2023 The 23rd International Conference on Control, Automation and Systems (ICCAS 2023)
Yeosu Venezia Hotel & Utop Marina Hotel, Yeosu, Korea, Oct. 17∼20, 2023

Indoor Pedestrian-Following System by a Drone with Edge Computing and Neural Networks: Part 1 - System Design

In-Chan Ryu[1], Jung-Il Ham[2], Jun-Oh Park[3], Jae-Woo Joeng[4], Sung-Chang Kim[5], Hyo-Sung Ahn[6,†]

[1,2,3,4,6]School of Mechanical Engineering, Gwangju Institute of Science and Technology, Gwangju, 61005, Korea
([1]inchanryu@gm.gist.ac.kr, [2]jungilham@gm.gist.ac.kr, [3]junoingist@gm.gist.ac.kr, [4]ju.jeong@gm.gist.ac.kr)
([6,†]hyosung@gist.ac.kr)

[5]Edge Computing Application Service Research Section, Electronics and Telecommunications Research Institute Gwangju, 61011, Korea
([5]sungchang@etri.re.kr)

Abstract: As the drone market continues to expand, the need for accurately determining a drone's position and orientation using a camera in GPS-denied environments becomes increasingly critical. This paper aims to achieve precise position and attitude data by incorporating SLAM to provide visual measurements for EKF, thereby ensuring the stability of drone operations. An experiment was conducted to execute commands from the ground control PC using the map created as a result of SLAM. The primary tools used for this purpose included the Pixhawk Orange, Jetson Nano, and the ZED-Mini camera. The research showcases the effectiveness of these tools and methods in enhancing indoor drone functionality.

Keywords: VPS, Simultaneous Localization and Mapping, Sensor Fusion, Extended Kalman Filter, Indoor Navigation, Drone

1. INTRODUCTION

In the rapidly growing drone market, which has seen an average annual growth rate of 29% over the past decade, there is a notable shift in interest towards drones equipped with autonomous flight capabilities for mission execution, rather than basic, low-priced drones primarily used for simple photography and videography.

The evolution of drone technology is driven by market demands and advancements in positioning sensors, control theory, and related fields. Drones are now applied across a diverse range of industries, from aerial photography and agricultural pesticide control to long-distance communication and versatile military applications.

In these diverse applications, autonomous flight is a fundamental requirement. Achieving this autonomy necessitates seamless exchange of location information between ground control centers and drones [1]. In outdoor environments, drones efficiently utilize GPS and IMU data from ground control centers to establish reliable location information exchange and tracking. However, challenges arise in GPS-denied environments like indoor spaces and areas with significant obstructions. In such settings, unstable location information transmission and reception can render autonomous flight unattainable. Furthermore, many drone navigation technologies primarily rely on precise outdoor sensors like GPS, limiting their effectiveness indoors.

To address the challenge of stable flight in GPS-denied environments, the drone industry has increasingly adopted Vision Positioning System (VPS) technology. This approach combines vision sensors, typically cameras, with existing sensors to enhance navigation capabilities [2]. By leveraging common feature points captured by onboard cameras, along with drone IMU data and movement measurements, drones can estimate their position accurately, thereby compensating for the limitations of traditional positioning sensors [3].

However, it's important to note that research into indoor positioning technologies using vision sensors, such as 3D mapping and VPS, has primarily been conducted for automotive applications, with relatively limited exploration in the context of indoor drones.

In this paper, we introduce a drone system that leverages the fusion of vision information and IMU data to achieve precise positioning within indoor environments. This work represents Part 1 of a series, and its structure is as follows: Section 2 introduces a flight control system with a specific focus on the Extended Kalman filter (EKF). Section 3 delves into Simultaneous Localization and Mapping (SLAM) techniques, with an emphasis on loop closing. The experimental setup and analysis are presented in Section 4, and Section 5 offers a comprehensive conclusion.

Part 2 [4] of this series will build upon the concepts explored in Part 1 and will focus on the development of a human-tracking drone system. It will incorporate additional technologies, including YOLO-v3 (You Only Look Once) object detection, and our proprietary monocular depth estimation technique, to further enhance drone capabilities and performance.

