therefore, for vision-based position measurement, measurement updates from the standard EKF algorithm were applied.

Algorithm 2 Pseudocode for IEKF Measurement Updates

$\hat{\mathbf{x}}_k^0 = \hat{\mathbf{x}}_{k|k-1}$ ▷ Initialization: Set iteration $i = 0$, and set predicted state $\hat{\mathbf{x}}_k^0 = \hat{\mathbf{x}}_{k|k-1}$

while $\|\hat{\mathbf{x}}_k^{i+1} - \hat{\mathbf{x}}_k^i\|/\hat{\mathbf{x}}_k^i \leq \varepsilon$ and $i < N_{max}$ **do** ▷ IEKF measurement update iterations

$\quad \mathbf{H}_k^i = \left.\frac{\partial \mathbf{h}}{\partial x}\right|_{x=\hat{\mathbf{x}}_k^i}$ ▷ Computes the Jacobian from the best available state estimate

$\quad \mathbf{K}_k^i = \mathbf{P}_{k|k-1}(\mathbf{H}_k^i)^T (\mathbf{H}_k^i \mathbf{P}_{k|k-1}(\mathbf{H}_k^i)^T + \mathbf{R}_k)^{-1}$ ▷ Compute the Kalman gain

$\quad \hat{\mathbf{x}}_k^{i+1} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k^i(\mathbf{z}_k - \mathbf{h}(\hat{\mathbf{x}}_k^i) - \mathbf{H}_k^i(\hat{\mathbf{x}}_{k|k-1} - \hat{\mathbf{x}}_k^i))$ ▷ Compute the state estimate

$\quad i = i + 1$ ▷ Repeat the IEKF measurement update until a stopping criterion is satisfied

end while

$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_k^{i+1}$ ▷ Update the final state $\hat{\mathbf{x}}_{k|k}$

$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k^i \mathbf{S}_k^i (\mathbf{K}_k^i)^T$ ▷ Update the final covariance matrix $\mathbf{P}_{k|k}$

In particular, in vision-based measurement, the measurement value may be an outlier measurement due to issues such as illumination or visibility changes. Therefore, even if an abnormal vision measurement occurs, a stable state estimation can be achieved by applying chi-square-based defect detection. Specifically, the statistical parameter for measurement fault detection can be defined as follows [33]:

$$\bar{\boldsymbol{\varepsilon}}_k = \boldsymbol{v}_k^T \mathbf{S}_k^{-1} \boldsymbol{v}_k. \qquad (37)$$

The hypothesis test for evaluating system failure based on chi-square fault detection is as follows:

$$\begin{cases} H_0 : \bar{\boldsymbol{\varepsilon}}_k < \lambda & \text{Normal} \\ H_1 : \bar{\boldsymbol{\varepsilon}}_k \geq \lambda & \text{Fault} \end{cases} \qquad (38)$$

where λ can be determined based on the desired confidence level of the chi-square distribution, and the 95% confidence level is used. Chi-square-based fault detection is suitable for detecting sudden faults but not for detecting slowly increasing or constant faults [34].

