Robust Visual SLAM with Integrated UWB Positioning for Construction Robotics

Jingyang Liu[1], Linxiaoyi Wan[2], and Joshua Bard[3]

[1,2,3]Carnegie Mellon University, Pittsburgh, 15206, PA, United States
Email: [1]jingyanl@andrew.cmu.edu, [2]linxiaow@andrew.cmu.edu, [3]jdbard@cmu.edu

***Abstract*—In this paper, we present a multimodal localization system for construction robotics operated in job sites where visual simultaneous localization and mapping (VSLAM) algorithms can yield inaccurate and inconsistent results due to environmental factors such as insufficient textures, occlusion, and unstable lighting conditions. The system integrates Ultra-Wideband (UWB) ranging and inertial measurement unit (IMU) with VSLAM as complementary sensing modalities. The integration of UWB and IMU can mitigate the cumulative drift of VSLAM and provide reliable localization when VSLAM fails. The fusion of multiple sensing modalities is based on the extended Kalman filtering (EKF) framework. We tested the localization system in real indoor environments with an industrial robot and validated it with ground truth data. The result shows that the proposed multimodal localization system achieves robust performance with increased accuracy under three test conditions — inconsistent lighting, regions with minimal features, and regions with repetitive patterns.**

I. Introduction

Indoor construction robotics can assist humans in a wide range of processes at different stages, including, installation, finishing, inspection, maintenance, and demolition. Among all these scenarios, location estimation is essential for robots to effectively navigate and perform tasks autonomously. Visual simultaneous localization and mapping (VSLAM) is one of the most adopted techniques for indoor localization. By extracting the visual features of a sequence of images as descriptors, the position of the camera can be tracked in real time. However, several environmental factors of indoor construction sites can pose challenges to VSLAM (Figure 1).

- **Inconsistent illumination**

Construction sites can rely on temporary lighting or natural light. For areas with low light (shadows, e.g.) and inconsistent illumination (turning on/off artificial lighting, e.g.), VSLAM may fail in feature tracking leading to drift and errors in localization.

- **Regions with minimal features**

VSLAM, such as ORB-SLAM [1], can yield impressive results in well-textured environments. However, construction job sites are not always texture rich. For example, concrete buildings with flat walls can pose challenges to feature detection when the wall surface shares similar color and reflectivity.

- **Regions with repetitive patterns**

Repetitive building components, such as scaffolding, can be self-similar with insufficient visual feature variations. VSLAM is prone to fail due to the lack of distinctive feature points, visual ambiguity, and false matches.

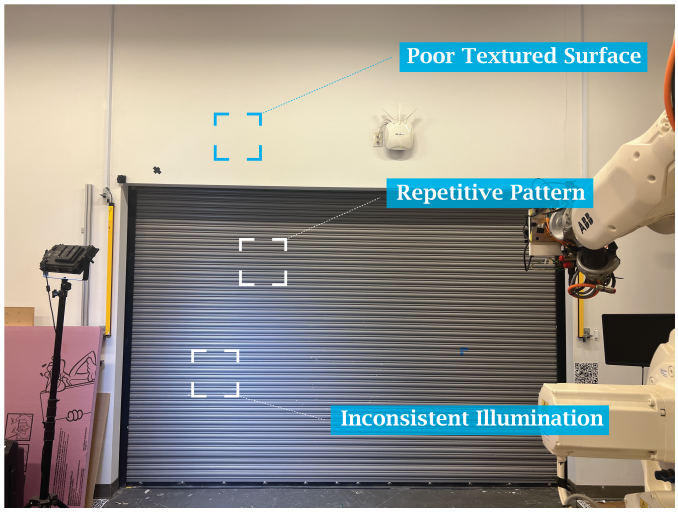

Fig. 1. Challenging scenarios on job sites for visual SLAM including (a) a vertical white wall surface without rich textures (b) a roller door with repetitive line patterns (c) areas with artificial lighting that may change over time.

In addition, indoor construction sites are not suitable for global navigation satellite systems (GNSS), as GNSS signals can be blocked or weakened by walls and other obstacles. To improve the robustness of VSLAM in such environments, this study proposes a multimodal localization system that integrates ultra-wideband (UWB) wireless localization and inertial measurement unit (IMU) into VSLAM systems to mitigate accumulative drift and enhance robustness. By leveraging the synergies between UWB, IMU, and VSLAM, our multimodal localization can overcome the limitations of each sensing modality, providing robust localization in cluttered construction environments. When the VSLAM fails to provide reliable estimation, the complimentary sensing modalities can function to maintain localization consistency. Additionally, UWB, as a global feature-independent localization system, can correct the cumulative drift caused by VSLAM, leading to enhanced accuracy in long-term localization tasks.

