Enhancing Indoor Mobile Robot Localization through the Integration of Multi-Sensor Fusion Algorithms

Rut Yatigul
Department of Production and Robotics Engineering
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
s6601021810081@email.kmutnb.ac.th

Noppadol Pudchuen
Department of Instrumentation and Electronics Engineering
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
noppadol.p@eng.kmutnb.ac.th

Aran Blattler
Department of Production and Robotics Engineering
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
aran.b@eng.kmutnb.ac.th

Wisanu Jitviriya
Department of Production and Robotics Engineering
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
wisanu.j@eng.kmutnb.ac.th

2024 1st International Conference on Robotics, Engineering, Science, and Technology (RESTCON) | 979-8-3503-4884-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/RESTCON60981.2024.10463559

***Abstract*— This paper presents an innovative approach that combines visual odometry, Inertial Measurement Unit (IMU), and wheel odometry, using the Extended Kalman Filter and Unscented Kalman Filter to enhance mobile robot localization and mapping. The primary goal is to improve localization accuracy and robustness while providing a cost-effective alternative to traditional SLAM methods that rely on expensive LIDAR and RGBD camera systems. By integrating visual odometry, IMU data, and wheel odometry our method not only enhances precision and robustness of the mobile robot localization system but also reduces the financial burden associated with high-end sensor hardware. This fusion creates a reliable solution, particularly suited for resource-constrained environments. Our research contributes to the democratization of SLAM technologies, making them more accessible to a wider range of applications. The results presented in this report showcase the potential of our approach. This paper highlights the pivotal role of multi-sensor fusion framework which significantly enhances localization accuracy and robustness.**

Keywords*— *Visual odometry, Inertial Measurement Unit (IMU), Extended Kalman Filter, Unscented Kalman Filter, Multi-sensor fusion, Simultaneous Localization and Mapping

I. INTRODUCTION

In the current era, Visual Simultaneous Localization and Mapping (SLAM) have emerged as prominent trends in mobile robotics. Autonomous mobile robots necessitate a robust system for precisely identifying their position and orientation to navigate their environment effectively. Various methods have been developed to enable Visual SLAM, such as RTAB-MAP [1], which utilizes depth cameras for precise position and orientation estimation, and ORB SLAM, applicable to monocular, stereo, or depth cameras [2]. One of the primary advantages of Visual SLAM lies in its capacity to collect comprehensive environmental information. By leveraging cameras, Visual SLAM captures not only geometric data but also detailed texture and color information. This results in the creation of maps that are not only rich in spatial information but also semantically meaningful. Such maps find applications in augmented reality, object recognition, and navigation in visually diverse environments. Furthermore, Visual SLAM offers cost-efficient localization solutions [3]. Cameras are generally more affordable than high-quality LiDAR sensors, making Visual SLAM an economically attractive choice for various applications, especially in consumer electronics, drones, and mobile robotics. However, Visual SLAM faces challenges. It can be sensitive to lighting conditions, relying on camera images. Variations in lighting, such as low-light conditions, shadows, or low contrasts, can significantly affect its performance. Additionally, Visual SLAM may encounter difficulties in texture-less environments [4], where there are no distinctive visual features for tracking and mapping. This is particularly problematic in featureless spaces with repetitive patterns, like blank walls or open skies. Other

Fig. 1. ORB SLAM2 output key frame.

limitations of Visual SLAM include a loss in localization accuracy during rapid motion [5] and interruptions in sensor signals when the camera is obstructed by walls, corners, or accidentally blocked by humans. These challenges necessitate the development of robust methods to enhance the accuracy and reliability of mobile robot localization using visual odometry.

This paper introduces a localization approach using Extended Kalman Filter and Unscented Kalman Filter to encounter challenge present in Visual SLAM. Implementation and tested on Turtlebot3 Burger with ROS, our method enhances robustness and accuracy by fusing data from RGB-D camera, IMU, and wheel odometry.

