IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 7 August 2022, accepted 28 August 2022, date of publication 1 September 2022, date of current version 12 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3203438

RESEARCH ARTICLE

Drift-Free Visual SLAM for Mobile Robot Localization by Integrating UWB Technology

HUEI-YUNG LIN[1,2], (Senior Member, IEEE), AND MING-CHI YEH[2]
[1]Department of Computer Science and Information Engineering, National Taipei University of Technology, Taipei 106344, Taiwan
[2]Department of Electrical Engineering, National Chung Cheng University, Chiayi 621301, Taiwan

Corresponding author: Huei-Yung Lin (lin@ntut.edu.tw)

This work was supported in part by the Ministry of Science and Technology of Taiwan under Grant MOST 106-2221-E-194-004.

ABSTRACT Visual simultaneous localization and mapping (vSLAM or visual SLAM) is an important technique for mobile robot localization in the global navigation satellite system denied (GNSS-denied) environments. However, the positioning accuracy could become unbearable due to the lack of image feature points when the robot navigates in a spacious indoor space. The drifting errors accumulated over time are generally inevitable and need to be mitigated by more sophisticated loop-closure algorithms. In this paper, we propose a drift-free visual SLAM technique for mobile robot localization by integrating the ultra-wideband (UWB) positioning technology. The basic concept is to utilize the global constraint of the UWB positioning to reduce the locally accumulated errors of visual SLAM localization based on the extended Kalman filtering (EKF) framework. In our experimental results, various SLAM approaches are performed in the indoor scenes, and the evaluation and comparison have demonstrated the feasibility of the proposed localization technique. By the integration of UWB positioning, the overall drift error of the robot navigation is reduced for more than 50%.

INDEX TERMS Simultaneous localization and mapping, robot vision, ultra-wideband, indoor positioning.

I. INTRODUCTION

Nowadays, the manufacturers have gradually considered smart factories and smart robots as some important aspects of the development trend. When the Industry 4.0 is ready for the world, robotics plays a key role for industrial automation. In the dangerous working environments or performing the tasks with high repetition, it is expect to replace the human operations with the robots capable of perception and proper reaction. This is commonly achieved by industrial robots with the visual servoing capability. Thus, sensing techniques and data processing algorithms are the core technologies for automated factories in the future. Among the various sub-systems used for mobile robots, the self-localization module is with specific importance in robot mobility. When production lines are transformed to automated industrial systems to increase the throughput with high reliability, it is necessary to provide precise positioning and sufficient accuracy for robot localization.

The associate editor coordinating the review of this manuscript and approving it for publication was Guilin Yang.

Over the past few decades, technologies for simultaneous localization and mapping (SLAM) have been widely investigated [1], [2]. The objective is to obtain the location information of the mobile robot and construct the map of the environment during the robot navigation at the same time. Some commonly used sensors include infrared, sonar, 2D and 3D light detection and ranging (LiDAR), monocular camera, stereo camera system, depth camera, and inertial measurement unit (IMU). In addition to the development of SLAM techniques with single sensors, the fusion of different SLAM approaches has been shown to provide better localization accuracy [3], [4]. Nevertheless, there still exist many unsolved problems even the sensor fusion methods are incorporated. Some typical examples include the slipperiness of wheel contacts and the influence of unknown external forces. These might introduce some errors which are difficult to mitigate by the SLAM systems using on-board sensors alone, and therefore more sophisticated loop closure detection needs to be adopted [5].

The most important problem for a SLAM system is to deal with the measurement errors in the position and orientation.

The uncertainties could be induced by the restricted measurement accuracy and the influence of the environment. In general, each measurement or feature extraction contain some levels of errors. Therefore, a total uncertainty of a SLAM technique will expand during a long range navigation. Consequently, the cumulative error due to the drifting in motion cannot be ignored. The inaccurate localization or positioning results might cause the incorrect operations for industrial robot applications.

