

Augmented Feature Point Initialization Method for Vision/Lidar Aided 6-DoF Bearing-Only Inertial SLAM

Sukchang Yun*, Byoungjin Lee*, Yeon-Jo Kim*, Young Jae Lee* and Sangkyung Sung[†]

Abstract – This study proposes a novel feature point initialization method in order to improve the accuracy of feature point positions by fusing a vision sensor and a lidar. The initialization is a process that determines three dimensional positions of feature points through two dimensional image data, which has a direct influence on performance of a 6-DoF bearing-only SLAM. Prior to the initialization, an extrinsic calibration method which estimates rotational and translational relationships between a vision sensor and lidar using multiple calibration tools was employed, then the feature point initialization method based on the estimated extrinsic calibration parameters was presented. In this process, in order to improve performance of the accuracy of the initialized feature points, an iterative automatic scaling parameter tuning technique was presented. The validity of the proposed feature point initialization method was verified in a 6-DoF bearing-only SLAM framework through an indoor and outdoor tests that compare estimation performance with the previous initialization method.

Keywords: Vision sensor, Lidar, Feature point initialization, 6-DoF inertial SLAM

1. Introduction

The performance of an inertial sensor-based bearing-only SLAM (Simultaneous Localization and Mapping) depends on the performance of the inertial measurements and the reliability of feature point information, i.e., the accuracy of the positions of feature points. Among three dimensional feature point initialization methods, a number of delayed initialization methods have been developed, which can determine the positions using a monocular vision sensor with time delay [1-3]. There has also been an undelayed initialization method to reduce the time delay by using multiple vision sensors [4]. For a representative delayed initialization method, Bailey [1] proposed a method of determining three dimensional positions of feature points by using positions and attitudes of a robot at two sequential points.

For the initialization method [1], the performance of estimating three dimensional positions heavily relies on the accuracy of the INS (Inertial Navigation System). After feature points are measured by a vision sensor at different time instances, three dimensional positions are determined by using the measured pixel coordinates of the feature points in those two images and the position and attitude of a vision sensor at each point. In this process, the position accuracy of feature points depends on the position and attitude accuracy of a vision sensor and the distance between the two points, that is, the length of baseline.

However, since the position and attitude accuracy of a vision sensor is computed from INS, the accuracy and length of baseline contain mutual tradeoff effects. Therefore, inertial sensors tend to yield both direct accumulation of sensor noise and indirect propagation of localization error in the navigation systems.

Researches related with integrating a vision sensor and a lidar have been studied in various fields. Especially, integrating a vision sensor and a two dimensional lidar in a mobile robot platform is a representative topic of works. Researches about detecting moving objects such as cars or pedestrians and determining positions of the targets have also been studied [5-6]. Moreover, path planning techniques of an UAV (Unmanned Aerial Vehicle) was developed with the sensor integration framework [7]. For determining target position, one may utilize ToF (Time of Flight) cameras, such as Kinect [8], DepthSense [9], Basler ToF [10] and so on. In a latest report about ToF cameras on Novus Light Technologies Today (light technology-specific web site), disadvantages of recent ToF cameras were pointed out as “new pulsed modulation cameras still suffer from the fact that the depth resolution that they offer at frame rates of 30 frames/sec is no more than +/- 1cm over working distances between 0.5 to 5m” [11]. Thus, for both indoor and outdoor applications, a lidar which produces longer range measurements and is robust to ambient light should be employed.

Recently, the research utilizing vision and lidar sensors has actively been studied in UGV (Unmanned Ground Vehicle) applications. Image information was used to detect roads or lanes, concurrently, and lidar information was used to determine drivable regions [12-13]. Additionally, vision and lidar sensors were also utilized to detect moving

[†] Corresponding Author: Dept. of Aerospace Information Engineering, Konkuk University, Korea. (sksung@konkuk.ac.kr)

* Dept. of Aerospace Information Engineering, Konkuk University, Korea. ({amerisan, younglee}@konkuk.ac.kr, schumir@naver.com, yeonjo.k@gmail.com)

Received: September 10, 2015; Accepted: May 16, 2016

objects such as cars and humans in order to improve the robustness against driving environments [14-15]. However, most researches aimed to detect objects, thus, a study on applying two sensors to calculating a 6-DoF (Degree of Freedom) navigation solutions has been rarely reported.

The researches mentioned in the previous paragraphs are based on the extrinsic calibration techniques which estimate rotational and translational relationships between a vision sensor and a lidar. A representative calibration method uses a single vision sensor and a two dimensional lidar with a checkerboard, and has two-stage procedures as follows: (1) estimating linear solutions, (2) refinement based on a non-linear optimization [16]. Based on the method, the calibration between a vision sensor and a multi-planar lidar system [17], and the calibration between multiple vision sensors and a lidar [18] were carried out, respectively.

For three dimensional application, well-documented work about extrinsic calibration between a camera and a three dimensional lidar can be found in [19]. The author proposed a method for fast extrinsic calibration between a camera and a three dimensional lidar with a checkerboard calibration target, and they provided a useful laser-camera calibration toolbox for Matlab. Estimating calibration parameters, they decoupled the problem into estimating a translation part and a rotation part independently, and then utilized an iterative optimization procedure. However, for each scan user should select a region of points that mostly lie on the planar calibration target. That is a bit bothersome work, like a well-known camera calibration Matlab toolbox. Pandey et al. [20] proposed an automatic targetless extrinsic calibration method between a camera and 3D laser scanner. The calibration algorithm uses a mutual information framework to estimate the statistical dependence between the laser scanner reflectivity values and image intensity values. However, this approach cannot apply directly to range sensors without associated reflectivity information. Recently, Moghadam et al. [21] proposes an extrinsic range-image calibration technique which utilizes strong line-to-line geometric constraints to jointly estimate the 6DoF transformation between the coordinate systems. The method requires no special calibration pattern and is not limited to any particular types or configurations of range and imaging sensors. However, human intervention is needed for correspondence of 2D-to-3D lines.

Additionally, in order to make two dimensional lidar information expand to three dimensional information, a lidar was mounted on a rotating platform, and calibration methods between the rotating lidar system and a vision sensor were presented [22-23]. In [22], an extrinsic calibration was conducted based on [19], but they used only the point correspondences which the user hand selects. Hur et al. [23] presented result of an extrinsic calibration between a vision sensor and a rotating lidar mounted on a one-axis gimbal platform. In this process, a single calibration tool was utilized, and linear solutions were

computed by applying direct linear transformation to estimate rotational parameters. However, the translational vector was measured manually. Furthermore, the offset effect between an actual sensing point of the lidar and the gimbal coordinate system was not considered.