2. FLIGHT CONTROL SYSTEM

In this section, The navigation filter was designed for indoor drone operation. Specifically, the navigation filter employs the Extended Kalman Filter (EKF), a widely uti-

† Corresponding Author

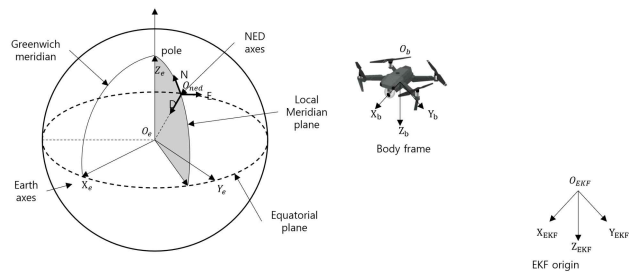

Fig. 1.: ECEF Coordinate System with Origin O_e and Axis X_e, Y_e, Z_e, NED frame with origin O_{ned} and axis N, E, D, body frame with origin O_b, center of mass of drone and axis X_b, Y_b, Z_b, EKF-Origin frame with origin O_{EKF} and axis $X_{EKF}, Y_{EKF}, Z_{EKF}$, orientation follows right-handed rule

lized method in drone navigation. The output of SLAM is used as a measurement input for the EKF to provide compensation. Further details about SLAM will be covered in the subsequent chapter. The EKF utilized here is based on popular open-source drone platforms, namely Ardupilot and PX4, and detailed information about the mechanism can be found in [5]. While the equations and applications for the drone are commonly understood, the task of tuning various design parameters is typically left to engineers, which can significantly impact system performance.

2.1 Reference Frame

A reference frame is imperative for drone control. An inertial frame where Newton's Law is conserved is used to describe the dynamics of drones. While various frames exist for describing Earth-relative motion, the Earth-Centered Earth-Fixed Frame (ECEF), which disregards Earth's rotation, is employed due to the drone's limited operational time and distance. Generally, the navigation system is formulated and computed in the NED frame. Meanwhile, the body frame is affixed to the drone's structure, originating from its center of mass. All sensor measurements affixed to the body are transformed into this body frame before fusion. Notably, Ardupilot's navigation system introduces an additional frame called EKF-origin, enhancing users' intuitive grasp of the drone's position [6]. The mentioned frames are illustrated in Fig. 1.

2.2 Extended Kalman Filter

For the purpose of localizing the drone's position and attitude, we employ an Extended Kalman Filter (EKF). This filter effectively combines data from the IMU, downward-facing rangefinder, and visual odometry. Given the extensive utilization of EKF in localizing autonomous vehicles, its equations and implementation have become well-established. However, the task of defining design parameters, such as the process noise covariance matrix Q and the measurement noise covariance matrix R, remains in the user's hand. This demands prior expert knowledge in signal processing and a comprehensive grasp of the underlying physical system.

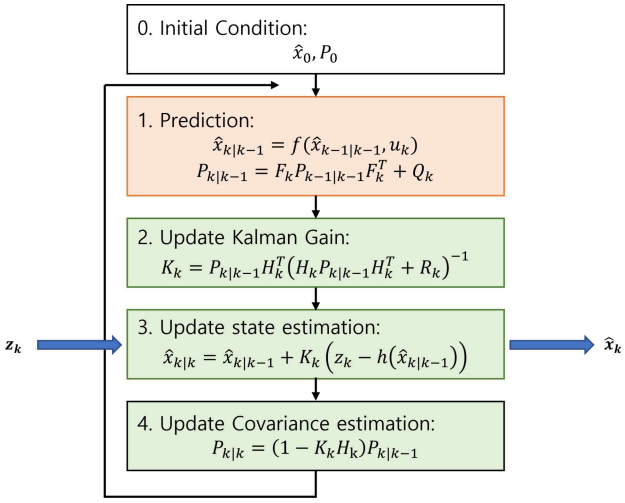

Fig. 2.: Block Diagram of the EKF, subscript k means discrete time k. $\hat{x}_0$ and P_0 are the initial estimated values of the state vector covariance matrix, $\hat{x}_{k|k}$ and $\hat{x}_{k|k-1}$ are updated and predicted state estimate, respectively, $P_{k|k}$ and $P_{k|k-1}$ are updated and predicted covariance estimate, respectively, K_k is Kalman gain, F_k and H_k are Jacobian of f and h respectively.