As the recent progress of image sensing technologies, low-cost cameras with high image quality have been widely used for robot navigation. Along with the maturity of computer vision algorithms, visual SLAM has become one of the most promising approaches for mobile robot localization. Since the vision-based methods heavily rely on feature point extraction, the correspondence matching could be failed due to the occlusion or illumination condition changes. Thus, visual SLAM techniques usually suffer from the limited accuracy caused by accumulation errors, especially in an open large space. Meanwhile, the GNSS is able to provide a certain degree of positioning accuracy independent of the location and navigation path. Nevertheless, the GNSS signals from satellites are not able to enter the buildings, which makes the approach not applicable to indoor localization. Alternatively, several technologies with local signal broadcast, such as Wi-Fi [6], radio-frequency identification (RFID) [7], ultra-wideband (UWB) [8] and Bluetooth [9], have been developed for indoor positioning.

The UWB positioning technology has many advantages over traditional methods such as GNSS, Wi-Fi and RFID. It is able to achieve high-precision range sensing owing to the high sampling rate and possible obstacle penetration of the signals [10]. One essential issue of an indoor UWB positioning system is to estimate the distance. The most common approach is based on the time of arrival (TOA). Several implementation challenges, including non-line-of-sight (NLOS), multi-path and electronic jamming, are the emerging problems to be addressed in the development [11].

In this paper, we propose a mobile robot localization technique by integrating various measurement and sensing techniques. The drifting errors introduced by the computation from proprioceptive sensors is mitigated through the integration with UWB positioning. This approach is designed to cope with the problem of current visual SLAM systems in terms of accuracy when operating in the featureless scenes. It aims for the mobile robot localization in a relatively spacious environment where the derivation of accurate visual SLAM results is more difficult. At the same time, the multi-path problem of UWB signals will have less influence on the positioning accuracy. Built upon the RTAB-Map (Real-Time Appearance-Based Mapping) SLAM framework with 3D data acquisition, UWB positioning is integrated to provide a global constraint. When the SLAM system starts, an independent 3D localization thread is also carried out by UWB simultaneously. The extended Kalman filter (EKF) is then adopted to fuse the localization results based on the confidence weighting [12]. In our proposed method, the localization failure and drifting errors can be monitored continuously, and the UWB positioning is adopted for the optimization of the SLAM output whenever the uncertainty is larger than a threshold. Several experiments are performed in the real-world environments, and the performance evaluation has demonstrated the feasibility of our approach for precise indoor localization.

II. RELATED WORK

In the robotics research community, many SLAM approaches have been proposed over the last decades [13]. Most classic approaches in the existing literature utilize laser rangefinders (lidars) or cameras for environmental data acquisition. Among various SLAM techniques, the lidar based approaches have been extensively investigated [14]. These techniques are frequently regarded as the main localization methods, and perform relatively well in robot navigation tasks. Generally speaking, 2D lidars are adopted for domestic applications such as cleaning robots or region exploration, while 3D lidars are utilized in the applications of self-driving vehicles or aerial robots. Those obstacles in the detectable region are perceived as 3D point clouds with the depth information. In the lidar-based techniques, the 3D data obtained from different locations are registered to a common coordinate frame for comparison, and then used to calculate their relative orientation and position [15]. The frame-by-frame transformation is then used for mobile robot localization and environment map construction.

Due to the advances of sensing technologies and machine learning algorithms, vision-based SLAM approaches have received considerable attentions in recent years. The existing methods uses the rich image information from the scenes for place recognition and location computation [16], [17]. Similar to the lidar-based approach, the visual SLAM techniques developed with depth cameras directly use the acquired point cloud data for 3D map construction and robot localization. Some recent popular SLAM systems using depth cameras for data acquisition include ORB-SLAM2 [18] and S-PTAM [19]. To reduce the drifting error under long distance navigation, the loop closure based on the feature description mapping is performed rather than using the bag-of-words. There also exist vision-based SLAM techniques incorporating the IMU measurements to improve the local trajectory accuracy. The implementation of maplab [20] and VINS-Mono [21] uses only an RGB camera and IMU to provide the visualization of environment maps. In this research, we adopt RTAB-Map [22] as our framework for the development of drift-free visual SLAM techniques. Different from the previous approaches which also take RGB-D images as input, it utilizes 'working', 'long-term' and 'short-term' for memory management. RTAB-Map can take care of online operations under large-scale and long-term conditions while successfully performing the closed-loop detection. Thus, it is commonly adopted as a building block for the development of advanced SLAM techniques.