In this paper, we build our VSLAM system upon the ORB-SLAM framework and combine the UWB and IMU localization systems with the VSLAM system through the extended Kalman filter (EKF). The proposed multimodal localization system is tested under three scenarios — inconsistent illumi-

nation, repetitive pattern, and insufficient features. Through validating with the ground truth, we found that the integration of UWB and IMU can enhance robustness when VSLAM fails at feature extraction or correspondence matching. Meanwhile, the cumulative drift caused by VSLAM can be reduced by UWB positioning. Real-world experiments demonstrated the feasibility and reliability of the system with applications in the domain of indoor localization for construction robotics.

II. Related Work

Ultra-wideband (UWB) as a wireless communication technique has been adopted for precise indoor localization in recent decades. Compared to other wireless communication technologies, such as radio frequency identification (RFID), Wi-Fi, and Bluetooth, UWB has advantages including (1) accuracy: UWB can achieve high ranging accuracy (sub-centimeter) even in harsh environments due to its resistance to multipath [2]. (2) robustness: UWB is more resistant to signal interference, since the signal can be transmitted simultaneously over multiple frequency bands [3]. (3) power consumption: UWB shows lower normalized energy consumption compared to other wireless communication protocols such as Wi-Fi and ZigBee [4]. For long-term localization, UWB can be a suitable option in terms of energy efficiency. With the aforementioned aspects taken into account, many studies have been conducted to integrate UWB into VSLAM to mitigate scale ambiguity, improve accuracy, and improve robustness [5] [6]. For example, UWB can effectively overcome scale ambiguity and scale drift when combined with monocular VSLAM systems [7] [8]. In GPS-denied environments, UWB can serve as indoor GPS to supplement visual SLAM by providing global constraint [9] or local correction [10]. For long-term localization, UWB can mitigate cumulative drift caused by visual SLAM [11]. For large-scale applications, UWB can be combined with LiDAR to improve localization accuracy [12] [13]. By tightly coupling multiple sensors such as IMU, monocular camera, and UWB, the optimization can be done in a joint manner for a more robust and accurate localization [14] [15]. Extended from previous works, this study proposes a novel multi-modal localization system for construction robotics operated in unstructured environments, the integration of UWB can potentially address the VSLAM failures caused by inconsistent illumination, insufficient features, and repetitive patterns which are commonly seen in job sites.

III. Method

We use three measurements — IMU, the peer-to-peer ranges between UWB nodes, and the range between the robot and the 3D position of the visual features as the source for indoor localization. IMU and UWB are first combined, and then we use EKF for the data fusion of VSLAM and the rectified UWB (Figure 2). We use a set of Marvelmind Super-MP-3D beacons for UWB localization and a Kinect sensor for VSLAM. The update rates for VSLAM and UWB are set to 30 Hz and 15 Hz respectively.

UWB localization uses ultra-wideband impulse radio to measure the distance between mobile and stationary base

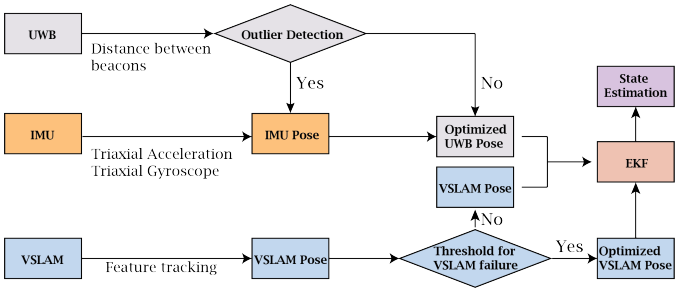

Fig. 2. System flowchart of the multimodal localization system

beacons. Localization of the target using UWB beacons requires (1) three known stationary beacons with known height or (2) four known stationary beacons [16]. In this study, we set up four stationary base beacons in a room of 5.0 m (W) × 15.0 m (L) × 2.8 m (H).The four base beacons are synchronized, and the position of each beacon is self-calibrated. The distance between the mobile beacon and the base beacon can be estimated by the Time-of-Arrival (ToA) of UWB signals. In each time slot, the mobile beacon transmits a short-duration and low-power pulse signal at a unique channel with a time stamp, and the stationary beacon responds with a time stamp after a predefined time. Since the UWB sensors are synchronized at a picosecond level, we can precisely calculate the distance based on the difference between two-time stamps and the speed of the signal. Then we can obtain the location of the mobile beacon using multilateration (Figure 3).