Considering the previous research history, this study proposes a novel feature point initialization method in a 6-DoF bearing-only inertial SLAM framework, which results in the enhanced navigation performance through the improved accuracy of feature point position. In order to improve the initial positions of feature points, a vision sensor and a lidar were tightly integrated during calibration period. In this process, an extrinsic calibration method which estimates rotational and translational relationships between heterogeneous sensors, and a feature point initialization method utilizing the estimated calibration parameters are sequentially adapted. The proposed system can operate both indoors and outdoors, and working range of the system depends on maximum detectable range of a lidar.

The rest of this paper is organized as follow: Section 2 presents an extrinsic calibration method between a vision sensor and a lidar and test results. In section 3, a feature point initialization method is proposed based on the estimated calibration parameters in section 2, and initialization test results are presented for proving the validity of the method. Besides, its practical adaptation in the SLAM framework is also summarized in the section. In section 4, an indoor and outdoor tests are described and experimental results are presented to verify the validity of the proposed feature point initialization method in a 6-DoF bearing-only SLAM by comparing with the existing initialization method. Finally, section 5 concludes the paper with a discussion of our current research.

2. Vision/Lidar Extrinsic Calibration

2.1. Extrinsic calibration

Generally, lidar information is provided in the form of a polar or spherical coordinates. The system consisting of camera and two dimensional lidar on a rotating gimbal unit is illustrated as Fig. 1. The system consisting of camera and two dimensional lidar on a rotating gimbal unit is illustrated as Fig. 1.

In the above system, two sensors are aligned to each other with a distance, \mathbf{t}_l^c , and attitude difference, \mathbf{C}_l^c . A point expressed in the lidar Cartesian coordinate system, \mathbf{p}_l^l , having its origin, O^l , and metric units, can be represented with respect to the camera coordinate system by using a three dimensional rotational matrix, \mathbf{C}_l^c , and a translational vector, \mathbf{t}_l^c , as shown in (1). The subscript of the position vector, \mathbf{P} , means sensor data (c : camera, l : lidar), and the superscript means expressed coordinate systems (P : pixel, c : camera, l : lidar).

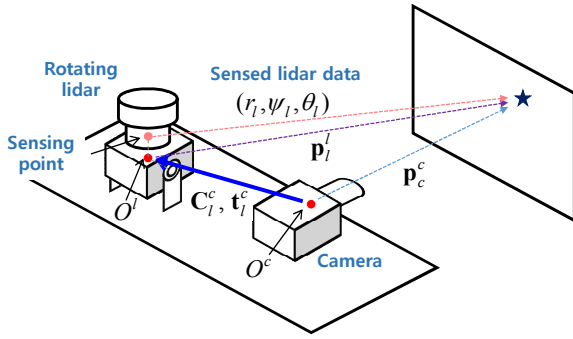


Fig. 1. Camera-rotating lidar system

$$\mathbf{p}_l^c = \mathbf{C}_l^c \mathbf{p}_l^l + \mathbf{t}_l^c \quad (1)$$

In the case of the lidar system which produces three dimensional measurements via rotating two dimensional lidar, there may be difference between the sensing point at which light sources are emitted and the rotational axis. In this case, the origin of the lidar coordinate system is a point where the sensing point and the rotational axis intersect perpendicularly, and the offset should be compensated. For the lidar system, raw measurements are expressed in the spherical coordinate system and consist of a range, r_l , a corresponding azimuth angle, ψ_l , and an elevation angle, θ_l , from the rotating anchor. Lidar measurements in the spherical coordinate system can be represented as a point, \mathbf{p}_l^l , in the lidar Cartesian coordinate system as follows.

$$\mathbf{p}_l^l = \begin{bmatrix} \cos \theta_l \cos \psi_l & 0 & 0 & 0 & 0 \\ -\sin \theta_l & 0 & 0 & \cos \theta_l d_y - \sin \theta_l d_z \\ \cos \theta_l \sin \psi_l & 0 & 0 & \sin \theta_l d_y + \cos \theta_l d_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_l \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (2)$$

In (2), the offset is represented by d_y and d_z . When rotating around the x-axis of the lidar coordinate system, X^l , the offset is described as shown in Fig. 2.

With the camera intrinsic parameters, \mathbf{p}_l^c can be transferred on the pixel coordinate system.

In (3), if it is assumed that skew coefficients of an image is zero, the intrinsic matrix, \mathbf{K} , is composed of a focal length and principal points. In the equation, $(\cdot)_n$ represents a normalized vector divided by the z-component of the vector in the parentheses.

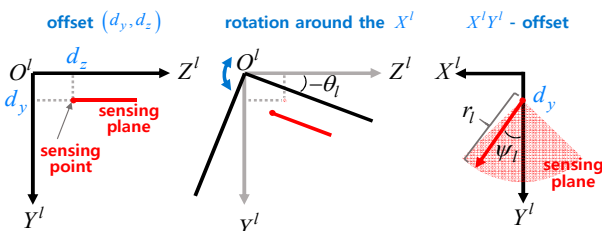


Fig. 2. Offset between the sensing point and the origin of the lidar coordinate system

$$\begin{aligned} \mathbf{p}_l^p &= \mathbf{K} \cdot (\mathbf{p}_l^c)_n \\ &= \mathbf{K} (\mathbf{C}_l^c \mathbf{p}_l^l + \mathbf{t}_l^c)_n \\ &= \begin{bmatrix} fl & 0 & cc_x \\ 0 & fl & cc_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\mathbf{C}_{l,(1-row)}^c \mathbf{p}_l^l + \mathbf{t}_{l,x}^c}{\mathbf{C}_{l,(3-row)}^c \mathbf{p}_l^l + \mathbf{t}_{l,z}^c} & \frac{\mathbf{C}_{l,(2-row)}^c \mathbf{p}_l^l + \mathbf{t}_{l,y}^c}{\mathbf{C}_{l,(3-row)}^c \mathbf{p}_l^l + \mathbf{t}_{l,z}^c} & 1 \end{bmatrix}^T \\ &= \begin{bmatrix} fl \frac{\mathbf{C}_{l,(1-row)}^c \mathbf{p}_l^l + \mathbf{t}_{l,x}^c}{\mathbf{C}_{l,(3-row)}^c \mathbf{p}_l^l + \mathbf{t}_{l,z}^c} + cc_x & fl \frac{\mathbf{C}_{l,(2-row)}^c \mathbf{p}_l^l + \mathbf{t}_{l,y}^c}{\mathbf{C}_{l,(3-row)}^c \mathbf{p}_l^l + \mathbf{t}_{l,z}^c} + cc_y & 1 \end{bmatrix}^T \end{aligned} \quad (3)$$