Except for the visual SLAM systems, there are also a number of well-studied signal-based methods for indoor localization. Some widely used technologies include Bluetooth, RFID, Wi-Fi and UWB [23]. In general, the Bluetooth technology consumes much less power compared to others, but it is not satisfactory in the localization accuracy (about 1 meter), even with a number of beacons installed in the environment [24]. RFID is based on the radio frequency identification technology. The electromagnetic (EM) transmission characteristics of radio frequency (RF) signals are used to realize the data communication between the RF tags and readers. There are many advantages of RFID, including fast data transmission, high-level security, less non-RF communication interference [25]. However, a large number of RFID devices are generally requires for the indoor positioning task, which is not suitable for many application scenarios.

In the early development of indoor positioning technologies, the nearest neighbor and cross positioning methods were adopted by Wi-Fi for implementation. The nearby Wi-Fi base stations (or hot spots) in a reachable distance are detected by the closest neighbor approach. If there exist multiple sources of signals identified in a small region, the localization accuracy can be further improved based on the estimation of the current location via the triangulation cross positioning. Since most indoor environments have the Wi-Fi infrastructure readily available, it is generally not necessary for additional hardware setup for positioning. Thus, Wi-Fi positioning techniques have been integrated into the SLAM framework in recent studies. As demonstrated in [26] and [27], the Wi-Fi positioning can be used to optimize the localization outputs from Lidar SLAM and RGB-D SLAM. The major drawback of the Wi-Fi implementation is its power usage requirement and the deployment of hotspots. Furthermore, the positioning results might be affected by the channel interference issue.

For the existing indoor localization techniques, the ultra-wideband requires more power consumption and is more expensive compared to other methods such as RFID, Wi-Fi and Bluetooth Low Energy (BLE). However, it is able to achieve the accuracy of 10 cm when optimized with some positioning algorithms. In the recent study, [28] integrated the sensing outputs from the RGB-D camera and UWB, and utilized the relation between the UWB positioning and the 3D point clouds to construct an environment map. By fusing the RGB-D and UWB information, it provides a stable system for map building and localization. Similarly, [29] integrated the UWB positioning with a 2D LiDAR. Their proposed method used the 2D LiDAR data to increase the accuracy of UWB positioning by using complete surrounding perception data of the environment. Alternatively, the depth measurement from UWB can also reduce the error accumulation in lidar-based SLAM techniques to provide complementary effects [30]. In [31], Magnago *et al.* have demonstrated the accuracy in large-scale indoor scenes for the localization technique based on odometry-assisted UWB ranging.

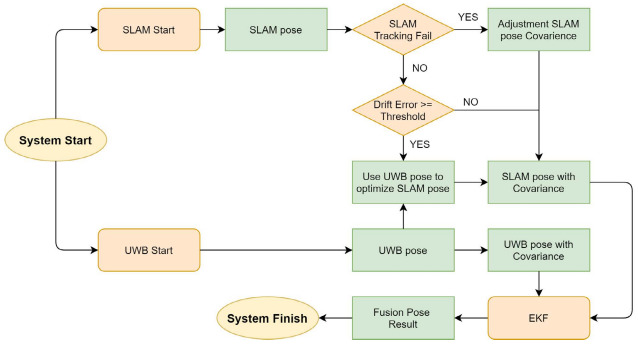


FIGURE 1. The system flowchart of the proposed drift-free visual SLAM technique.

For the EKF algorithm used in this work, it is based on the implementation in [3] and [32]. From the modern control theory and statistical data processing, the noise corrupted state vectors of a system can be estimated by the iteration of data measurements. Thus, the objective is to derive the optimal value of the current system state by the one estimated at the previous time stamp and the present measurement. When adopted for the mobile robot localization, this is to compute the 3-D state vector representing the position. By defining the noise with Gaussian distribution, the current state vector and the prediction error covariance matrix are obtained according to the kinematic model derived from Newtonian mechanics.