$$(x', y', z') = \sum_{i=1}^{4} \left[\sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} - r_i \right]^2 \quad (1)$$

$$r_i = \frac{(T_{i,cycle} - T_{i,reply}) \times C}{2} \quad (2)$$

Where i is the index of the base beacons, (x_i, y_i, z_i) is the coordinate of i th stationary beacon which is known after self-calibration.(x', y', z') and (x, y, z) are the unknown coordinates of the mobile beacon before and after the error minimization of multilateration. r_i is the distance between each base beacon and the mobile beacon. $T_{i,cycle}$ is the time between the mobile beacon sending the signal and receiving the response. $T_{i,reply}$ is the time between a base beacon receiving the signal and sending the response. C is the speed of light$(3 \times 10^8\ m/s)$.

The multipath interference caused by the reflections of UWB signals off walls and non-line-of-sight (NLOS) can affect signal accuracy, leading to inaccurate localization. IMU can provide more accurate short-term localization invariant to external factors. If the relative position between the UWB and IMU is larger than a threshold, e.g. 0.25 m, we use the predicted position of IMU to correct the UWB localization and remove the outlier. As the update frequency rate of IMU is higher than UWB, the integration of IMU data is processed after two frames of UWB data are processed.

VSLAM tracks features such as corners and edges in the environment and estimates the movement of the camera between consecutive frames. We use ORB-SLAM2 [17] for tracking the RGB-D camera's position and building the map of the environment. The RGB-D camera and UWB systems

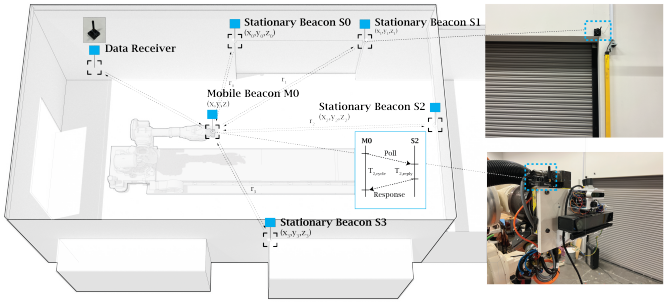

Fig. 3. Infrastructure layout of the UWB positioning system

are synchronized to a common time reference, and thus the localization information of the two systems can be aligned and paired for data fusion. As UWB localization is a general non-linear system, we use the extended Kalman filter (EKF) framework to couple UWB and VSLAM localization for more robust and accurate estimation.

The EKF is conducted in two steps — time updates and measurement updates. Once initialized, EKF predicts the system state at the next step and the uncertainty of the prediction. When the measurement is received, the system updates the prediction and the uncertainty of the current state.

The state vector x consists of position and velocity (v_x, v_y, v_z)can be defined as:

$x = [p_x, p_y, p_z, v_x, v_y, v_z]^T \in \mathbb{R}^6$

The prediction of the next state can be defined as

$$\overline{x}_{k+1,k} = f(\overline{x}_{k,k}, u_k) \tag{3}$$

$$P_{k+1,k} = F_{k,k} P_{k,k} F_{k,k}^T + Q_{k,k} \tag{4}$$

where F is the state transition matrix. $P_{k,k}$ and $Q_{k,k}$ are the covariances of $x_{k,k}$,$w_{k,k}$ respectively. u_k is the control input.$\overline{x}_{k+1,k}$ is a predicted system state vector at time step $k+1$, $\overline{x}_{k,k}$ is an estimated system state vector at time step k .

In the update step, we can obtain the measurement z_k to calculate the Kalman gain G_k. We then update the state covariance matrix by

$$G_k = P_{k,k-1} H_k^T (H_k P_{k,k-1} H_k^T + R_k)^{-1} \tag{5}$$

where R_k is the covariance matrix of measurement uncertainty.H_k is the Jacobian matrix of measurement, which can be defined as partial derivatives of the measurement function with respect to each state variable.

$$H_k = \partial h(i) / \partial x(j) \tag{6}$$

where i is the index of the measurement component. j is the index of the state variable. H_k needs to be computed at each time step.