For the calibration, \mathbf{C}_l^c and \mathbf{t}_l^c are estimated by minimizing the cost function which is a function of the pairs of common feature points. The problem minimizing k pairs of feature points can be formulated as shown in (4). This non-linear least squares problem is generally solved by using a Levenberg-Marquardt algorithm [24], etc.

$$\arg \min_{\mathbf{C}_l^c, \mathbf{t}_l^c} \sum_k \left(\left\| \mathbf{p}_{c,k}^p - \mathbf{p}_{l,k}^p \right\|^2 \right) \quad (4)$$

The accuracy of the estimated calibration parameters depends on the distribution of feature points. During calibration process, the most essential thing is to extract common feature points for efficient data association. For a clear association between feature points, calibration tools were newly proposed, with which feature points are easily detectable.

The calibration procedures are summarized as follows. First, for each calibration tool, sensor data is acquired from a vision sensor and a lidar within the sensing range of sensors. Then, feature points are extracted using acquired data, and then data association is done using the geometrical characteristics of the calibration tool. Finally, the non-linear least square problem in (4) is solved using the augmented associated common feature point sets.

2.2. Calibration test and results

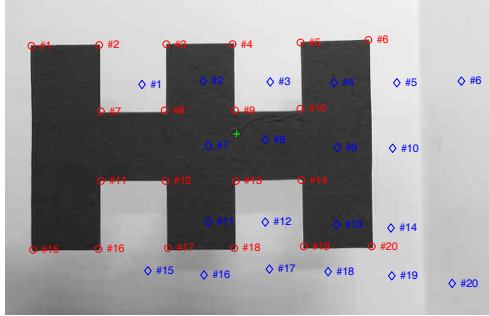
In order to verify performance of the presented calibration method, the calibration procedures were applied to real sensor measurements. The detailed test environments and working conditions are described in the section 4.1.

First, vision/lidar calibration was conducted by using calibration tool 1. After extracting feature points by Harris corner detection algorithm, data association was fulfilled with the geometric characteristics of the calibration tools, by numbering feature points which lie on edges of the plane in both image and lidar data from upper left to lower right.

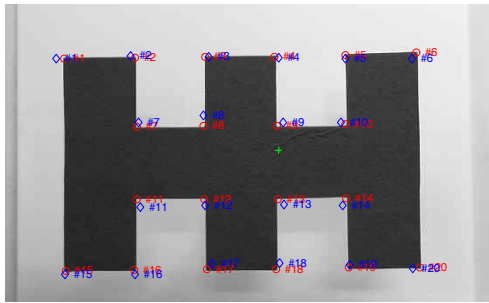
The estimation accuracy of the calibration parameters closely depends on measurement noises of a vision sensor and lidar, and position errors of the extracted feature points. Therefore, in order to increase the estimation accuracy,

Table 1. Vision/lidar calibration results

Calibration parameters	Values
Rotational angle (Euler angle from C_f^c , [degree])	[1.38 0.66 -0.95]
Translational vector (t_f^c , [m])	[-0.090 -0.017 -0.005]



(a) Before calibration



(b) After calibration

Fig. 3. Results of projecting feature points on lidar data to those on image domain

feature point set was augmented with additional feature point information, which was acquired in a situation the calibration tool 2 was further from the sensor module than the calibration tool 1, in other words, depth sensitivity of the feature point set was improved.

Finally, vision/lidar calibration was conducted with the augmented data set using both calibration tool 1 and 2. The estimated calibration parameters are summarized in Table 1.

In order to verify the properness of the estimated calibration parameters, it was applied to the acquired data using calibration tool 1. Fig. 3 (a) shows the results of projecting the feature points from lidar onto the image domain without calibration parameter information (that is, $C_f^c = \mathbf{I}$, $t_f^c = \mathbf{0}$). In the figure, the feature points on image data are represented by red circle mark ‘o’, while those on lidar data are represented by blue diamond mark ‘◇’. As seen in the figure, the matching errors between the point sets are significantly large due to lack of calibration parameter information. Fig. 3 (b) shows the projection results with the estimated calibration parameters. In the figure, it is clearly seen that all feature points are successfully matched.

Even though nonlinear optimization methods can produce highly reliable estimates, performance is directly affected by quality of a set of data points. In this calibration test, an average matching error is 7 pixels, which mainly arises inevitably from the low resolution of the lidar and the consequent errors in lidar feature extraction. Calibration parameter errors due to the data quality degrade performance of feature initialization, however that can be overcome with the iterative back-projection presented as in the following section.

3. Vision/Lidar-Based Feature Point Initialization for 6-dof Bearing-only SLAM

3.1 Feature point initialization

A feature point initialization is a process that determines three dimensional positions with the two dimensional coordinates. With the estimated calibration parameters, all points in pixel coordinate system can be transferred to three dimensional coordinates in camera coordinate system. If the calibration between two sensors was completed, (5) can be established.

$$\mathbf{p}_c^p = \mathbf{p}_l^p \quad (5)$$

A point extracted as a feature point in the pixel coordinate system, \mathbf{p}_c^p , can be represented by multiplication of the normalized vector of the transferred point in the camera coordinate system, \mathbf{p}_c^c , and the camera intrinsic matrix, \mathbf{K} , as shown in (6).

$$\mathbf{p}_c^p = \mathbf{K}(\mathbf{p}_c^c)_n \quad (6)$$

The normalized point, $(\mathbf{p}_c^c)_n$, can be obtained by rearranging the Eq. (6).

$$\mathbf{K}^{-1}\mathbf{p}_c^p = \begin{bmatrix} \mathbf{p}_{c,x}^c & \mathbf{p}_{c,y}^c & 1 \\ \mathbf{p}_{c,z}^c & \mathbf{p}_{c,z}^c & \end{bmatrix}^T \quad (7)$$

The purpose of the feature point initialization is to find \mathbf{p}_c^c itself. For this, new variables a and b are defined as (8).

$$a = \frac{\mathbf{p}_{c,x}^c}{\mathbf{p}_{c,z}^c}, \quad b = \frac{\mathbf{p}_{c,y}^c}{\mathbf{p}_{c,z}^c} \quad (8)$$