III. PROPOSED APPROACH

The system flowchart of the proposed drift-free visual SLAM technique is illustrated in Figure 1. In this work, we adopt RTAB-Map (Real-Time Appearance-Based Mapping) as the basic SLAM framework for the development [33], and incorporate the pre-established UWB positioning system for global localization. As shown in the figure, the SLAM system on-board the robot for self-localization operates simultaneously with the UWB global positioning. By fusing the multimodal sources for robot localization, the high accurate results with drifting corrections can be obtained based on the weightings of confidence levels. In addition, the data fusion of SLAM and UWB is carried out via EKF (extended Kalman filter). If the measurement difference between two systems is greater than a threshold, a relocalization procedure is performed based on the UWB global information.

The RTAB-Map module adopted in this work is based on a general RGB-D SLAM framework. It is also capable of using memory management to perform the closed-loop detection. The essential task is to make it possible to accomplish a long range navigation with the online loop closure detection. In the general SLAM framework, the computations of visual odometry for frame-to-frame and frame-to-map are usually carried out separately. The input images are acquired from stereo or RGB-D cameras, and the frame-to-map feature extraction is performed without directly matching for the nearest neighbors. Under the situation of feature loss, the ratio of the first and second nearest neighbors is compared and

used to perform the matching. When applying the frame-to-frame approach, the optical flow is adopted for feature matching computation. Due to the discontinuous and non-uniform properties of optical flow fields, it is possible to detect the dynamic object or camera movement. In this method, the requirement for the extraction of feature descriptors is removed to better calculate the camera motion trajectory.

During the motion prediction process, the key frame position and the associated feature map in the current frame are derived using the action in the previous frame. In this process, the search window adopted for feature matching can be restricted to a suitable range. Consequently, better matches can be obtained even in the environments with repetitive patterns or dynamic objects. During the motion estimation process, PnP (perspective-n-point) and RANSAC (random sample consensus) algorithms are utilized to compute the transformations of image features, the current and key frames [34], [35]. It is then followed by a local bundle adjustment for the key frames and using the geometric consistency for camera pose updates.

To provide a global constraint for localization, we incorporate the UWB positioning technique in this work. When performing the two-dimensional computation, at least three base stations are required. But it is necessary to have at least four base stations for the calculation of three-dimensional position. For the current development, the methods based on time-of-flight (TOF), time-of-arrival (TOA) and time-difference-of-arrival (TDOA) are commonly adopted. In the two-way time-of-flight (TW-TOF) implementation, independent time stamps are generated from the initialization of all UWB modules. It is assumed that a request pulse signal at T_{a1} is transmitted by the transmitter of module A at its time stamp. Similarly, a pulse signal at T_{b1} is received by the receiver of module B at its time stamp. By processing the signals, a responsive signal transmitted by module B at the time stamp T_{b2} is received by module A at its time stamp T_{a2}. The time of flight associated with the pulse signals between these modules is used to compute the distance by

$$S = \frac{C}{2}[(T_{a2} - T_{a1}) - (T_{b2} - T_{b1})] \tag{1}$$

where C is the speed of light [36]. If the positioning is carried out with multiple modules, it is required to calculate the distances from the label to those modules. In this configuration, the 3D coordinates will be derived using a trilateral method.

Figure 2 depicts an illustration of TDOA. It is based on TOA, but improved by the less requirements on synchronization in time. The tag position is determined by computing the time difference between the signals arriving the modules. Once the tag is under the coverage of a base station, the signals will be transmitted actively from the positioned tag to the positioned base station. The two positions of the positioning tag are then computed by the positioning system based on the difference of arrival times of the signals received by the base stations. According to this calculation, the coordinates of the tag are obtained by the intersections of multiple

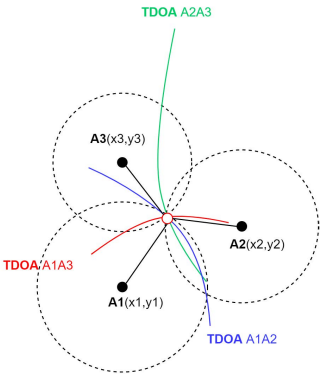


FIGURE 2. A schematic diagram of the UWB positioning using the TDOA (time-difference-of-arrival) approach. This method is an improvement of TOA, and has the advantage of lower time synchronization requirements.