We can then update the estimate with measurements by

$$\overline{x}_{k,k} = \overline{x}_{k,k-1} + G_k(z_k - H_k \overline{x}_{k,k-1}) \tag{7}$$

And update the uncertainty by

$$P_{k,k} = (I - G_k H_k) P_{k,k-1} \tag{8}$$

In this work, the implementation can be described as follows

- Initialization

We initialize the state vector by using an ArUco marker as a global reference frame for UWB and VSLAM (as robots can start from a known location). We set the covariance matrix based on prior empirical knowledge.

- Prediction

The measurement z_k at timestamp k contains the position and heading offset obtained from the UWB / IMU and VSLAM, and the predicted measurement $\overline{z}_{k+1,k}$ is defined as

$$\overline{z}_{k+1,k} = h(\overline{x}_{k+1,k}) + v \tag{9}$$

where $H \in \mathbb{R}^{8\times6}$ is the measurement model. $\overline{x}_{k+1,k} \in \mathbb{R}^{6\times1}$ is the predicted next position at time stamp k . v represents the uncertainty in the measurement. We can then obtain the innovation sequence $i_k$defined as

$$i_k = z_{k+1,k} - z_k \tag{10}$$

The innovation sequence can be used to calculate the Kalman gain, which determines the weight given to the predicted state estimate and the actual measurement to update the state estimate.

- Estimation

The estimation can be done through Equation (7) The Kalman gain and state covariance are updated accordingly.

- Update

After estimation, we can update the state vector based on the measurements using the Kalman gain. The state vector contains information about the location and velocity of the target.

VSLAM localization can fail due to environmental factors such as inconsistent illumination and insufficient features. We incorporate a data pre-processing step into the pipeline before the data fusion to identify VSLAM failures. We defined an empirical threshold e.g. 0.2 m. If the relative distance between the VSLAM and UWB localization is larger than the threshold, the VSLAM localization is considered as an outlier and replaced by a new position that shares the same heading offset with the UWB localization at an offset distance of the threshold value.

IV. RESULTS

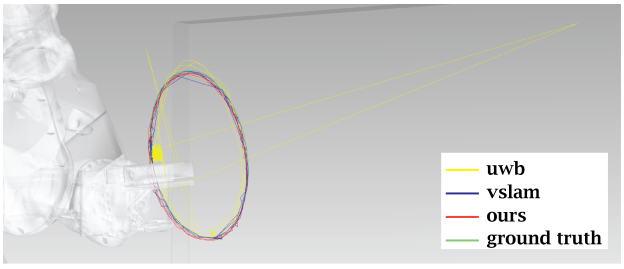

Fig. 4. Trajectories obtained using UWB-only, VSLAM-only, and the proposed multimodal localization system.

Table I gives the average absolute distances between estimated and ground truth trajectories over time stamps. From the