II. METHODOLOGY

Exploring indoor localization systems highlights visual odometry as a popular algorithm. However, relying on a single sensor for autonomously operating robot localization entails considerable risk and unreliability, given the susceptibility of sensors to noise. Visual sensors (stereo cameras) have their own set of limitations. For instance, they may struggle to operate optimally in low-light or low-textured environments, as SLAM algorithms depend on feature points and feature matching between consecutive frames to precisely determine the position and orientation of the mobile robot. Another drawback of visual odometry is the degradation of localization accuracy during rapid movements due to motion blur in the images. This paper addresses these challenges by introducing a localization system that strategically fuses inputs from multiple sensors. In the initial step, data is collected from various sensors to determine an approximate position and orientation. Subsequently, in the second step, the fused position and orientation are precisely estimated using both the Extended Kalman Filter and Unscented Kalman Filter.

A. *Mobile Robot Localziation System*

In recent years, SLAM has gained widespread adoption due to its compatibility with low-cost hardware, making it applicable to monocular, stereo, and RGB-D cameras while maintaining high-precision measurements. A well-known camera for indoor localization is the RealSense D435 RGB-D camera, commonly used in indoor environments. However, for this research, the researcher will exclusively employ a stereo camera pair of RealSense D435 RGB-D camera with the intention of implementing a more cost-effective stereo setup using web cameras in the future. Among the SLAM algorithms available, ORB-SLAM2 stands out as a feature-based SLAM system, which can be employed with mono, stereo, and RGB-D cameras, providing high-precision localization and mapping capabilities.

ORB-SLAM2 processing consists of two key stages: the front end and the back end, working in tandem to accurately estimate the camera's position and construct a comprehensive map of its surroundings.

The front end is responsible for extracting distinctive visual features from the camera image. These features are typically represented as "keypoints" and encompass ORB (Oriented FAST and Rotated BRIEF) features. Feature matching utilizes the keypoints obtained during the feature extraction process to establish correspondence between consecutive frames. This critical step allows for the estimation of camera motion by tracking how these features move between frames. The camera pose estimation relies on the matching features and their relative motion. ORB-SLAM2 determines the camera's position and orientation at each frame concerning the previous frame, employing the Perspective-n-Point (PnP) algorithm. This algorithm links 2D image points to 3D world points, effectively establishing the camera's precise location and orientation. Loop detection plays a crucial role in identifying instances when the camera revisits previously observed areas. This feature significantly contributes to correcting drift and enhancing the overall consistency of the map.

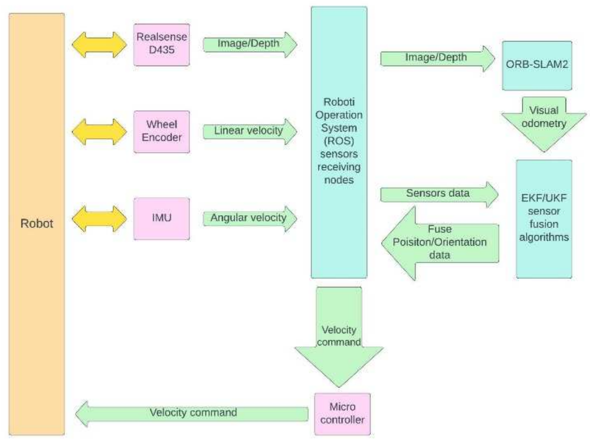


Fig. 2. System architecture diagram of proposed technique.

The back end includes a map building which use information from front end and build 3D map of environments. Bundle adjustments is used to refine the accuracy of map and camera pose estimation. Optimization is used to optimize the entire system correcting accumulated errors and refine map and camera trajectory. Finally, Loop closure correction when loop is detected at the front end, the back-end work to adjust map and camera pose to observed loop align correctly with previous visited area, eliminating drift.

As for the system provided in Fig. 2. The ORB-SLAM mainly provides absolute position and orientation, IMU provides angular velocity and acceleration, Wheel odometry provides linear velocity.