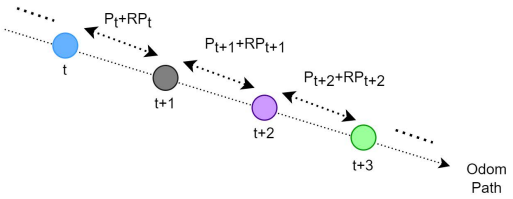


FIGURE 3. The localization trajectory in the global scale is split to a frame-by-frame basis to estimate the relative position of the robot movement.

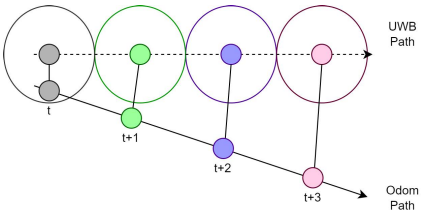


FIGURE 4. The difference of positioning results between the odometry and UWB techniques is used for localization adjustment. If it is larger than a preset value, the measurement is restricted according to the time frame.

hyperbolas. To have this method operate successfully, one crucial prerequisite is the synchronization in time among the different modules. In general, TDOA is more complicated for deployment and setup because of the time synchronization requirement between different modules. However, it only requires to transmit the signals instead of waiting for the responses from the tags. Consequently, more spare time for the localization of other tags can be achieved by using this approach.

Since the UWB positioning system measures the distance and performs the localization using TDOA, this approach can be considered as GNSS positioning in the indoor environment. Thus, the drifting error due to the long range mobile robot navigation can be avoided. The measurement error will be fixed within a certain range with no accumulation, and the localization estimates will approximate the ground-truth globally. As illustrated in Figure 3, the relative poses among multiple frames are generally estimated by SLAM based localization techniques. To apply UWB measurements to the

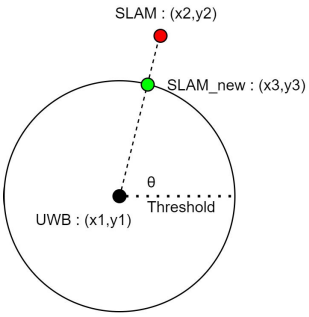


FIGURE 5. The schematic diagram for the UWB error correction. If the difference between the newly observed point and the previous position estimated by UWB is larger than a preset value, new coordinates will be used for the localization update.

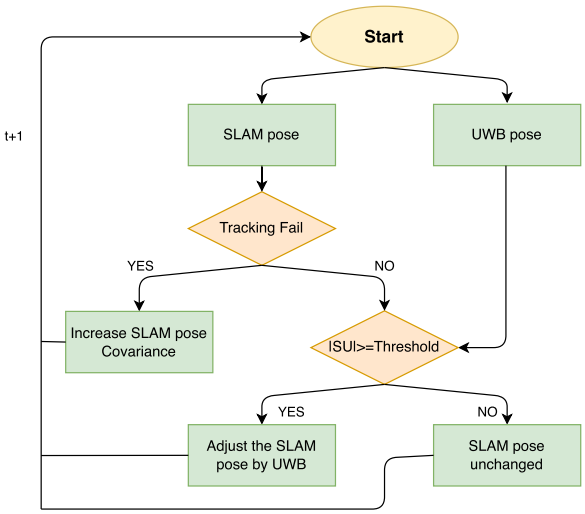


FIGURE 6. The flowchart of the proposed UWB-based SLAM drifting error correction approach.

localization estimates computed by SLAM techniques without losing short-term accuracy, the relative poses estimated from the consecutive SLAM frames are extracted and split to multiple ones. Let the localization position at time t be P_t, and the transformation between the time instants t and $t+1$ be RP_t. Then the relative position can be computed as $(\Delta x, \Delta y)$ frame-by-frame. The computation processes of UWB and SLAM at the same time are aligned in time, and the localization results are cross-referenced. If the difference between the estimates is large than a pre-defined value, then the positioning result will be limited according to the time frame as illustrated in Figure 4.