IV. EXPERIMENTS

To verify the performance of the proposed sensor fusion algorithm, an experimental test was performed using a motorized motion stage capable of linear and rotational motion. The overall experimental setup using a network camera, IMU, and the motion stage is shown in Fig. 4. In the experiments, a commercial vision sensor from Hanwha Techwin (XNZ-L6320), an IMU from Xsens (MTi-680G), and motion stages from Thorlabs, Ins (PT1/M-Z8, MLJ150/M, KRRMTE/M) were used. The distance between the camera and the target planar marker installed on the stage was set to 10 m. The images
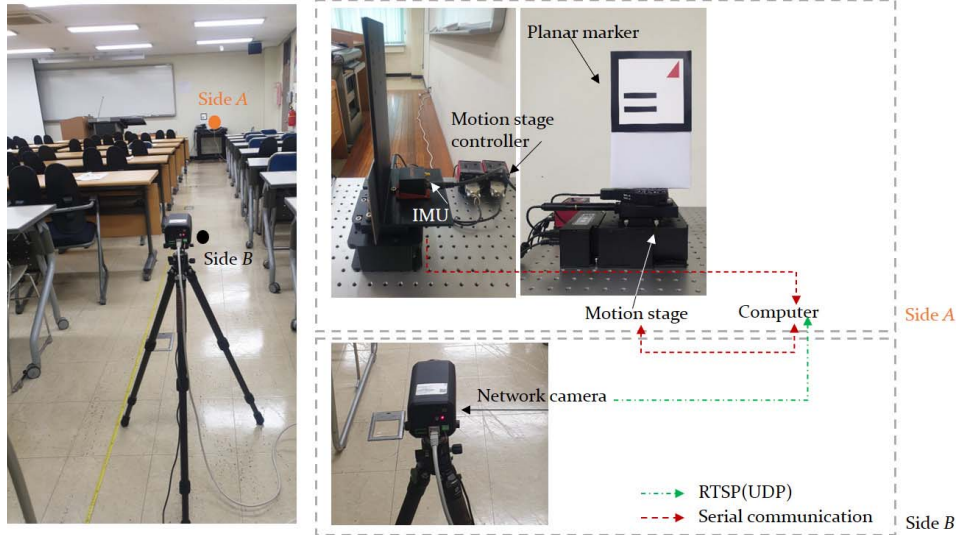
Fig. 4. Configuration of the experiments. The distance between the two sides is set to 10 m.

obtained from the vision sensor and the accelerations, angular rates, and magnetic fields from the IMU were transferred via a real-time streaming protocol (RTSP) and serial communication, respectively. User datagram protocol (UDP) communication, which transmits data between computers over a network using IP without establishing a connection a priori, was used to eliminate the physical connection with the main computing device and to expand the monitoring point by expanding the markers to multiple numbers.

A motion stage provides motorized linear translation and yaw motion to create motion between the two sides. Considering the behavior of port structures that are increasingly deteriorating, vertical settlement and rotational movement in the yaw direction were simulated. In a vision-based displacement measurement system, measurement errors may occur due to hardware limitations such as temporal aliasing or rolling shutter phenomena [35], [36], [37], [38]. Temporal aliasing artifacts are likely to appear when the sampling frequency is lower than twice the maximum frequency of the original signal [35], [36]. A rolling shutter is a method of image capture that does not capture an entire scene at a single instant in time but rather rapidly scans across the scene [37], [38]. In the experiments, it is assumed that the aforementioned hardware limitations do not occur with a stationary camera and low-frequency movements of motion stage within 15 Hz, which is half of the sampling frequency.

The 6-DOF displacements were calculated based on the marker image, and the accuracy was increased using the QIEKF with the accelerations, angular rates, and magnetic fields of the three axes obtained from the IMU (see Figs. 5 and 7). A low-pass filter, a conventional filtering method, was applied to the displacement results from the vision sensor for fairer comparison with the QIEKF with chi-square-based defect filtering. The low-pass filter on the time domain with the smoothing factor, α, is as follows:

$$\bar{\mathbf{x}}_k = \alpha \bar{\mathbf{x}}_{k-1} + (1 - \alpha) \mathbf{x}_k \tag{39}$$

where $\bar{\mathbf{x}}_k$ and $\mathbf{x}_k$ are the kth filtered output and raw-data input, respectively. In this experiment, α is set to 0.8. The estimated results using the QIEKF were compared with the values obtained from the motion stage, which can be considered the ground truth. In the sensor fusion, the covariance of the translational displacement along the Z-axis and the pitch and yaw were set to be larger than the other components by considering the sensitivity of the vision-based displacement measurement system as described in Section II-B (see Fig. 3). In the experimental tests, therefore, the process noise of $\mathbf{Q}$ was set to diag([0.015 · $\mathbf{I}_{3\times3}$, 0.001 · $\mathbf{I}_{3\times3}$, 0.001 · $\mathbf{I}_{3\times3}$, 0.35 · $\mathbf{I}_{3\times3}$]). The measurement noises of $\mathbf{R}_k^a$, $\mathbf{R}_k^m$, $\mathbf{R}_k^\psi$, and $\mathbf{R}_k^p$ were empirically set to 0.35 · $\mathbf{I}_{3\times3}$, 0.05 · $\mathbf{I}_{3\times3}$, 0.09, and diag([0.25, 0.25, 0.5]), respectively, through trial and error. In addition, the value of $N_{\max}$ was set to 20, which is a sufficiently large value empirically through trial and error. Since the IMU sensor used in this study can provide only the roll, pitch, and yaw attitudes as an attitude heading reference system (AHRS), only the attitude of the IMU was compared in the subsequent analysis.