\mathbf{p}_c^c can be expressed by introducing a scaling parameter, k , as below.

$$\mathbf{p}_c^c = k[a \quad b \quad 1]^T \quad (9)$$

Then, the point in the lidar coordinate system, \mathbf{p}_l^l , can

be computed by using the estimated calibration parameters. At this step, the offset between the sensing point and the origin of the lidar coordinate system is compensated as (10).

$$\mathbf{p}_c^l = \mathbf{C}_c^l \left(k \begin{bmatrix} a & b & 1 \end{bmatrix}^T - \mathbf{t}_l^c \right) - \begin{bmatrix} 0 & d_y & d_z \end{bmatrix}^T \quad (10)$$

To find the scaling parameter, k , first, roughly estimated $\hat{\mathbf{p}}_c^l$ is computed by applying an arbitrary \hat{k} .

$$\hat{\mathbf{p}}_c^l = \mathbf{C}_c^l \left(\hat{k} \begin{bmatrix} a & b & 1 \end{bmatrix}^T - \mathbf{t}_l^c \right) - \begin{bmatrix} 0 & d_y & d_z \end{bmatrix}^T \quad (11)$$

In order to find the closely matched lidar measurements with the above $\hat{\mathbf{p}}_c^l$, the estimated azimuth angle, $\hat{\psi}_l$, and the elevation angle, $\hat{\theta}_l$, are computed as (12) and (13).

$$\hat{\psi}_l = \arctan_2(\hat{\mathbf{p}}_{l,x}^l, \hat{\mathbf{p}}_{l,z}^l) \quad (12)$$

$$\hat{\theta}_l = \arctan \left(\frac{-\hat{\mathbf{p}}_{l,y}^l}{(\hat{\mathbf{p}}_{l,x}^l)^2 + (\hat{\mathbf{p}}_{l,z}^l)^2} \right) \quad (13)$$

With the estimated azimuth and elevation angles, the range measurement, \hat{l}_l , can be found, thus, the point in the Cartesian coordinate system, $\hat{\mathbf{p}}_l^l$, is obtained. Next, reprojecting $\hat{\mathbf{p}}_l^l$ to the pixel coordinate system yields $\hat{\mathbf{p}}_l^p$, and by making feedback of the difference between $\hat{\mathbf{p}}_l^p$ and the extracted feature point, \mathbf{p}_c^p , the scaling parameter, \hat{k} , is updated. Repeating the process of computing \hat{k} and $\hat{\mathbf{p}}_l^p$ can provide more accurate $\hat{\mathbf{p}}_l^l$, which is referred to an automatic scaling parameter tuning technique in our work. Finally, the three dimensional position in the camera coordinate system is computed as follows.

$$\hat{\mathbf{p}}_c^c \approx \hat{\mathbf{p}}_l^c = \mathbf{C}_c^c \mathbf{p}_l^l + \mathbf{t}_l^c \quad (14)$$

3.2 Initialization test and results

For verifying the performance of the presented feature point initialization method, the method was applied to the acquired data using calibration tool 2. After choosing arbitrary points, \mathbf{p}_c^p , the three dimensional positions in the camera coordinate system of the selected points were computed via the presented feature initialization method.

Arbitrary points was chosen and represented by red circle mark 'o'. The matched points on lidar data, \mathbf{p}_l^l , computed by the initialization method were represented by blue diamond mark '◇'.

Fig. 4 shows the results of the matching point updates as \hat{k} automatically changes in the case of one of the selected feature points.

According to the iteration numbers, which were represented by 'i=iteration number', the localized points were represented by green 'x', and the corresponding

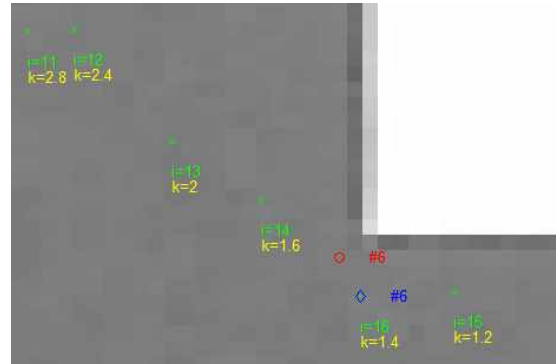


Fig. 4. Updates of matched feature points on lidar data in pixel coordinates according to \hat{k} changed automatically

scaling parameters, \hat{k} , were written by 'k= \hat{k} '.

3.3 Practical use of initialization in 6-DoF SLAM

The EKF-based SLAM is a well-known approach in the SLAM problem and is represented in the form of a state space equation with additive Gaussian noises. For 3-DoF vehicles, the process model that expresses the motion of the vehicle is represented as (15) [25].

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_k) + \mathbf{w}_k \quad (15)$$

In the above equation, \mathbf{x}_k is an augmented state vector, which consists of the position and heading information of the vehicle and a number of feature point positions. \mathbf{u}_k is a control input vector applied at $k-1$ time, which propagates \mathbf{x}_{k-1} to \mathbf{x}_k . $f(\cdot)$ represents the vehicle kinematics, and \mathbf{w}_k denotes an additive zero mean uncorrelated Gaussian noise with covariance \mathbf{Q}_k . The measurement model is represented as (16).

$$\mathbf{z}_k = h(\mathbf{x}_k) + \mathbf{v}_k \quad (16)$$

In (16), $h(\cdot)$ represents the relationship between the state vector and the measurement vector, and \mathbf{v}_k denotes an additive zero mean uncorrelated Gaussian noise with covariance \mathbf{R}_k . With the above models, SLAM can be demonstrated using the standard EKF framework [26-27].

This paper focuses on 6-DoF bearing-only SLAM based on inertial sensors. The system consists of a low-cost IMU whose performance is under tactical grade and bearing sensors. For relative navigation without previously known information like landmark, the performance of the navigation system heavily depends on the inertial bias errors. Therefore, the augmented state vector is defined as (17).

$$\mathbf{x} = \left[\delta \mathbf{P}^n \quad \delta \mathbf{V}^n \quad \delta \Psi \quad \delta \mathbf{b}_a \quad \delta \mathbf{b}_g \quad \delta \mathbf{m}_1^n \quad \dots \quad \delta \mathbf{m}_j^n \right] \quad (17)$$

In (17), $\delta \mathbf{P}^n$ and $\delta \mathbf{V}^n$ represent the position and velocity errors of the vehicle in the navigation frame, and $\delta \mathcal{P}$ represents the attitude error in the body frame. The bias errors of acceleration and angular rate are denoted by $\delta \mathbf{b}_a$ and $\delta \mathbf{b}_g$, respectively, and the position error of the j^{th} feature point is represented by $\delta \mathbf{m}_j^n$. In addition, the INS error model, which is based on the psi-angle approach [28-29], is used as the process model, where the errors of sensor bias and feature point position are modeled as random constants.