2.2.1 EKF algorithm

The System equation and measurement equation for EKF is shown as below.

$$\dot{x}_{k+1} = f(x_k, u_k) + w_k$$
$$y_k = h(x_k) + v_k$$

where, $x = \begin{bmatrix} q & v_{NED} & P_{NED} & \Delta\theta_{bias} & \Delta vel_{bias} \end{bmatrix}^T$ is the state vector, q is quaternion rotation from the body frame to the NED frame. P_{NED} and V_{NED} denote the drone's position and velocity in the NED frame respectively.

The system equation $f(x_k, u_k)$ and measurement equation $h(x_k)$ are described in [7], alongside the treatment of IMU data. Here, $\Delta\theta_{bias}$ and Δvel_{bias} represent changes in the drone's angle and velocity over the sampling period. The process noise, denoted as w, possesses a zero mean, variance of Q, and exhibits white noise characteristics. Similarly, the measurement noise, denoted as v, is represented as white noise with a zero mean and variance of R. A visual representation of the EKF algorithm is illustrated in Fig. 2 in the form of a block diagram.

2.2.2 design parameters

The design parameters of EKF are listed in Table. 1 In Table 1, M_NSE and P_NSE represent measurement noise and process noise, respectively. VISO stands for visual odometry.

2.3 Control System

For the drone control system, the controller of Ardupilot is used with a fine-tuned gain parameter for the custom drone. The control diagram and detailed information of the controller are described at [8].

Table 1.: Design Parameters for the EKF

variable	value	variable	value
EK3_POSNE_M_NSE	0.5	EK3_ALT_M_NSE	2
EK3_VELNE_M_NSE	0.5	EK3_VELD_M_NSE	0.3
EK3_RNG_M_NSE	0.5	EK3_YAW_M_MSE	0.5
VISO_POS_M_NSE	0.1	VISO_YAW_M_NSE	0.087266
EK3_GYRO_P_NSE	0.015	EK3_ACC_P_NSE	0.003
EK3_GBAIS_P_NSE	0.001	EK3_ABIAS_P_NSE	0.003

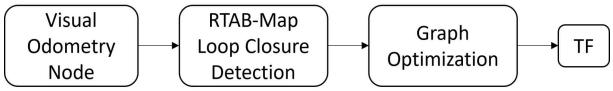

Fig. 3.: Visual Representation of Loop Closing Process

3. SLAM

3.1 Loop Closing

In our approach, we employed RTAB-Map (Real-Time Appearance-Based Mapping), a graph-based SLAM technique, to perform loop closure and global mapping [9].

RTAB-Map takes an estimated vision-based position, obtained without loop closure, as its input. We used the vision position provided by the ZED-mini camera. However, this vision-based position alone is insufficient for performing global 3D mapping and loop closure. This is where global SLAM, such as RTAB-Map, comes into play. RTAB-Map is a classic GraphSLAM method designed to accomplish global mapping and detect previously visited locations. GraphSLAM tackles the graph optimization problem by consolidating information and providing the best possible estimate, considering all the collected data. This approach enables the creation of a more accurate map compared to maps generated using other methods, such as EKFSLAM [10].

The process involves graph optimization, which aims to minimize the discrepancy between the estimated location and the location observed by the sensors. The result of this optimization is conveyed as a ROS TF (Transformation), which contains information about the transformation from the drone's starting point to its current position. Figure 3 provides a simplified schematic illustrating the process, detailing how visual odometry data from the ZED camera undergoes the SLAM process and is output as a TF measurement used for EKF estimation.

The remainder of this section is about how this optimization works for vision position updates.

3.2 Graph Optimization

The concept briefly outlined in this section is closely aligned with the approaches described in [11]. Let e_{ij} represent a function that computes the discrepancy between the expected observation $\hat{z_{ij}}$ and the actual observation z_{ij}.:

$$e_{ij} = z_{ij} - \hat{z}_{ij}$$

Here, z represents a measurement of the state vector x, while $\hat{z}$ represents predicted measurements based on the state vector.