As shown in Fig. 5, the QIEKF-based displacement results are reliably estimated even if the vision-based displacements deviate significantly due to erroneous detection of marker feature points. As shown in the figure, the innovation-based chi-square defect detection method can detect abrupt errors in measurements with a small number of calculations based on the discrepancy between the actual and predicted measurements. By applying this algorithm, excessive measurements that can be considered outliers are removed from the QIEKF results.

More specifically, Fig. 6(a) and (b) represents the innovation sequence $\boldsymbol{v}_k$ with the corresponding standard deviation of the innovation covariance matrix (i.e., $\pm 3(\mathbf{S})^{1/2}$) and $\bar{\boldsymbol{\varepsilon}}_k = \boldsymbol{v}_k^T \mathbf{S}_k^{-1} \boldsymbol{v}_k$ with the threshold $\boldsymbol{\lambda}$, respectively. The estimated translational displacement in the Y-axis of approximately 12 and 90 s extracted based on the vision sensor are outlier measurements because the innovation value has a much
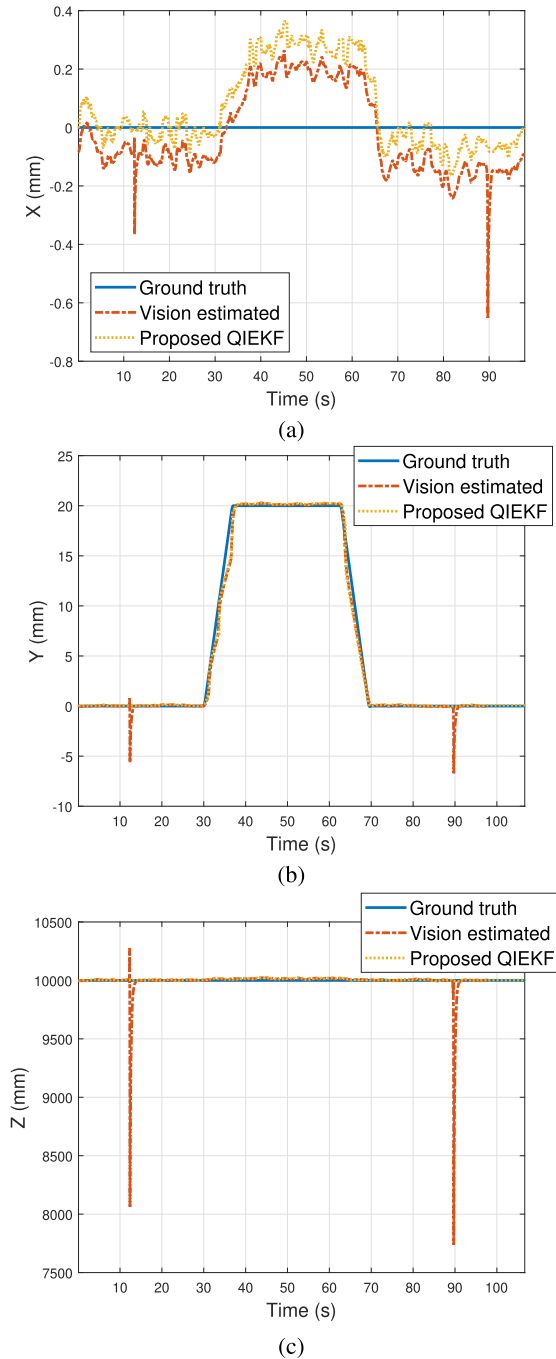
Fig. 5. Translational displacements along (a) *X*, (b) *Y*, and (c) *Z* axes via vision and IMU sensor fusion.