Due to the inherent limitation of the UWB positioning technique, large errors are always present in the vertical direction. Thus, we consider the difference between the points in the x and y directions for location restriction and update. Given two points (x_1, y_1) and (x_2, y_2), we update the new positioning result by

$$\Delta_{xy} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{2}$$

The idea of the overall error correction based on the measurements of SLAM and UWB is illustrated in Figure 5. If the newly created position from SLAM and the current position estimate obtained by UWB have a difference larger than a value τ, then we use the new position coordinates for an update. The new localization position (x_3, y_3) has the same direction as (x_2, y_2), but provided an offset given by the threshold as follows

$$x_3 = \begin{cases} x_2 + \Delta x, & \text{if } \Delta_{xy} < \tau \\ x_1 + \tau cos\theta, & \text{if } \Delta_{xy} \geq \tau \end{cases} \tag{3}$$

$$y_3 = \begin{cases} y_2 + \Delta y, & \text{if } \Delta_{xy} < \tau \\ y_1 + \tau sin\theta, & \text{if } \Delta_{xy} \geq \tau \end{cases} \tag{4}$$

where Δx and Δy are the displacements in the x and y directions respectively, and Δ_{xy} is the 2-D displacement in the Euclidean distance. While the error correction is carried out in the x and y directions, the drift error in the z axis is filtered out by EKF with the 2D mode setting.

The localization sensor fusion of SLAM and UWB positioning is further carried out using the extended Kalman filter (EKF). We assume that the probabilities of localization follow the Gaussian distribution initially for iterative updates, the truncation operations in Eqs. (3) and (4) are then used to derive the final result. The Kalman filter was initially used to cope with a linear-quadratic equation by finding the solutions recursively [37]. It is a problem for estimating the current state (or the state of process) of a linear dynamic system under the perturbation of white noise. To properly estimate or predict the current state, the Kalman filter proceeds recursively by utilizing the current measurements and the previous states. Since the Kalman filter performs in a recursive way till the optimal estimation state is achieved, it is commonly regarded as a powerful technique to minimize the error of state estimates. The EKF is an extended version of the Kalman filter to deal with non-linear models. In EKF, it includes an extra linearization model in the prediction step and the calculation of partial derivatives of the state variables.

Given an $n \times 1$ process state vector x_k, an $m \times 1$ measurement vector z_k, and the control input u_k, where k denotes the time stamp, then a general non-linear system and the measurement model can be described by

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) + w_k \\ z_k &= h(x_k) + v_k \end{aligned} \tag{5}$$

where the random variables w_k and v_k represent the Gaussian white noise and the measurement noise, respectively. Let P_k, Q_k and R_k be the covariance matrices for x_k, w_k and v_k, respectively, the EKF algorithm is carried out with two steps: *prediction (time update)* and *correction (measurement update)*. In the prediction update step, the state projections and error covariance estimates are computed from

$$\begin{aligned} \hat{x}_{k+1}^{-} &= f(x_k, u_k) \\ P_{k+1}^{-} &= F_k P_k F_k^T + G_u Q_k G_u^T \end{aligned} \tag{6}$$

where

$$F_k = \frac{\partial f}{\partial X} = \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial \phi} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial \phi} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} & \frac{\partial f_3}{\partial \phi} \end{pmatrix} \tag{7}$$

and

$$G_u = \frac{\partial f}{\partial u} = \begin{pmatrix} \frac{\partial f_1}{\partial u_1} & \frac{\partial f_1}{\partial u_2} \\ \frac{\partial f_2}{\partial u_1} & \frac{\partial f_2}{\partial u_2} \\ \frac{\partial f_3}{\partial u_1} & \frac{\partial f_3}{\partial u_2} \end{pmatrix} \tag{8}$$

In the measurement update step, the measurement z_k becomes available and EKF calculates the Kalman gain matrix. It is then incorporated with the measurement innovation to derive the estimated state x_k, followed by the state error covariance matrix update. The general scheme for the measurement update is given by

$$\begin{aligned} K_k &= P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \\ \hat{x}_k &= \hat{x}_k^- + K_k(z_k - h(\hat{x}_k^-)) \\ P_k &= (I - K_k H_k) P_k^- \end{aligned} \tag{9}$$

where H is the Jacobean matrix of the measurement and computed by $H = \partial h / \partial X$, and $\hat{x}_k$ is an estimate of the state vector x_k. The discrepancy between the predicted and observed measurements is commonly referred to as the measurement innovation or the residual, and denoted by $r_k = z_k - h(\hat{x}_k^-)$. A major disadvantaged of EKF is that if the estimated process and measurement noise are not precisely modeled, the filter would diverge and lead to a system inconsistence. Practically, well-defined error covariance models are seldom achieved so the values are set manually. It may take a lot of time with trial and error before getting an expected model.