The azimuth and elevation angles of a feature point are the components in the measurement vector for bearing-only SLAM framework [26-27]. In this system, the measurement model is generally represented as in (18), in which the measurement noise \mathbf{v} is assumed to be an additive zero mean uncorrelated Gaussian noise with covariance \mathbf{R} .

$$\mathbf{z} = \begin{bmatrix} \varphi \\ \mathcal{E} \end{bmatrix} = h(\mathbf{x}) + \mathbf{v} = \begin{bmatrix} \tan^{-1}\left(\frac{y^c}{x^c}\right) \\ \tan^{-1}\left(\frac{z^c}{\sqrt{(x^c)^2 + (y^c)^2}}\right) \end{bmatrix} + \mathbf{v} \quad (18)$$

where $\mathbf{m}^c = [x^c \ y^c \ z^c]^T = \mathbf{C}_b^c (\mathbf{C}_n^b (\mathbf{m}^n - \mathbf{p}^n) - \mathbf{p}_c^b)$

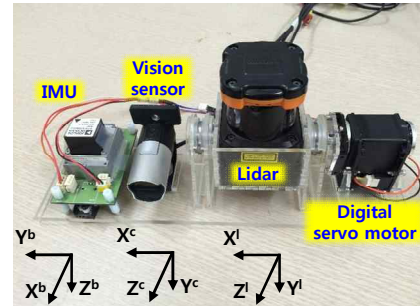
In (18), the azimuth and elevation angles are denoted by φ and \mathcal{E} , respectively. \mathbf{m}^c is the position of a feature point in the camera coordinate system, computed using the vehicle position \mathbf{p}^n , position of the feature point \mathbf{m}^n in the navigation coordinate system, lever arm vector \mathbf{p}_c^b between the camera and the center of the vehicle body, and direction cosine matrices, namely the vehicle attitude \mathbf{C}_b^n and camera attitude \mathbf{C}_b^c . The feature point initialization method which this paper is concerned with is utilized to determine the initial positions of feature points.

4. Experiments

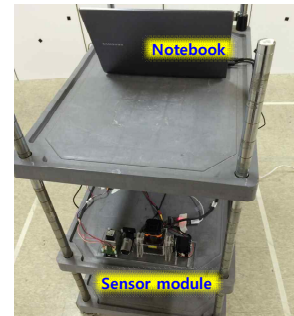
For verifying the performance of the presented feature point initialization method, vision/lidar calibration was done with real sensor measurements, and three-dimensional positions of feature points were estimated by using the calibration parameters in section 2.2 and 3.2. In this section, based on the calibration and initialization results, navigation performance of the 6-DoF bearing-only SLAM was compared with that based on the existing delayed initialization method [1].

4.1. Experimental environments

In order to acquire real sensor measurements, an integrated sensor module consisting of an IMU (Inertial Measurement Units), a vision sensor and a two-dimensional lidar was constructed and a notebook-based data acquisition



(a) Integrated sensor module



(b) DAQ system on a moving cart

Fig. 5. Data acquisition system

system was implemented. As a moving cart mounting the DAQ system moves along the specific trajectory, experimental data was acquired. In this test, three-dimensional lidar measurements was generated by rotating the two dimensional lidar. In order to do so, one-axis rotating system, whose rotation range is ± 30 degrees, operating by a servo motor was utilized. At the calibration stage, rotating speed was set to 20 degree/s in order that lidar data has a minimum detectable angle of 0.5 degrees in a vertical direction, and at the navigation stage, rotating speed was set to 64 degree/s. Output rates of the IMU, vision sensor, and lidar were set to 100, 30, 40 Hz, respectively. Sensing axes of the vision sensor and lidar (Z^c and Z^l) were parallel to the forward axis of the IMU (X^b). The origin of the lidar coordinates is the point at which the sensing point and the rotational axis intersect perpendicularly and thus the sensing point is located at -0.0405 meters in the direction of Y^l -axis from the origin. Fig. 5 shows the integrated sensor module and its installation on a moving cart.

The specifications of the sensors are presented in Table 2.

For vision/lidar calibration, sensor measurements were acquired with calibration tool 1 and 2. In our work, Harris corner detection algorithm was utilized to extract feature points in images [30].

For dynamic test, an artificial wall whose dimension was 3.15 by 4.3 by 0.9 meters was constructed. A number of black square patches were randomly attached on the wall in order to easily extract feature points. Trajectory of the

Table 2. Sensor specifications

Sensor	Model	Main features	
IMU	ADIS16385 (Analog Devices)	Gyro. initial bias error	$\pm 3 \text{ }^\circ/\text{s}$ ($\pm 1\sigma$)
		Gyro. bias stability	Z-axis: $6 \text{ }^\circ/\text{h}$ (1σ) X, Y-axis: $21 \text{ }^\circ/\text{h}$ (1σ)
		Acc. initial bias error	$\pm 10 \text{ mg}$ ($\pm 1\sigma$)
		Acc. bias stability	$50 \text{ } \mu\text{g}$ (1σ)
Vision sensor	LifeCam Studio (Microsoft)	Resolution	640 x 480
		Type	Color
Lidar	UTM-30LX (Hokuyo)	Scan angle	270°
		Detection range	0.1~30 m
		Angular resolution	0.25°
		Range resolution	0.001 m

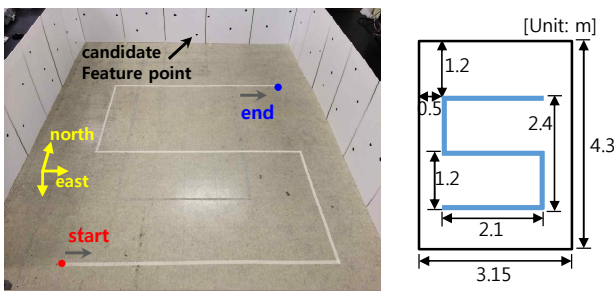


Fig. 6. Experimental environments

moving cart and experimental environments are depicted in Fig. 6.