Algorithm 1 Algorithm for Graph Optimization Estimated Pose Update

$\text{OPTIMIZE}(\mathbf{x})$:
  while(!converged)
    $(\mathbf{H}, \mathbf{b}) = \textbf{buildLinearSystem}(\mathbf{x})$
    $\Delta\mathbf{x} = \textbf{solveSparse}(\mathbf{H}\Delta\mathbf{x} = -\mathbf{b})$
    $\mathbf{x} = \mathbf{x} + \Delta\mathbf{x}$
  end
  return x

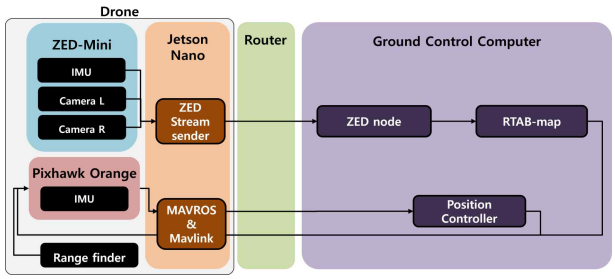

Fig. 4.: Diagram Illustrating the Overall System

The state that minimizes e_{ij} can be determined by solving the problem of minimizing squared error:

$$x^* = argmin_x \sum_i e_{ij}^T(x_i, x_j)\Omega_{ij}e_{ij}(x_i, x_j)$$

Where Ω refers to the information matrix, which is the inverse of the covariance matrix of the measurements.

It has been established that the graph optimization problem described above, when linearized, is equivalent to minimizing the squared error function [11].

$$F(x + \Delta x) = c + 2b^T\Delta x + \Delta x^T H \Delta x$$

Where, Where, $c = \sum_{ij} e_{ij}^T\Omega_{ij}e_{ij}$, $b^T = \sum_{ij} e_{ij}^T\Omega_{ij}J_{ij}$ and $H = \sum_{ij} J_{ij}^T\Omega_{ij}J_{ij}$. In this context, J_{ij} represents the Jacobian matrix of $e_{ij}(x)$ computed with respect to the state vector x.

The quadratic form can be minimized with respect to Δx by solving the linear system $H\Delta x^* = -b$. When we determine the matrices H and b, it's important to note that while b may lose sparsity, H retains its sparsity. This sparsity property of H enables the efficient solution of the linear system. Algorithm 1 outlines the process for updating the estimated pose using the matrices H and b.

4. EXPERIMENT

4.1 Experimental Setup

In real-world experiments, we have successfully developed a system enabling autonomous drone operations within GPS-denied indoor environments. Our drone setup incorporates key components including the Pixhawk Orange, Jetson Nano, and the ZED camera. A visual depiction of the drone's physical configuration is presented

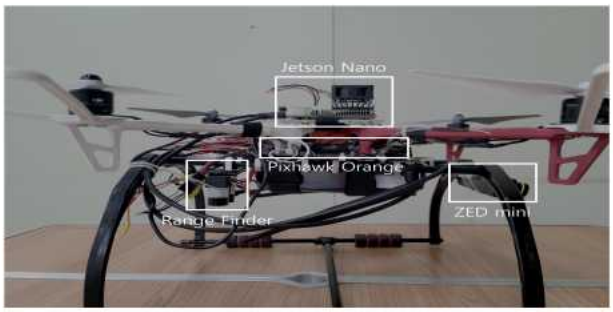

Fig. 5.: Drone Side View Perspective

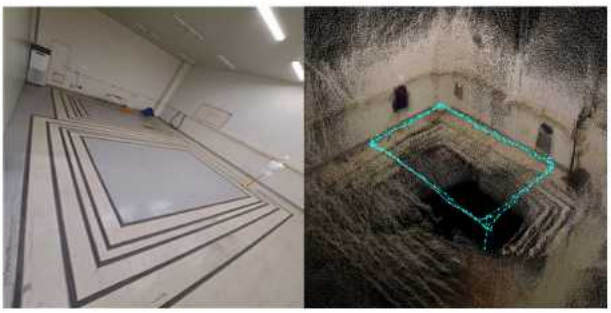

Fig. 6.: View of the Actual Experiment Place and the Map Generated by SLAM

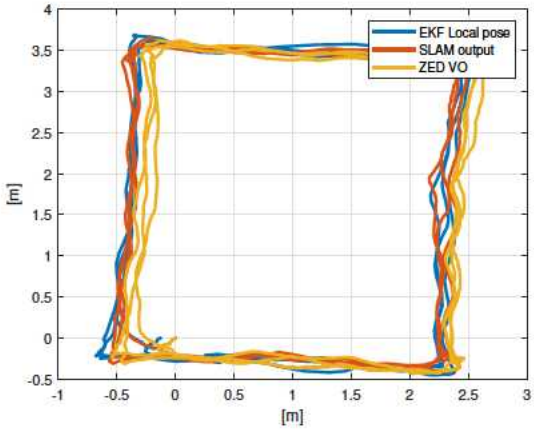

Fig. 7.: TF, Visual Odometry (VO), and EKF Position Comparison

in Fig. 5. To establish seamless communication and configuration, multiple communication channels have been integrated. The holistic communication process is illustrated in Fig. 4.