B. *Sensor Fusion Algorithms*

The Extended Kalman Filter (EKF) and Unscented Kalman Filter referenced from [6] is currently employed for sensor integration on mobile robots. For instance, wheel odometry and an Inertial Measurement Unit (IMU) each exhibit substantial accumulative error. Visual odometry, reliant on lighting conditions and environmental texture, is also susceptible [7]. Depending solely on a single sensor for robot pose calculation poses a significant risk of localization errors or loss of position data [10]. An EKF typically consists of two parts [8]. In the case of nonlinear systems, the state equation can be expressed as Eq. (1).

$$x_k = f(x_{k-1}, u_k) + \omega_{k-1} \quad (1)$$

where x_k represents the system state of the mobile robot at time k, f is a nonlinear state transition function, u_t serves as the speed control input function, and ω_{k-1} accounts for a process noise [11]. The primary scope of this research is to accurately estimate the position of the mobile robot within a 2-D 3m x 3m environment. Furthermore, the measurement model can be conceptualized as Eq. (2).

$$z_k = h(x_k) + v_k \tag{2}$$

where, z_k represents the measurement at time k, h is a nonlinear sensor model that map the state into measurement space, v_k denotes the measurement noise. The Extend Kalman filter approximates the non-linear matrix h by employing a Taylor series expansion centered around the estimated state vector, thereby accommodating the nonlinearity inherent in the sensor model as presented in Eq. (3).

$$h[\hat{x}_{k|k-1}] \approx h[\hat{x}_{k|k}] + \frac{\partial h[\hat{x}_{k|k}]}{\partial \hat{x}_{k|k}} (\hat{x}_{k|k-1} - \hat{x}_{k|k}) \tag{3}$$

Linear transformations frequently introduce significant errors in the estimated state vector, potentially leading to filter divergence.

The Unscented Kalman Filter is derived from unscented transformation, a method for estimating the statistics of a random variable undergoing a non-linear transformation [9]. A set of $2 * n_X + 1$ weight samples is deterministically selected to accurately capture the true mean and variance of the prior random variable as shown in Eq. (4).

$$n_X = n_x + n_w + n_v \tag{4}$$

Where, n_X is the number of samples, n_x is the number of process state, n_w is the dimension of ω_k and n_v is the dimension of v_k. The Unscented Kalman Filter approximates the nonlinear observation matrix by employing unscented transformation as described in Eq. (5).

$$h[\hat{x}_{k|k-1}] \approx \sum_{i=0}^{2*n_x} W_i * h[X^x_{i,k|k-1}] + X^v_{i,k} \tag{5}$$

W_i are the weights, $X^x_{i,k|k-1}$ are sigma points, $X^v_{i,k}$ are sigma point use to describe measurement noise.

III. Experimental Result

A. Environmental and equipment

Fig. 3 shows the Turtlebot3 Model Burger which was used in the experiment. Turtlebot3 is a popular and compact differential drive mobile robot. The forward movement direction of the car is in positive direction of x-axis, the leftward movement direction is the positive y-axis, and the vertically upward direction is the positive direction in z-axis, which satisfies the right-hand rule. The placement of TurtleBot components is illustrated in Fig. 3.

The testing environment is a 3m x 3m square room, illustrated in Fig. 4. The robot's path consists of two distinct routes. Firstly, there is the square path, as depicted in Fig. 5. The mobile robot maintains a constant speed of 0.03 meters per second, initially moving in the positive x-direction for 2.4 meters. It then executes a -90 degree yaw, transitioning to continuous motion in the negative y-direction for 2.2 meters.

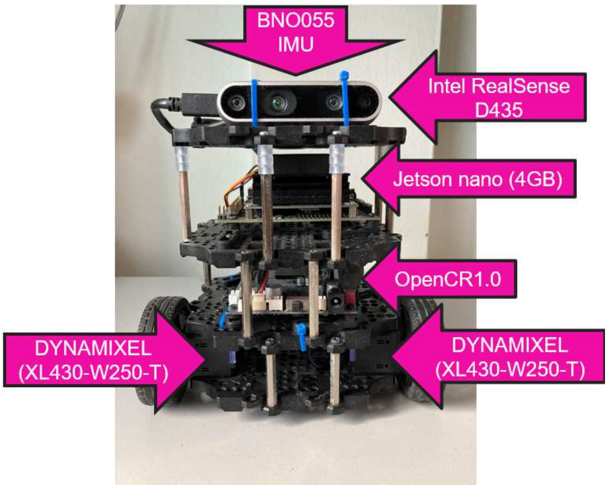


Fig. 3. Turtlebot3 Model Burger.