Although the test has a limitation that true position information cannot be obtained due to manual motion input, it is assumed that the error of the reference trajectory is below about 0.2 meters. Thus, for the accurate manual motion, the trajectory was constructed by combining straight lines. Basically, the test aims to show availability improvement of feature points with the proposed vision/lidar fusion, and their effect on the SLAM performance. Therefore, in the author’s opinion, the indoor simple test is enough to verify the purposes. At the initial navigation stage, INS initial alignment is conducted during 17 seconds, and then the moving cart moves along the reference trajectory until 34 seconds, and finally keeps stationary condition until 39 seconds. No any intentional attitude input was applied in order to maintain tracking time of feature points as long as possible. In this test, KLT (Kanade-Lucas-Tomasi) algorithm was utilized to detect and track multiple feature points [31].

4.2. Experimental results

In this test, when the presented feature point initialization method was applied, the scaling parameter was set to a fixed value, which is an average value considering test environments. The automatic scaling parameter tuning technique in the previous section can be actually utilized once three dimensional lidar data is constructed. For this, the motion change of a vehicle should be considered when

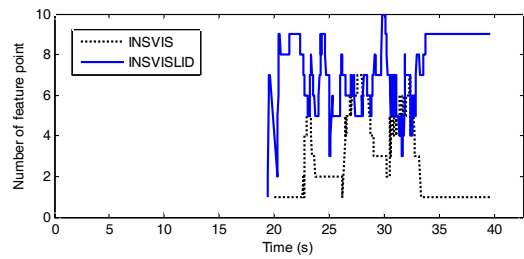
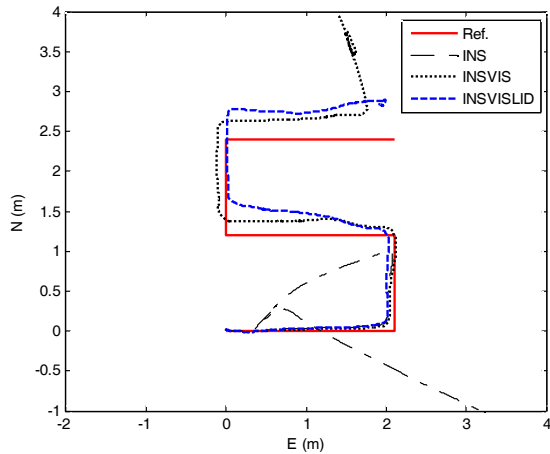


Fig. 7. The number of initialized feature points

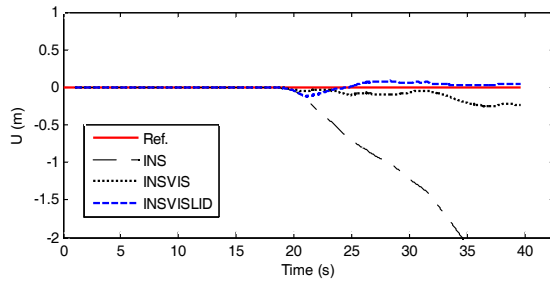
constructing the database. However, because of the tradeoff between the rotation speed and the vertical angle resolution, the automatic scaling parameter tuning method cannot be utilized, therefore, the feature points close to the emitted ray of the lidar within a certain angular threshold are only initialized.

The maximum number of tracked feature points was set to ten. A number of black square patches as candidate feature points were attached uniformly in order to easily track feature points, consequently, ten tracked points was maintained during all periods. The presented feature point initialization method has the advantage of improving availability of the feature points in a SLAM navigation filter. Thus, the navigation performance of a bearing-only SLAM using the presented feature initialization method were compared with that using the existing delayed initialization method. Fig. 7 shows the number of initialized feature points depending on the initialization methods. In the case of using the delayed initialization method, the number of initialized points was represented by ‘INSVIS’ and varies from one to seven, including noticeable periods of having points below two. Whereas, the number of initialized feature points of the SLAM based on the presented initialization method, represented by ‘INSVISLID’ in the figure legend, varies from one to ten, therefore, improving availability of the feature points.

Fig. 8 (a) shows horizontal position estimation results. In the case of INS, the position error gets larger from the point (north 1m, east 2m) and diverges rapidly. On the other hand, two SLAM systems provide considerably accurate estimates than INS. Although those seem to be having almost identical estimation performance during the dynamic periods considering the error of the reference trajectory, the error of ‘INSVIS’ becomes larger after the dynamic motion stops. This results from a lack of the feature point availability. As seen in the above fig. 0, For ‘INSVISLID’, nine feature points can be used for SLAM, whereas, for ‘INSVIS’, the only one can be used during the final stationary periods. Fig. 8 (b) shows altitude estimation results, similar to the case of the horizontal position estimation. Considering no motion inputs are applied in the direction of up-axis, two SLAM systems provide much more accurate estimates than INS. However, it is remarkable that ‘INSVISLID’ provide stable estimation performance even during the stationary periods after 34s.



(a) Horizontal position



(b) Altitude

Fig. 8. Results of position estimation

By applying the presented feature point initialization method to a bearing-only SLAM, it was hard to verify the notable navigation performance improvements as compared with the existing system during the dynamic periods. That is because the quality of feature point information between the two cases is not much different, and also the simulation time is somewhat short. Rapid attitude change or high-dynamic motion of the moving cart can reduce the tracking time of feature points moving in FOV (Field Of View) of the vision sensor, that is, feature point availability can get extremely lower. Consequently, feature points can disappear in FOV just as soon as feature points are initialized, or even before the initialization when the delayed initialization method is used. For this reason, the test scenarios and environments were constructed considering that the SLAM using the delayed initialization method could operate. Therefore, the more dynamic the motion of the moving cart is, the more effective applying the proposed feature point initialization method to the SLAM can be.

Fig. 9 shows velocity estimation results. Considering estimation performance during the stationary periods, ‘INSVISLID’ provides more accurate velocity estimates. Especially, the relatively large error in the north velocity comes from the characteristics of having low sensitivity of the vision sensor in the direction of optic axis.