Fig. 6. Performance of the chi-square fault detection method. (a) Innovation analysis of vision-based *Y*-axis displacement and (b) statistical parameter for measurement of fault detection with a threshold of 95% confidence level.

larger value than the standard deviation value. The outlier measurement value detected through the proposed chi-square fault detection is not used for the measurement update of the proposed QIEKF algorithm, as shown in Fig. 6(b). Therefore, it can be seen that the proposed QIEKF algorithm outputs reliable pose estimation values regardless of the outlier measurement values of the pose measurements extracted from the vision sensor.

Additionally, the results of the QIEKF can be confirmed through Fig. 7, which shows that the signal drift of the IMU is eliminated as a result of the proposed sensor fusion.
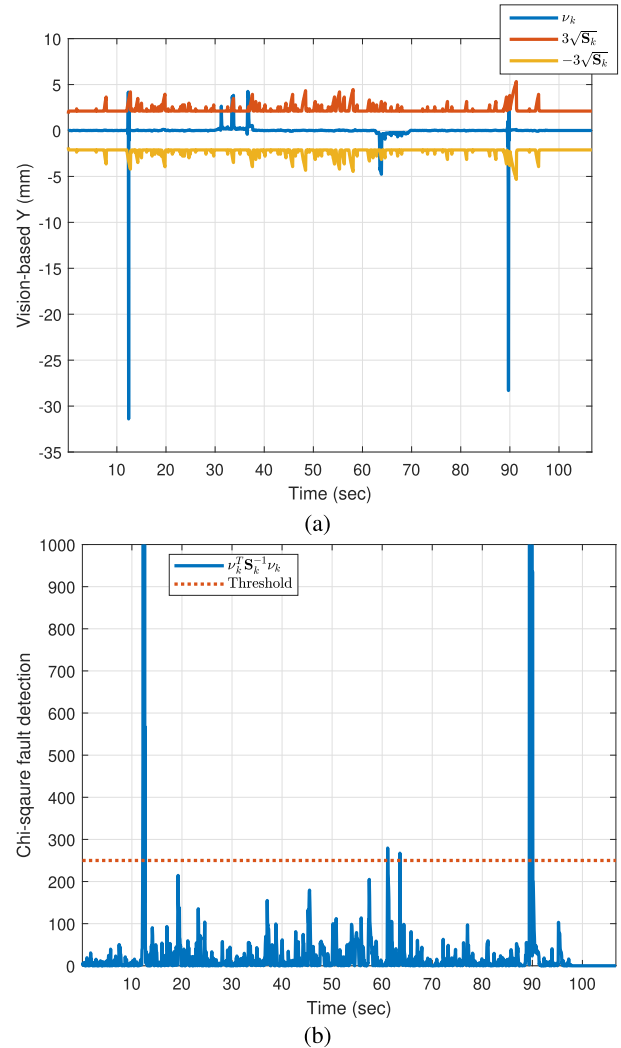
The gravitational vector measured by the accelerometer contains only roll and pitch information, and by comparing the measured magnetometer value to the magnetic field in the north, east, and down (NED) coordinate system at that location, three-axis information about roll, pitch, and yaw can be obtained. However, the measured magnetic field can be easily distorted and interfered with by external electromagnetic disturbances induced by time-varying currents in ferrous objects, motors and transmission wires. Therefore, we designed the proposed filter to apply more weight to the attitude measurements calculated based on the vision sensor.