In the conventional localization framework, it is usually assumed that the environment map is available. The extended Kalman filter estimates the robot position by integrating the information acquired from odometers, map and exteroceptive sensors. In this work, our implementation consists of the following steps: pose prediction, observation, measurement prediction, matching and estimation.

- *Pose Prediction:* The robot pose at time $k+1$ is predicted based on the previous position and orientation (at time k) and its movement with the control input u_k, and is given by

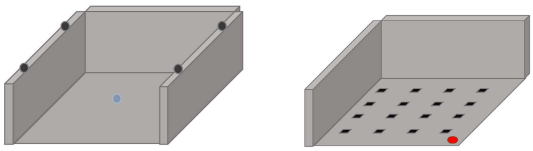
$$\begin{aligned} \hat{x}_k^- &= f(x_k, u_k) \\ P_{k+1}^- &= F_k P_k F_k^T + G_u Q_k G_u^T \end{aligned} \tag{10}$$

  where $F_k = \partial f / \partial X$ and $G_u = \partial f / \partial u$.
- *Actual Observation:* In the second step the robot obtains the sensor measurements z_{k+1} at time $k+1$.
- *Observation Prediction:* From the predicted robot pose $\hat{x}_{k+1}^-$ and the currently available map, a predicted measurement $\hat{z}_k$ is generated by

$$\hat{z}_k = \begin{pmatrix} \sqrt{(\hat{x} - x_i)^2 + (\hat{y} - y_i)^2} \\ \tan^{-1} \frac{\hat{y} - y_i}{\hat{x} - x_i} - \hat{\phi} \end{pmatrix} \tag{11}$$

(a) (b)

FIGURE 7. (a) A Pioneer P3-DX mobile robot with an aluminum extrusion rack is used in the experiment. (b) The evaluation is carried out in a large indoor environment.



(a) (b)

FIGURE 8. (a) The UWB setting adopted in our experiment with anchors placed at the same height. (b) The AprilTags are placed in a rectangular grid for localization comparison.

  The state error between the actual and predicted measurements is $v_k = z_k - \hat{z}_k$, where v_k is the innovation sequence (or residual) with the covariance $S_k = H_k P_k^- H_k^T + R_k$, and $H_k = \partial h / \partial X$
- *Matching:* In the matching procedure, an assignment is processed from measurements to the landmarks and then store in the map.
- *Estimation:* The estimation of the state is given by

$$\begin{aligned} \hat{x} &= \hat{x}_k^- + K_k(z_k - h(\hat{x}_k^-)) \\ K_k &= P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \\ P_k &= (I - K_k H_k) P_k^- \end{aligned} \tag{12}$$

  where K_k is the Kalman gain and P_k is the new state covariance matrix.

The flowchart of the proposed UWB-based SLAM drifting error correction approach is illustrated in Figure 6. When both the SLAM and UWB systems start the positioning process, the localization status is continuously monitored. If a localization failure of the visual SLAM is detected, the coefficients of the covariance matrix are changed to a large value to represent its low confidence. Otherwise, the difference between the SLAM and UWB positioning results is calculated, and a threshold is used to determine whether the drifting error correction will be performed. The noise covariance matrix of the EKF is then updated in the diagonal elements. Since the estimates from SLAM and UWB are our main concerns, the weights for the position and orientation (6 DoFs) are set as larger values compared to the velocities and angular velocities in 3 directions.

IV. EXPERIMENTS

The proposed drift-free visual SLAM with UWB technique is implemented on a mobile robot system, and tested in the real

world environment. Figure 7 shows the mobile robot and the indoor space for our experiments. We construct an aluminum extrusion rack and install it on a Pioneer P3-DX robot to place the sensing and computing devices (See Figure 7(a)). An Intel Realsense D455 depth camera and a laptop computer are placed on top for sensor data acquisition and SLAM computation. For the UWB positioning, a tag is mounted below the camera system. We create a new local coordinate system, and define the relationship among the camera and UWB coordinate frames. It is used to ensure that there is no conflict between the localization results obtained from visual SLAM and UWB positioning. The initial test on visual SLAM is carried out in the scenes with sufficient feature points, and good localization results can be derived with high stability. When the visual SLAM is performed in a spacious environment, however, the localization failure problem occurs. Thus, our experiments focus on the evaluation of UWB-assisted localization techniques in an open space with several navigation paths. In the current setting, the update rates for UWB and SLAM are 50 Hz and 30 Hz, respectively.