As aforementioned in section 4.1, the true position and

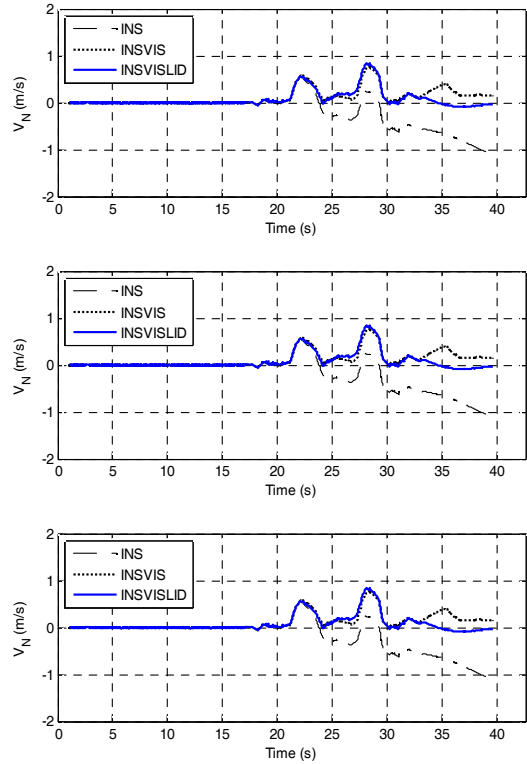


Fig. 9. Results of velocity estimation



Fig. 10. System installation and experimental environments

Table 3. RMSE analysis results

System	Position (m)			Velocity (m/s)		
	North	East	Down	North	East	Down
INSVIS	0.171	0.065	0.021	0.089	0.034	0.023
INSVISLID	0.018	0.020	0.005	0.032	0.010	0.010

velocity cannot be obtained during the dynamic periods. However, it is obvious that true velocity is zero after the dynamic motion ends. In this respect, RMSE (Root-Mean-Square Error) values during the stationary periods were compared as shown in Table 3.

In order to analyze estimation performance during dynamic motion periods, outdoor test was also conducted with an accurate reference navigation system of SPAN® UIMU-HG1700 (Novatel Inc.). Fig. 10 shows the system installation and experimental environments.

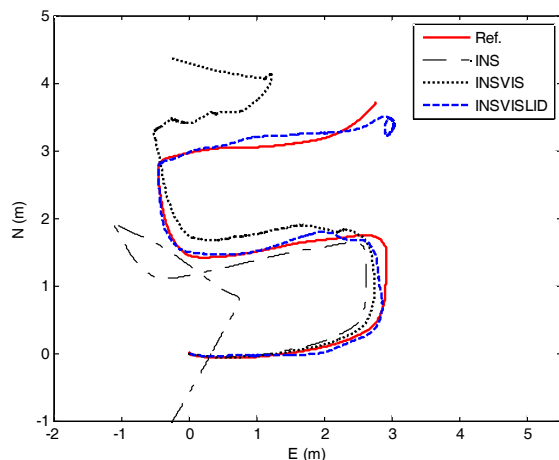


Fig. 11. Horizontal position estimation results

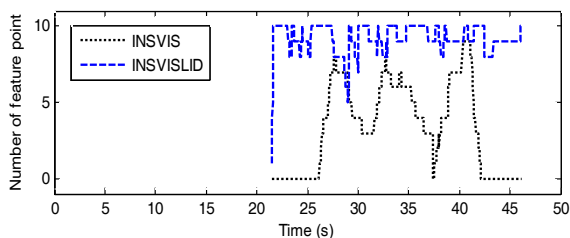


Fig. 12. The number of initialized feature points

Fig. 11 shows horizontal position estimation results. In contrast with the indoor test, it is obvious that ‘INSVISLID’ provide reliable estimation performance during the whole time. Three dimensional position errors (RMSE) of ‘INSVISLID’ and ‘INSVIS’ were 0.18m and 0.87m, respectively. As a result, it was proven that the presented system shows performance improvement more than four times due to the availability improvement of feature points as seen in Fig. 12.

5. Conclusion

This paper proposed a novel method of finding three dimensional position of feature points in order to improve performance of a 6-DoF bearing-only SLAM. For this, a calibration method which estimates the rotational and translational relationship between a lidar and a vision sensor was presented, so all points in the lidar coordinate system could be transferred to points in the pixel coordinate system. With the estimated calibration parameters, the three dimensional positions of the two dimensional coordinates in the pixel coordinate system were determined in the manner of performing the calibration procedure reversely. In this process, automatic scaling parameter tuning technique was presented in order to improve the accuracy of the positions.

For verifying performance of the proposed feature point

initialization method, navigation performance of the 6-DoF bearing-only SLAM based on the method was compared with that based on the existing delayed initialization method. As a result of the test, navigation performance was enhanced by using the proposed initialization method due to the improvements of the availability of feature points. In the indoor test, considering that true velocity is zero after the dynamic motion ends, stationary performance was analyzed. For analyzing dynamic performance, outdoor test was also conducted. As a result, it was proven that the presented system shows performance improvement more than four times due to the availability improvement of feature points.

Although attitude errors could be obtained in the outdoor test, those were not suitable to do error analysis because the magnitude was as small as the installation angle error (‘INSVISLID’: 0.59deg., ‘INSVIS’: 1.08deg., [3d RMSE]). In order to analyze performance of attitude estimation, it is necessary to expand the trajectory scale of the moving cart and apply various motion inputs to the cart for improving observability in SLAM filter.

Considering the tracking time become shorter as the motion of a vehicle is more dynamic, a SLAM for high dynamic air vehicles can be realized based on the proposed method.

Acknowledgements

This research was supported by the Korea Institute for Advancement of Technology(MOTIE, No.R0004086) and National Research Fund(NRF-2016R1A2B4010719).

References

- [1] T. Bailey, “Constrained initialisation for bearing-only SLAM,” in *IEEE International Conference on Robotics and Automation*, 2003, pp. 1966-1971.
- [2] A. J. Davison, “Real-time simultaneous localisation and mapping with a single camera,” in *International Conference on Computer Vision.*, 2003, vol. 2, pp. 1403-1410.
- [3] A. J. Davison, Y. G. Cid, and N. Kita, “REAL-TIME 3D SLAM WITH WIDE-ANGLE VISION,” in *IFAC/EURON Symposium on Intelligent Autonomous Vehicles*, 2004.
- [4] N. M. Kwok and G. Dissanayake, “An efficient multiple hypothesis filter for bearing-only SLAM,” *2004 IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IEEE Cat. No.04CH37566)*, pp. 736-742, 2004.
- [5] C. Premebida, O. Ludwig, and U. Nunes, “LIDAR and Vision-Based Pedestrian Detection System,” *J. F. Robot.*, vol. 26, no. 9, pp. 696-711, 2009.
- [6] L. Huang and M. Barth, “Tightly-Coupled LIDAR