The ZED camera is connected to the Jetson Nano to stream video to the ground control computer through the network. We used 100 Mbps Wi-Fi for transferring ROS topics and streaming video over machines. The video had a resolution of 2560x720 and 15 fps. When streaming video, hardware acceleration is used in Jetson Nano to perform real-time encoding and decoding with minimal overhead. The streaming module uses the RTP protocol to send and receive the video feed [12].

The Jetson Nano communicates to the Pixhawk using the telem2 port of the Pixhawk which uses UART interface protocol to transfer log data [13]. The ground control computer communicates with the Jetson Nano using ROS [14].

The ZED node passes the VO topic to the RTAB-Map running on the ground control computer [15]. RTAB-Map uses this VO topic to generate the global map and update the pose. The updated pose was then published as ROS TF and is used for the Ardupilot as a visual measurement for pose estimation.

Using the estimated pose, position control is performed to determine guidance commands at the ground control PC. The Position control later uses local pose information of an object to determine guidance command which will be dealt with in the second part of the series. The guidance command is delivered to the MAVROS node and on-board control is performed accordingly.

The drone's flight time is around 15 to 20min with a fully charged battery with this setup.

4.2 Sensor Fusion Results

Exploring SLAM's role as EKF measurements: SLAM processes vision and camera data for location and mapping. ZED camera provides vision, but revisits need 3D mapping and loop closure [16]. RTAB-Map gives compensated position shared via TF, with a 3D global map. TF is VO for EKF. SLAM's TF is map-based, EKF's data is NED. Merging demands crucial coordinate transformation due to navigation's lack of Earth sensors. SLAM and map framing align naturally [17]. Unfortunately, assessing accuracy needs costly equipment. So, achieving precise output remains tough. Contrasting SLAM's loop closure with VO's drift drawback is achieved by comparing their performance along a known track.

Fig. 6 visually presents both the experimental environment and the map generated by the SLAM process. In this experiment, we manually moved the drone along a predetermined path to acquire comprehensive data for map creation. The trajectory of the drone is depicted on the map with aqua-colored dots and lines. Despite the apparent scarcity of distinctive feature points in the experiment location, the SLAM-generated map is effectively visualized using RVIZ, effectively capturing the environment's characteristics.

Fig. 7 graphically displays the TF produced by SLAM, the camera's vision-based position, and the ultimate fused position obtained by EKF, utilizing TF as a measurement input. The visual representation of the vision position exhibits drift, due to the accumulation of errors that tend to manifest in positioning systems without the presence of coordinate-based measurements.

Conversely, the output derived from SLAM, specifically the TF, illustrates a loop-closing effect, effectively correcting the drone's current position within the map by recognizing previously visited locations. This advantageous loop-closing effect is equally evident in the output

Table 2.: Details of the Drone Hardware Configuration

Type	Amount	Name	Size	Weight
Power	1	DJI F450 FRAME KIT + Landing gear	-	518
Propeller	4	Self-Lock Propeller (Universal Type)	$\phi = 249$mm	13
Battery	1	PT-B5200N-UKP55	$44 \times 153 \times 25.5$	385
Motor	4	2312E 960KV(CW/CCW)	$90 \times 70 \times 30$	69.5
Electronic Speed Controller	4	DJI E305 430 Lite	54x24x9	27
Flight Controller	1	Pixhawk Orange	$94.5 \times 44.3 \times 22.3$	35
Stereo Camera	1	ZED	$124.5 \times 30.5 \times 26.5$	63
Processor	1	Jetson Nano	$91.3 \times 118.7 \times 35.8$	348

obtained through EKF.

On a different note, the advantage of synergizing SLAM and EKF becomes conspicuous when steering the drone with rapid movements. VO operates relatively slowly but exhibits minimal drift, whereas the IMU, operating at a higher frequency, contains more substantial drift. When integrated, these two sources of data can swiftly provide the necessary information to control the drone effectively.