As shown in the results from the proposed filter, the rotational displacements about the three axes achieve reliable estimation by fusing data from the vision sensor and the IMU. In particular, the yaw angle estimation results from the proposed QIEKF algorithm are the most accurate and reliable by combining observations from different sensors, such as vision, inertial, and magnetic sensing. To quantitatively verify the proposed algorithm, the estimated pose based on the

TABLE I
RMSE OF TRANSLATIONAL AND ROTATIONAL DISPLACEMENTS, RESPECTIVELY

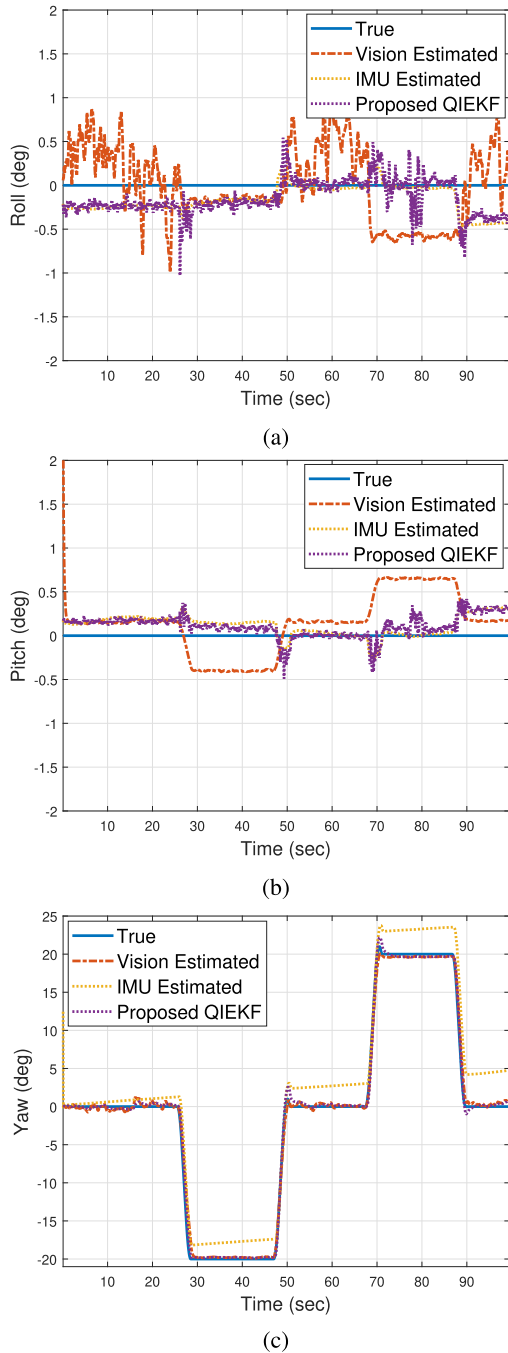| Case	Axis	QIEKF	Vision	Case	Axis	QIEKF	Vision	IMU
	X	0.14	0.15		Roll	0.24	0.42	0.22
Translational disp. (mm)	Y	0.66	0.76	Rotational disp. (deg)	Pitch	0.16	0.34	0.16
	Z	9.53	114.58		Yaw	0.46	0.68	2.75

Fig. 7. Rotational displacement results. (a) Roll, (b) pitch, and (c) yaw via vision and IMU sensor fusion.

smaller than 10 mm and 0.5° for translational and rotational cases, respectively, after fusing the vision with the IMU.

V. CONCLUSION

In this article, a QIEKF method for fusing a vision sensor with an IMU is proposed. By using the feature points of a 2-D planar marker, the 6-DOF displacements between a camera and a marker can be estimated. A vision-based displacement measurement system, however, has drawbacks in that displacement error may occur due to measurement noise and is highly limited to environmental conditions such as light or weather. An IMU sensor, which can measure accelerations, angular rates, and magnetic fields, measures the behavior of the system with high accuracy; however, there are several problems with signal drift and misalignment. To solve the aforementioned problems, a QIEKF for fusing different types of sensors is proposed. The QIEKF can be summarized as follows.