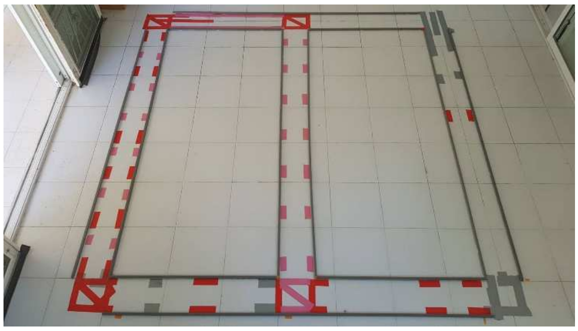
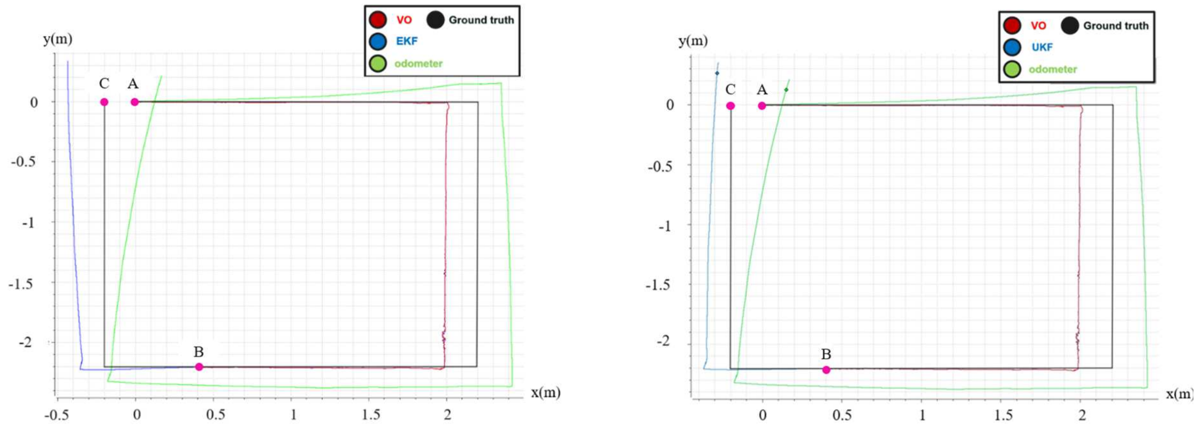
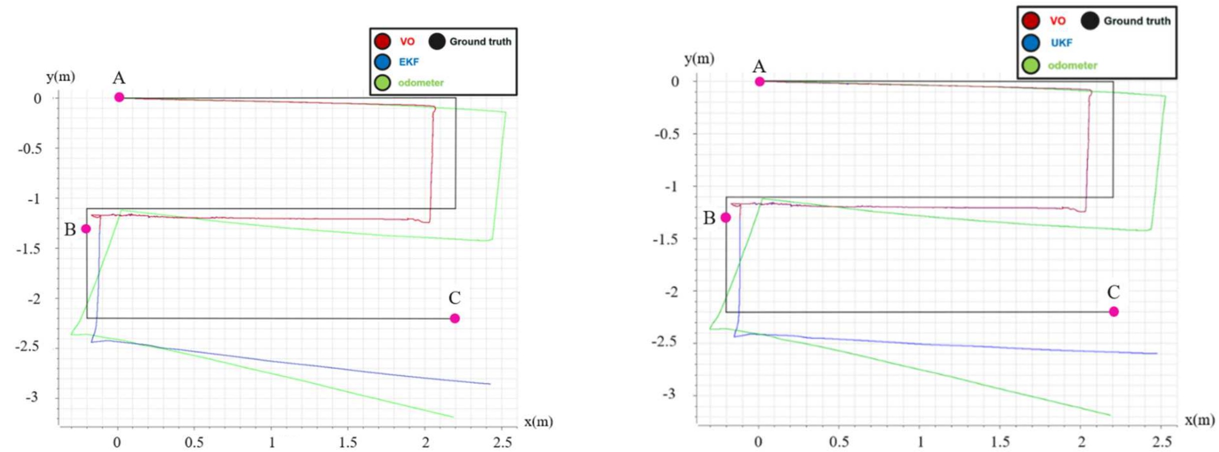