4.3 Drone Tracking with Fusion Data

Implementing vision-based techniques on drones poses more challenges compared to ground robots, particularly when it comes to extracting navigation information from VO. The integration of VO with an IMU is a common practice due to the imperative need for frequent navigation data to sustain control loops. Any inaccuracies present in this data can exert a substantial impact on drone behavior, particularly during aerial operations [18].

Additionally, instances may arise where external forces cause the drone to undergo abrupt motions that surpass the frame capture rate of the camera. This can lead to a degradation in the performance of capturing feature points, hampering accurate tracking. Consequently, it becomes crucial to stabilize the drone's motion and ensure suitable speed to enable the proper functioning of the vision-based technique.

To evaluate the effectiveness of the EKF in conjunction with SLAM for drone navigation, an experiment was conducted. The experimental setup was structured as outlined below:

The map frame was established with its origin coinciding with the initial position of the drone. The x-axis of the map frame was aligned with the drone's initial heading direction, while the y-axis was positioned 90° to the right of the x-axis.

Using this map frame origin as reference, the vertex of a square was defined, determining the drone's position, and the corresponding heading angle was designated as a waypoint. These waypoints were sequentially assigned to the drone in a counterclockwise direction. Upon reaching each waypoint, the drone adjusted its heading angle by 90° to align its body frame's x-axis with the subsequent waypoint.

Once the drone approached a waypoint within specified distance and angle thresholds, which were set at 0.2m and 10° respectively, the next waypoint was dispatched.

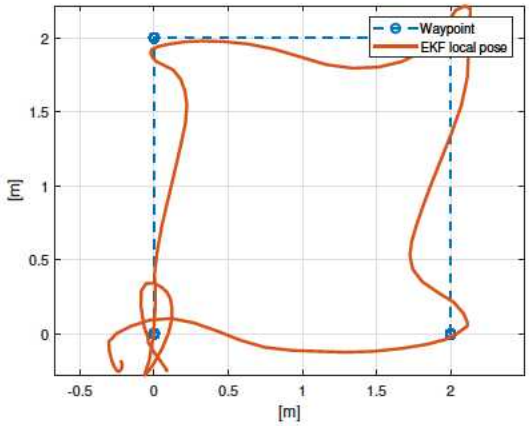

Fig. 8.: Drone Position During Tracking of Square Waypoints

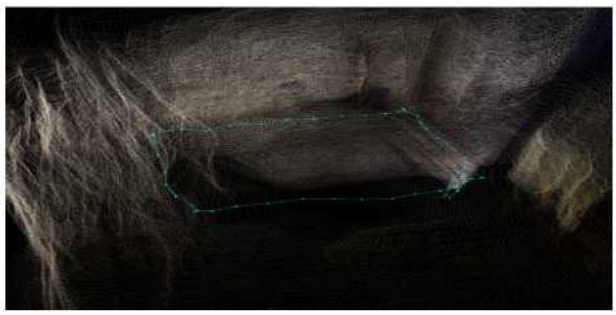

Fig. 9.: 3D Map Generation During Waypoint Tracking

This process was repeated for a total of 9 waypoints.

Fig. 8 visually presents the waypoints as defined by the experimental scenario, alongside the drone's position. The figure effectively demonstrates the drone's stable movement as it successfully tracks the assigned waypoints. The apparent arc in the trajectory, rather than a straight line, results from the process of transitioning to the next waypoint. This transition occurs before the drone's heading angle reaches a steady state, thus retaining the influence of the transient response during heading adjustments while moving towards the next waypoint. The map generated by SLAM while the drone diligently follows the waypoints is illustrated in Fig. 9, which represents that the integration of EKF with SLAM remains effective even when the drone is autonomously navigating and fulfilling missions.

5. CONCLUSION

This paper explores the development of an indoor drone flight control system. With the drone market expanding rapidly, the need for reliable drone operation in GPS-deprived environments grows. The research combines IMU-based EKF to achieve stable drone position and attitude information.

The study underlines the importance of autonomous flight technology in a thriving drone market. Challenges arise indoors due to obstacles and unreliable GPS signals. To address this, the research integrates VPS technology and vision sensors, ensuring stable drone flight even in GPS-denied areas.