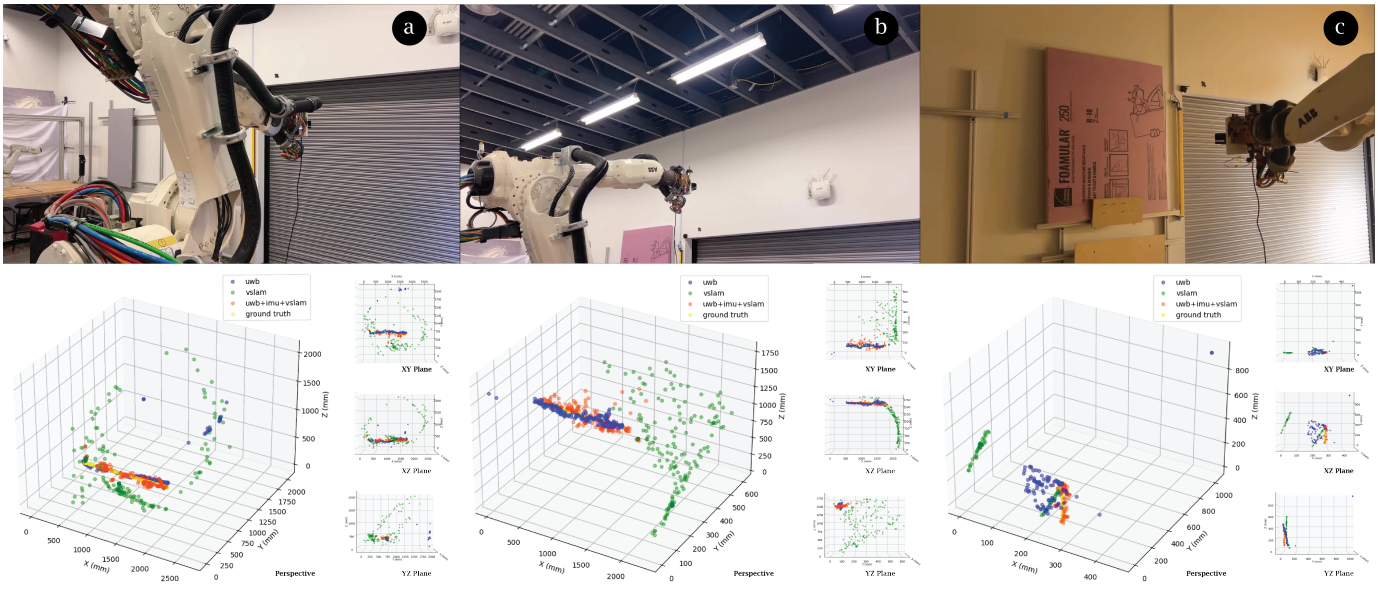

Fig. 5. Performance comparison of VSLAM-only, UWB-only and the proposed multimodal localization system (UWB+IMU+VSLAM) under three test conditions — (a) repetitive pattern, (b) poor textured surface and (c) inconsistent illumination.

TABLE I
THE AVERAGE TRANSLATION ERROR ALONG X, Y AND Z AXIS

	VSLAM-Only	UWB-Only	Ours
x (m)	0.18	0.53	0.13
y (m)	0.12	0.13	0.09
z (m)	0.27	0.33	0.23

table, we can see that the integration of VSLAM can provide a reliable short-term localization, which can significantly reduce the impact of the UWB range measurement outliers caused by environmental factors such as noise, interference, and occlusion (Figure 4). When the environment contains rich, unique visual features and the trajectory is short (the trajectory length in this experiment is 1864 mm) and closed, VSLAM localization can provide an accurate estimation of relative translation and orientation.

We used three environmental interference, including inconsistent illumination, repetitive pattern, and insufficient features, to test the robustness of the multimodal localization system.

- **Regions with repetitive patterns**

Components, such as roller doors, contain repetitive visual features which can pose challenges to the feature tracking in VSLAM localization. We use a roller door as a target object to test if the multimodal localization system can remain robust when symmetric or repetitive scene patterns bring ambiguities for feature correspondence.

- **Regions with minimal features**

Some areas of perpendicular walls in the test environment are blank or poorly textured. Insufficient visual features inliners can reduce the accuracy of VSLAM. We designed a trajectory passing through a blank wall to evaluate the multimodal localization system when distinct visual features are insufficient.

- **Inconsistent illumination**

To test the influence of illumination on the system's robustness, we set up two controllable illumination conditions — the dark environment and the normal environment. We define artificial indoor light of 300 lux as a normal environment and 30 lux as a dark environment. We can test the impact of inconsistent illumination on the multimodal localization system by switching between two illumination conditions.

In each visually challenging environment, we tested three localization approaches including VSLAM-only, UWB-only, and the multimodal system (UWB, IMU and VSLAM). As shown in Figure 5, when the lighting condition changes from normal to dark, VSLAM localization fails due to poor feature extraction and matching. When the lighting condition returns to normal, the accumulated errors lead to significant drift. Meanwhile, we can see that the UWB-only localization is affected by Non-Line-of-Sight (NLOS) when the UWB signal is obstructed. The attenuation of UWB signals leads to noise and inaccuracy in UWB-only localization. However, IMU sensors, as a complementary modality, optimize UWB localization when the current UWB measurement is identified as an outlier. The localization trajectory obtained using the multimodal system remains robust when both VSLAM and UWB localization fail.

Both repetitive patterns of the roller door and the poorly textured blank wall affect the accuracy of VSLAM. Small errors in orientation or position estimation can lead to significant accumulated drift. When VSLAM fails to provide accurate estimation, The combination of UWB and IMU provides reliable and globally consistent range measurement and corrects the drift. The pre-processing step for outlier removal before EKF fusion effectively avoids the interference of VSLAM failures on the pose estimation. The result illustrates that our multimodal localization system can enhance the system's robustness and overcome the limitations of each localization system by its own.