1) The proposed algorithm first integrates the angular velocity of the gyroscope to predict the attitude. Then, the attitude is updated via vector matching in the NED frame of measurements from the accelerometer and magnetometer. The estimated gyro biases are compensated from angular rates from the gyroscope to mitigate the drift error.
2) The attitude is updated using the yaw and the position values calculated from the vision sensor as measurements. Here, the state transition model of the position is modeled as a Gaussian random distribution.
3) An iterative EKF was applied to the measurement update to reduce the influence on the nonlinearity of the nonlinear measurement model corresponding to the accelerometer and magnetometer.
4) In addition, the influence of outliers from vision sensor measurements was minimized through a chi-square-based defect detection algorithm.

Experimental tests were performed to verify the performance of the proposed system. By considering the sensitivity of the vision-based displacement measurement system due to measurement noise in image processing techniques, the covariance of the translational displacement along the Z axis and the pitch and yaw were set to be larger than the other components. The results show that the proposed QIEKF method estimates 6-DOF displacements with the lowest RMSE in comparison with the results from the vision and IMU sensors. The proposed system can be applied in the field, such as port structures, where the behavior, especially the settlements, must be measured regularly.

root-mean-square error (RMSE) is evaluated against the ground truth. The RMSE in both cases is summarized in Table I. The RMSE for the proposed sensor fusion method is

References

[1] H. Jeon, Y. Bang, and H. Myung, "A paired visual servoing system for 6-DOF displacement measurement of structures," *Smart Mater. Struct.*, vol. 20, no. 4, Apr. 2011, Art. no. 045019.

[2] Y. Weng, J. Shan, Z. Lu, X. Lu, and B. F. Spencer, "Homography-based structural displacement measurement for large structures using unmanned aerial vehicles," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 36, no. 9, pp. 1114–1128, 2021.

[3] Y. H. Hong, S. G. Lee, and H. S. Lee, "Design of the FEM-FIR filter for displacement reconstruction using accelerations and displacements measured at different sampling rates," *Mech. Syst. Signal Process.*, vol. 38, no. 2, pp. 460–481, Jul. 2013.

[4] H. Jeon, S. Choi, J. U. Shin, Y. Kim, and H. Myung, "High-speed 6-DOF structural displacement monitoring by fusing ViSP (visually servoed paired structured light system) and IMU with extended Kalman filter," *Struct. Control Health Monit.*, vol. 24, no. 6, p. e1926, 2017.

[5] X. Meng, G. W. Roberts, A. H. Dodson, E. Cosser, J. Barnes, and C. Rizos, "Impact of GPS satellite and pseudolite geometry on structural deformation monitoring: Analytical and empirical studies," *J. Geodesy*, vol. 77, no. 12, pp. 809–822, Jun. 2004.

[6] J.-W. Park, J.-J. Lee, H.-J. Jung, and H. Myung, "Vision-based displacement measurement method for high-rise building structures using partitioning approach," *NDT & E Int.*, vol. 43, no. 7, pp. 642–647, 2010.

[7] H. Jeon, Y. J. Kim, W. Myeong, and H. Myung, "One-way ViSP (Visually servoed paired structured light system) for structural displacement monitoring," *Smart Mater. Struct.*, vol. 26, no. 8, Aug. 2017, Art. no. 085044.

[8] C.-Z. Dong, O. Celik, F. N. Catbas, E. J. O'Brien, and S. Taylor, "Structural displacement monitoring using deep learning-based full field optical flow methods," *Struct. Infrastruct. Eng.*, vol. 16, no. 1, pp. 51–71, Jan. 2020.

[9] R. Janeliukstis and X. Chen, "Review of digital image correlation application to large-scale composite structure testing," *Compos. Struct.*, vol. 271, Sep. 2021, Art. no. 114143.

[10] H. Jeon, Y. Kim, D. Lee, and H. Myung, "Vision-based remote 6-DOF structural displacement monitoring system using a unique marker," *Smart Struct. Syst.*, vol. 13, no. 6, pp. 927–942, Jun. 2014.

[11] A. Smyth and M. Wu, "Multi-rate Kalman filtering for the data fusion of displacement and acceleration response measurements in dynamic system monitoring," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 706–723, 2007.