The paper discusses EKF intricacies and its relevance in drone navigation. It also highlights SLAM's role in boosting drone navigation in indoor settings.

The experimental setup employs Pixhawk Orange, Jetson Nano, and the ZED-Mini camera, showcasing practical implementation. Results indicate promising real-world potential, particularly in challenging navigation scenarios.

In summary, this research offers insights into drone technology's future. Integration of advanced methods and sensors enhances drone navigation in demanding environments.

6. ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT)(2022R1A2B5B03001459), and was also supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government [22ZK1100, Honam region regional industry-based ICT convergence technology advancement support project].

REFERENCES

[1] T. Tomic, K. Schmid, P. Lutz, A. Domel, M. Kassecker, E. Mair, I. L. Grixa, F. Ruess, M. Suppa, and D. Burschka, "Toward a fully autonomous uav: Research platform for indoor and outdoor urban search and rescue," *IEEE robotics & automation magazine*, vol. 19, no. 3, pp. 46–56, 2012.

[2] H. Bavle, P. De La Puente, J. P. How, and P. Campoy, "Vps-slam: Visual planar semantic slam for aerial robotic systems," *IEEE Access*, vol. 8, pp. 60704–60718, 2020.

[3] L. V. Santana, A. S. Brandao, and M. Sarcinelli-Filho, "Outdoor waypoint navigation with the ar. drone quadrotor," in *2023 international conference on unmanned aircraft systems (ICUAS)*, pp. 303–311, IEEE, 2015.

[4] I.-C. Ryu, J.-I. Ham, J.-O. Park, J.-W. Joeng, S.-C. Kim, and H.-S. Ahn, "Indoor drone pedestrian following system with edge computing and neural networks. part 2: Development of tracking system and monocular depth estimation," Proceedings of the 23rd International Conference on Control, Automation and Systems (ICCAS 2023), Yeosu, Korea, Oct. 17∼20, 2023.

[5] A. D. Team, "Ekf origin." https://ardupilot.org/dev/docs/mavlink-get-set-home-and-origin.html (2023/02/10).

[6] A. Hendrix, "Extended kalman filter (ekf)." https://ardupilot.org/copter/docs/common-apm-navigation-extended-kalman-filter-overview.html (2023/01/20).

[7] R. Paul, "Ekf process and observation models." https://github.com/PX4/PX4-ECL/blob/master/EKF/documentation/Process(2023/01/12).

[8] R. Nick, "Arducopter flight controllers." https://nrotella.github.io/journal/arducopter-flight-controllers.html (2023/01/22).

[9] M. Labbé and F. Michaud, "Rtab-map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation," *Journal of Field Robotics*, vol. 36, no. 2, pp. 416–446, 2019.

[10] J.-B. Song and S.-Y. Hwang, "Past and state-of-the-art slam technologies," *Journal of Institute of Control, Robotics and Systems*, vol. 20, no. 3, pp. 372–379, 2014.

[11] G. Grisetti, R. Kümmerle, C. Stachniss, and W. Burgard, "A tutorial on graph-based slam," *IEEE Intelligent Transportation Systems Magazine*, vol. 2, no. 4, pp. 31–43, 2010.

[12] STEREOLABS, "Zed video streaming." https://www.stereolabs.com/docs/video/streaming/ (2022/10/12).

[13] ShubraChowdhury, "Robotics and nvidia jetson." https://github.com/ShubraChowdhury/Robotics-And-NVIDIA-Jetson (2023/01/12).

[14] A. Hendrix, "Ros multiple machines." http://wiki.ros.org/ROS/Tutorials/MultipleMachines (2023/02/1).

[15] A.-C. documentation, "Zed stereo camera for non-gps navigation." https://ardupilot.org/copter/docs/common-zed.html (2022/11/12).

[16] M. Labbe and F. Michaud, "Online global loop closure detection for large-scale multi-session graph-based slam," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2661–2666, IEEE, 2014.

[17] A. Addison, "Coordinate frames." https://automaticaddison.com/coordinate-frames-and-transforms-for-ros-based-mobile-robots/ (2022/08/12).

[18] J. García, J. M. Molina, and J. Trincado, "Real evaluation for designing sensor fusion in uav platforms," *Information Fusion*, vol. 63, pp. 136–152, 2020.