Fig. 4. Testing circuit



(a). EKF (b). UKF

Fig. 5. Result of the multi method localization route 1



(a). EKF (b). UKF

Fig. 6. Result of the multi method localization route 2

This sequence is repeated, forming a square loop motion. Secondly, the robot follows the z-path, as shown in Fig. 6. Here, the robot moves in the positive x-direction for 2.4 meters, performs a -90 degree yaw, moves in the negative y-direction for 1.1 meters, executes another -90 degree right yaw, and then performs an inverted motion of equal length in both the x and y axes, completing a Z-shaped trajectory. Noting a -0.2 meter initial deviation in both the square and Z motion patterns due to setup limitations of the Turtlebot3 and the testing environment, the research employs a Turtlebot3 model Burger, which underwent a processor upgrade from Raspberry Pi 3 to the Jetson Nano developer kit for enhanced image processing performance. The integration of an Intel RealSense D435 RGB-D camera into the robot facilitates visual odometry processing, while the GY-BNO055 IMU provides absolute yaw data and angular velocity information. Furthermore, the TurtleBot utilizes a DYNAMIXEL (XL430-W250-T) for dead reckoning.

B. Fusion localization result

We conducted experiments by recording the synchronized position from three odometry sources: visual odometry, wheel odometry, and fusion odometry. It is evident that the fusion of Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF) yields results that closely align with the actual path. The positional accuracy of wheel odometry, while generally reliable, tends to accumulate errors at each turning point. On the other hand, visual odometry, despite its high accuracy, is susceptible to instability due to various factors such as environmental shadows, texture variations, and rapid camera motion, visual blockage which will be the main thing that tested on this experiment leading to potential position tracking loss and inaccuracies. However, the implementation of fusion algorithms significantly improves the localization accuracy and robustness. The integration of EKF and UKF helps mitigate the errors associated with wheel and visual odometry, resulting in a more precise and reliable representation of the robot's path. Upon closer examination of the accuracy differences between EKF and UKF, it becomes apparent that UKF exhibits a slightly superior accuracy when compared to both methods against the actual robot path. This can be explained by a few reasons for example linearization approach, Extended Kalman Filter use first-order Taylor series expansion around the current mean and covariance estimate. This linearization process may introduce errors, especially if the nonlinearity is high or state space is not well behaved while Unscented Kalman Filter uses a set of carefully chosen sample point (sigma points) that can capture mean and covariance information more accurately. By propagating these sigma points through the nonlinear functions, UKF avoids the need for explicit linearization which in turn provides a more accurate approximation. It should be noted that with the better localization performance of UKF it also comes with trade off which is more requirement of computational power compared to EKF. The experimental results are illustrated in Fig. 5 and Fig. 6. Positioning results reveal a common issue with the wheel odometer (odometer) (depicted in green), namely directional drifting. This problem is notably observed after the robot executes a turn. Visual odometry (VO) (depicted in red) appears to be more consistent with the ground truth compared to the wheel odometer. However, it exhibits a lack of robustness in challenging visual conditions, such as areas with shadows, low-texture environments, visual obstruction, and rapid camera movements. The fusion positioning using Unscented Kalman Filter (UKF) / Extended Kalman Filter (EKF) (depicted in blue) successfully mitigates the directional offset observed in the wheel odometer and compensates for missing positioning data in areas where visual odometry may falter. To further assess performance, a series of equidistant sampling points are selected along the path. The Root Mean Square Error (RMSE) is chosen as a numerical method to quantify the error of each sensor and fusion method. The errors for each sensor and fusion method are presented in detail in Tables 1 and 2.