- and Computer Vision Integration for Vehicle Detection,” in *IEEE Intelligent Vehicles Symposium*, 2009, pp. 604-609.
- [7] P. Moghadam, W. S. Wijesoma, and D. J. Feng, “Improving Path Planning and Mapping Based on Stereo Vision and Lidar,” in *International Conference on Control, Automation, Robotics and Vision*, 2008, no. December, pp. 384-389.
- [8] “Kinect Specification.” [Online]. Available: <https://dev.windows.com/en-us/kinect/hardware>.
- [9] “DepthSense Camera Specification.” [Online]. Available: <http://www.softkinetic.com/Products/DepthSenseCameras>.
- [10] “Basler ToF Camera Specification.” [Online]. Available: https://www.baslerweb.com/media/documents/BAS1409_ToF_EN_web.pdf.
- [11] D. Wilson, “Time of Flight Cameras Take on Industrial Applications,” *Novus Light Technologies Today*, 15-Jun-2015.
- [12] G. Cui, J. Wang, and J. Li, “Robust multilane detection and tracking in urban scenarios based on LIDAR and mono-vision,” *IET Image Process.*, vol. 8, no. 5, pp. 269-279, 2014.
- [13] Q. Li, L. Chen, M. Li, S. L. Shaw, and A. Nüchter, “A sensor-fusion drivable-region and lane-detection system for autonomous vehicle navigation in challenging road scenarios,” *IEEE Trans. Veh. Technol.*, vol. 63, no. 2, pp. 540-555, 2014.
- [14] A. Cherubini, F. Spindler, and F. Chaumette, “Autonomous Visual Navigation and Laser-Based Moving Obstacle Avoidance,” *IEEE Trans. Intell. Transform. Syst.*, vol. 15, no. 5, pp. 2101-2110, 2014.
- [15] T. Ju, W. Lu, K. Chen, and J. Guo, “Vision-Based Moving Objects Detection for Intelligent Automobiles and a Robustness Enhancing Method,” in *2014 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW)*, 2014, vol. 1, pp. 75-76.
- [16] Q. Zhang and R. Pless, “Extrinsic Calibration of a Camera and Laser Range Finder (improves camera calibration),” in *IEEE/RJS International Conference on Intelligent Robot Systems*, 2004, pp. 2301-2306.
- [17] L. Huang and M. Barth, “A Novel Multi-Planar LIDAR and Computer Vision Calibration Procedure Using 2D Patterns for Automated Navigation,” in *IEEE Intelligent Vehicles Symposium*, 2009, pp. 117-122.
- [18] O. Naroditsky, A. P. Iv, and K. Daniilidis, “Automatic Alignment of a Camera with a Line Scan LIDAR System,” in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 3429-3434.
- [19] R. Unnikrishnan and M. Hebert, “Fast Extrinsic Calibration of a Laser Rangefinder to a Camera,” in *Tech. Report CMU-RI-TR-05-09, Robotics Institute*, 2005.
- [20] G. Pandey, J. R. Mcbride, S. Savarese, and R. M. Eustice, “Automatic Targetless Extrinsic Calibration of a 3D Lidar and Camera by Maximizing Mutual Information,” in *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012, pp. 2053-2059.
- [21] P. Moghadam, M. Bosse, and R. Zlot, “Line-based extrinsic calibration of range and image sensors,” in *IEEE International Conference on Robotics and Automation*, 2013, no. c, pp. 3685-3691.
- [22] D. Scaramuzza, A. Harati, and R. Siegwart, “Extrinsic self calibration of a camera and a 3D laser range finder from natural scenes,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007, pp. 4164-4169.
- [23] S. Hur, D. H. Shim, and J. Kim, “Integrated navigation system using camera and gimbaled laser scanner for indoor and outdoor autonomous flight of UAVs,” *2013 IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pp. 3158-3163, Nov. 2013.
- [24] J. J. Moré, “The Levenberg-Marquardt algorithm: Implementation and theory,” in *Proceedings of the Biennial Conference*, 1977, pp. 105-116.
- [25] H. Durrant-whyte and T. Bailey, “Simultaneous Localization and Mapping: Part I,” *IEEE Robot. Autom. Mag.*, vol. 13, no. 2, pp. 99-108, 2006.
- [26] M. W. M. G. Dissanayake, P. Newman, S. Clark, H. F. Durrant-whyte, and M. Csorba, “A Solution to the Simultaneous Localization and Map Building (SLAM) Problem,” *IEEE Trans. Robot. Autom.*, vol. 17, no. 3, pp. 229-241, 2001.
- [27] J. Kim and S. Sukkarieh, “Autonomous airborne navigation in unknown terrain environments,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 3, pp. 1031-1045, 2004.
- [28] D. Benson, “A Comparison of Two Approaches to Pure-Inertial and Doppler-Inertial Error Analysis,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-11, no. 4, pp. 447-455, Jul. 1975.
- [29] S. Yun, Y. J. Lee, C. J. Kim, and S. Sung, “Integrated Navigation Design Using a Gimbaled Vision/LiDAR System with an Approximate Ground Description Model,” *Int. J. Aeronaut. Sp. Sci.*, vol. 14, no. 4, pp. 369-378, Dec. 2013.
- [30] C. Harris and M. Stephens, “A Combined Corner and Edge Detector,” *Proceedings Alvey Vis. Conf.* 1988, pp. 147-151, 1988.
- [31] C. Tomasi and T. Kanade, “Detection and Tracking of Point Features,” in *Technical Report CMU-CS-91-132, Carnegie Mellon University*, 1991, no. April.



Sukchang Yun He received his B.S. and Ph.D. degrees in Aerospace Information Engineering from Konkuk University, Seoul, Korea, in 2009 and 2015, respectively. His research interests include INS, sensor fusion, Lidar/Vision based navigation algorithm.



Byoungjin Lee He received his B.S. degree in Aerospace Information Engineering from Konkuk University, Seoul, Korea, in 2010. He is currently pursuing the Ph.D degree in Konkuk University. His research interests include the integrated avionics, indoor positioning, navigation and control system for unmanned vehicles.



Yeon-Jo Kim He received B.S degree in Aerospace Engineering from Sejong University, Seoul, Korea. His research interests are vision-based navigation and obstacle avoidance.



Young Jae Lee He is a Professor in the Department of Aerospace Information Engineering at Konkuk University. His research interests include integrity monitoring of GNSS signal, GBAS, RTK, attitude determination, orbit determination, and integrated navigation-related engineering problems.



Sangkyung Sung He is an Associate Professor of the Department of Aerospace Information Engineering, Konkuk University. His research interests include inertial sensors, integrated navigation and its application to mechatronics and unmanned systems.