[12] H. Sohn, K. Kim, J. Choi, G. Koo, and J. Chung, "Development of a high accuracy and high sampling rate displacement sensor for civil engineering structures monitoring," in *Proc. Int. Conf. Exp. Vib. Anal. Civil Eng. Struct.* Springer, 2017, pp. 62–70.

[13] J.-W. Park, D.-S. Moon, H. Yoon, F. Gomez, B. F. Spencer, Jr., and J. R. Kim, "Visual-inertial displacement sensing using data fusion of vision-based displacement with acceleration," *Structural Control Health Monitor.*, vol. 25, no. 3, p. e2122, Mar. 2018.

[14] T. Wu et al., "Accurate structural displacement monitoring by data fusion of a consumer-grade camera and accelerometers," *Eng. Struct.*, vol. 262, Jul. 2022, Art. no. 114303.

[15] J. W. Park, D. S. Moon, H. Yoon, F. Gomez, B. F. Spencer, and J. R. Kim, "Real-time structural displacement estimation by fusing asynchronous acceleration and computer vision measurements," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 37, no. 6, pp. 688–703, 2022.

[16] J. Diebel, "Representing attitude: Euler angles, unit quaternions, and rotation vectors," *Matrix*, vol. 58, nos. 15–16, pp. 1–35, 2006.

[17] W. Youn, H. Choi, A. Cho, S. Kim, and M. B. Rhudy, "Aerodynamic model-aided estimation of attitude, 3-D wind, airspeed, AOA, and SSA for high-altitude long-endurance UAV," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 6, pp. 4300–4314, Dec. 2020.

[18] W. Youn and H. Myung, "Robust interacting multiple model with modeling uncertainties for maneuvering target tracking," *IEEE Access*, vol. 7, pp. 65427–65443, 2019.

[19] B. M. Bell and F. W. Cathey, "The iterated Kalman filter update as a Gauss-Newton method," *IEEE Trans. Autom. Control*, vol. 38, no. 2, pp. 294–297, Feb. 1993.

[20] G. Welch and G. Bishop, "An introduction to the Kalman filter," Tech. Rep., 1995.

[21] J. Min, Y. Bang, H. Bang, and H. Jeon, "Port structure inspection based on 6-DOF displacement estimation combined with homography formulation and genetic algorithm," *Appl. Sci.*, vol. 11, no. 14, p. 6470, Jul. 2021.

[22] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[23] Y. Yan, Z. Mao, J. Wu, T. Padir, and J. F. Hajjar, "Towards automated detection and quantification of concrete cracks using integrated images and lidar data from unmanned aerial vehicles," *Struct. Control Health Monit.*, vol. 28, no. 8, p. e2757, 2021.

[24] A. M. Sabatini, "Quaternion-based extended Kalman filter for determining orientation by inertial and magnetic sensing," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 7, pp. 1346–1356, Jul. 2006.

[25] W. Youn, "Magnetic fault–tolerant navigation filter for a UAV," *IEEE Sensors J.*, vol. 20, no. 22, pp. 13480–13490, Nov. 2020.

[26] W. Youn and S. Andrew Gadsden, "Combined quaternion-based error state Kalman filtering and smooth variable structure filtering for robust attitude estimation," *IEEE Access*, vol. 7, pp. 148989–149004, 2019.

[27] W. Youn, M. B. Rhudy, A. Cho, and H. Myung, "Fuzzy adaptive attitude estimation for a fixed-wing UAV with a virtual SSA sensor during a GPS outage," *IEEE Sensors J.*, vol. 20, no. 3, pp. 1456–1472, Feb. 2020.

[28] M. B. Rhudy, M. L. Fravolini, Y. Gu, M. R. Napolitano, S. Gururajan, and H. Chao, "Aircraft model-independent airspeed estimation without Pitot tube measurements," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 3, pp. 1980–1995, Jul. 2015.