V. Conclusion

This paper presented a robust multimodal localization approach for indoor construction robotics. The approach integrates multiple sensor readings from RGB-D cameras, UWB and IMU to mitigate drift errors caused by each single sensor. We tested the multimodal localization system under three commonly seen conditions in job sites, including inconsistent ambient lighting, surfaces with repetitive patterns, and poor textured surfaces. In the experiments, the multimodal localization system demonstrates robustness when VSLAM localization fails. When both UWB and VSLAM provide reliable localization, the UWB can reduce the drift error caused by VSLAM, and VSLAM can enhance the accuracy when the UWB is affected by noise in a local region. The limitation of the system lies in the coverage area of UWB sensors. In the experiment, we set up four UWB beacons to cover a room of 75 m^2. The increased area may involve more stationary UWB beacons. In future work, an optimization approach for effective UWB beacon layout can be investigated to enhance the scalability of the localization system for large-scale construction sites such as tunnels. The strategically placed UWB beacons can effectively increase the robustness of the localization system. As a complementary global constraint, UWB can provide reliable localization when sensors such as light detection and ranging (LiDAR) and GNSS fails due to signal blocking, insufficient geometric features, or object occlusion.

Acknowledgement

Carnegie Mellon University's GSA/Provost GuSH Grant funding was used to support this project.

References

[1] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[2] D. Dardari, A. Conti, U. Ferner, A. Giorgetti, and M. Z. Win, "Ranging with ultrawide bandwidth signals in multipath environments," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 404–426, 2009.

[3] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1067–1080, 2007.

[4] Jin-Shyan Lee, Yu-Wei Su, and Chung-Chou Shen, "A comparative study of wireless protocols: Bluetooth, UWB, ZigBee, and Wi-Fi," in *IECON 2007-33rd Annual Conference of the IEEE Industrial Electronics Society*, 2007, pp. 46–51.

[5] Thien-Minh Nguyen, T. H. Nguyen, M. Cao, Z. Qiu, and L. Xie, "Integrated uwb-vision approach for autonomous docking of uavs in gps-denied environments," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 9603–9609.

[6] H. Xu, L. Wang, Y. Zhang, K. Qiu, and S. Shen, "Decentralized visual-inertial-uwb fusion for relative state estimation of aerial swarm," in *2020 IEEE international conference on robotics and automation (ICRA)*, 2020, pp. 8776–8782.

[7] T. H. Nguyen, Thien-Minh Nguyen, and L. Xie, "Tightly-coupled ultra-wideband-aided monocular visual SLAM with degenerate anchor configurations," *Autonomous Robots*, vol. 44, no. 8, pp. 1519–1534, 2020.

[8] F. Liu, J. Zhang, J. Wang, H. Han, and D. Yang, "An UWB/vision fusion scheme for determining pedestrians' indoor location," *Sensors*, vol. 20, no. 4, p. 1139, 2020.

[9] Huei-Yung Lin and Ming-Chi Yeh, "Drift-Free Visual SLAM for Mobile Robot Localization by Integrating UWB Technology," *IEEE Access*, vol. 10, pp. 93 636–93 645, 2022.

[10] Y. Cao and G. Beltrame, "VIR-SLAM: Visual, inertial, and ranging SLAM for single and multi-robot systems," *Autonomous Robots*, vol. 45, pp. 905–917, 2021.

[11] F. J. Perez-Grau, F. Caballero, L. Merino, and A. Viguria, "Multi-modal mapping and localization of unmanned aerial robots based on ultra-wideband and RGB-D sensing," in *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, 2017, pp. 3495–3502.

[12] Y. Song, M. Guan, W. P. Tay, C. L. Law, and C. Wen, "Uwb/lidar fusion for cooperative range-only slam," in *2019 international conference on robotics and automation (ICRA)*, 2019, pp. 6568–6574.

[13] C. Wang, H. Zhang, Thien-Minh Nguyen, and L. Xie, "Ultra-wideband aided fast localization and mapping system," in *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, 2017, pp. 1602–1609.

[14] T. H. Nguyen, Thien-Minh Nguyen, and L. Xie, "Range-focused fusion of camera-IMU-UWB for accurate and drift-reduced localization," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1678–1685, 2021.

[15] Thien-Minh Nguyen, S. Yuan, M. Cao, Y. Lyu, T. H. Nguyen, and L. Xie, "Ntu viral: A visual-inertial-ranging-lidar dataset, from an aerial vehicle viewpoint," *The International Journal of Robotics Research*, vol. 41, no. 3, pp. 270–280, 2022.

[16] E. Fernando, O. D. Silva, G. K. Mann, and R. G. Gosine, "Observability analysis of position estimation for quadrotors with modified dynamics and range measurements," in *2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, 2019, pp. 2783–2788.

[17] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE transactions on robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.