Table 1: Fusion localization error of route 1 (square) (unit: m)

Sensor Type	Root Mean Square Error	
	A to B	B to C
Visual	0.204339	Tracking loss
Odometer	0.240426	0.220264
EKF	0.204054	0.193348
UKF	**0.203757**	**0.148670**

Table 2 Fusion localization error of route 2 (Z) (unit: m)

Sensor Type	Root Mean Square Error	
	A to B	B to C
Visual	0.202272	Tracking loss
Odometer	0.289652	0.495693
EKF	0.202040	0.375073
UKF	**0.200991**	**0.274328**

IV. Conclusion and Future Work

This paper examines the effectiveness of a real-time indoor mobile robot localization system through a sensor fusion approach, incorporating both Extended Kalman Filter and Unscented Kalman Filter techniques. Despite the distinct functionalities of individual sensors, their limitations are effectively addressed through the fusion process, resulting in a robust and accurate localization system. Experimental data is sourced from ORB-SLAM2 (providing absolute pose information), GY-BNO055 IMU (supplying absolute orientation data and angular velocity), and DYNAMIXEL (XL430-W250-T) (contributing linear velocity data).

Looking ahead, future work will explore alternative sensor fusion methods and implement machine learning techniques to identify and rectify faulty sensor signals, ultimately improving the overall reliability of the system. Additionally, there is a consideration for integrating external positioning systems to enhance localization accuracy.

Acknowledgements

We extend our sincere gratitude to the researchers of the Center for Innovative Robot and Advanced Precision Systems (iRAPs) and the Invigorating Robot Activity Project (iRAP) for their invaluable technical support. Their expertise and assistance have played a pivotal role in the success of this project.

Reference

[1] Das, S. (2018). Simultaneous localization and mapping (SLAM) using RTAB-MAP. *arXiv preprint arXiv:1809.02989*.

[2] Mur-Artal, R., Montiel, J. M. M., & Tardos, J. D. (2015). ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE transactions on robotics*, *31*(5), 1147-1163.

[3] Zhao, Y., Smith, J. S., & Vela, P. A. (2020). Good graph to optimize: Cost-effective, budget-aware bundle adjustment in visual slam. *arXiv preprint arXiv:2008.10123*.

[4] Gomez-Ojeda, R. (2020). Robust visual slam in challenging environments with low-texture and dynamic illumination.

[5] Piao, J. C., & Kim, S. D. (2017). Adaptive monocular visual–inertial SLAM for real-time augmented reality applications in mobile devices. *Sensors*, *17*(11), 2567.

[6] Moore, T., & Stouch, D. (2016). A generalized extended kalman filter implementation for the robot operating system. In *Intelligent Autonomous Systems 13: Proceedings of the 13th International Conference IAS-13* (pp. 335-348). Springer International Publishing.

[7] Chen, G., & Hong, L. (2023). Research on Environment Perception System of Quadruped Robots Based on LiDAR and Vision. *Drones*, *7*(5), 329.

[8] Ribeiro, M. I. (2004). Kalman and extended kalman filters: Concept, derivation and properties. *Institute for Systems and Robotics*, *43*(46), 3736-3741.

[9] Wan, E. A., & Van Der Merwe, R. (2000, October). The unscented Kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)* (pp. 153-158). Ieee.

[10] Chen, G., & Hong, L. (2023). Research on Environment Perception System of Quadruped Robots Based on LiDAR and Vision. *Drones*, *7*(5), 329.

[11] Chen, K., Yang, J., Jiang, S., & Xiong, C. (2023, April). Multi-Sensor Fusion Tomato Picking Robot Localization and Mapping Research. In *Journal of Physics: Conference Series* (Vol. 2477, No. 1, p. 012057). IOP Publishing.