[29] M. B. Rhudy, Y. Gu, J. N. Gross, and H. Chao, "Onboard wind velocity estimation comparison for unmanned aircraft systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 1, pp. 55–66, Feb. 2017.

[30] L. Liang-Qun, J. Hong-Bing, and L. Jun-Hui, "The iterated extended Kalman particle filter," in *Proc. IEEE Int. Symp. Commun. Inf. Technol.*, vol. 2, Oct. 2005, pp. 1213–1216.

[31] K. J. Havlí and O. Straka, "Performance evaluation of iterated extended Kalman filter with variable step-length," in *Proc. J. Phys., Conf.*, vol. 659, 2015, Art. no. 012022.

[32] M. Bloesch, M. Burri, S. Omari, M. Hutter, and R. Siegwart, "Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback," *Int. J. Robot. Res.*, vol. 36, no. 10, pp. 1053–1072, 2017.

[33] B. Brumback and M. Srinath, "A chi-square test for fault-detection in Kalman filters," *IEEE Trans. Autom. Control*, vol. AC-32, no. 6, pp. 552–554, Jun. 1987.

[34] R. Wang, Z. Xiong, J. Liu, J. Xu, and L. Shi, "Chi-square and SPRT combined fault detection for multisensor navigation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 3, pp. 1352–1365, Jun. 2016.

[35] H. Yoon, H. Elanwar, H. Choi, M. Golparvar-Fard, and B. F. Spencer, "Target-free approach for vision-based structural system identification using consumer-grade cameras," *Struct. Control Health Monitor.*, vol. 23, no. 12, pp. 1405–1416, Dec. 2016.

[36] D. Feng and M. Q. Feng, "Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection—A review," *Eng. Struct.*, vol. 156, pp. 105–117, Feb. 2018.

[37] H. Yoon, V. Hoskere, J.-W. Park, and B. Spencer, "Cross-correlation-based structural system identification using unmanned aerial vehicles," *Sensors*, vol. 17, no. 9, p. 2075, Sep. 2017.

[38] Y. Dai, H. Li, and L. Kneip, "Rolling shutter camera relative pose: Generalized epipolar geometry," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4132–4140.

Haemin Jeon received the B.S. degree from Handong Global University, Pohang, South Korea, in 2008, and the M.S. and Ph.D. degrees in civil and environmental engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2010 and 2014, respectively.

From 2014 to 2016, she was a Senior Researcher with the Agency for Defense Development (ADD), Daejeon. Since 2016, she has been a Professor at the Department of Civil and Environmental Engineering, Hanbat National University, Daejeon. Her current research interests include structural health monitoring, multisensor fusion, automation in construction, and machine learning.

Jiyoung Min received the B.S., M.S., and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2006, 2008, and 2012, respectively, all in civil and environmental engineering.

Since 2013, she has been a Senior Researcher with the Department of Structural Engineering Research, Korea Institute of Civil Engineering and Building Technology (KICT), Gyeonggi-do, South Korea. Her current research interests include concrete deterioration, nondestructive inspection, structural health monitoring, signal processing, and multisensor fusion.

Wonkeun Youn received the B.S. degree from Handong Global University, Pohang, South Korea, in 2008, and the M.S. degree in mechanical engineering and the Ph.D. degree in robotics from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2010 and 2020, respectively.

From 2011 to 2021, he was a Senior Researcher with the Unmanned Aircraft System Research Division, Korea Aerospace Research Institute (KARI), Daejeon. Since 2021, he has been an Assistant Professor at Chungnam National University. His current research interests include multisensor fusion, Bayesian estimation theory, multimodal target tracking, and global positioning system (GPS)/inertial navigation system (INS)-based navigation.

Hyuntae Bang received the B.S. and M.S. degrees from Hanbat National University, Daejeon, South Korea, in 2019 and 2021, respectively, all in civil and environmental engineering. He is currently pursuing the Ph.D. degree with Chungnam National University, Daejeon.

His current research interests include structural health monitoring, artificial intelligence, simultaneous localization and mapping, and robot navigation.