The principle of UWB for indoor positioning is the same as GNSS for outdoor positioning using multiple satellites. In the indoor environment, several UWB base stations are placed at some given locations with coordinate information, and the UWB tags are mounted on the mobile robot. Figure 8(a) depicts a general configuration of the UWB setting. Signal pulses are emitted from the tag at a fixed frequency, and the continuous communication is established between the base stations and the tag. Simultaneously, the tag location is computed by the localization algorithm using the ranging data from the base stations. For the environment for our experiments, the visual SLAM system is performed in an open and large indoor area as shown in Figure 7(b), and achieving high accuracy localization is a very challenging task. Thus, we also deploy several AprilTags as landmarks in the open space for localization assistance [38]. Figure 8(b) depicts the current AprilTag setting, where 16 landmarks are attached on the ground and arranged as a rectangular grid.

When a robot navigates around a spacious indoor environment, it is inevitably to result in localization drifting errors even some landmarks such as AprilTags are adopted. In the experiments, large drifting errors can still occur when a mobile platform moves near the walls where image features are easily detected. Figure 9 shows the trajectories obtained using the SLAM and UWB positioning systems (presented in red and blue curves), respectively. Figure 10 illustrates the localization results derived by the proposed technique with the fusion of SLAM and UWB. The trajectories obtained with the threshold settings using 2 m and 0.5 m are shown in Figures 10(a) and 10(b), respectively. The processing frame rates for RTAB-Map and UWB are 20 Hz and 50 Hz, respectively. Under the EKF fusion, the localization computation is at about 15 Hz.

Because the UWB positioning technique is relatively stable in the global scale but with less accuracy in a local region, we set the covariance matrix parameter to 0.5 in the

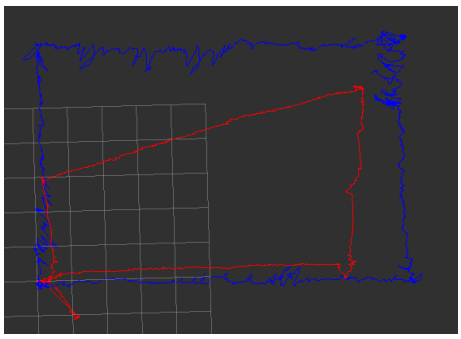
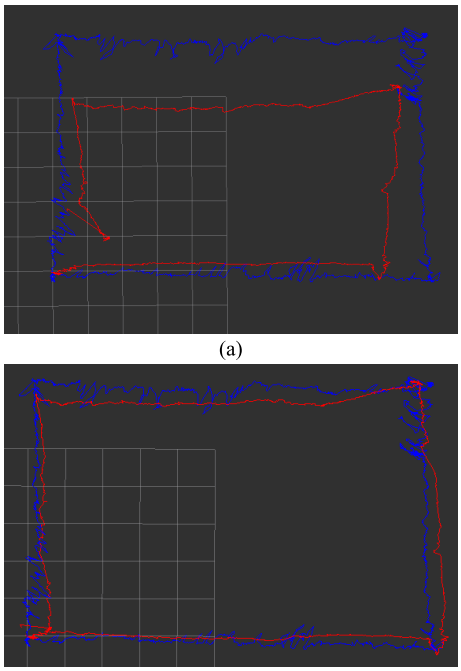


FIGURE 9. The trajectories obtained from RTAB-Map SLAM and the UWB positioning (marked in red and blue). Each block in the figure is 1 square meter, and horizontal and vertical axes are the x- and y- axis, respectively.



(a)

(b)

FIGURE 10. The localization results derived using the our approach by the integration of UWB positioning and visual SLAM technique under different parameter settings. (a) The output using the threshold of 2 meters. (b) The output using the threshold of 0.5 meters.

experiment. It is followed by setting the SLAM covariance matrix parameter to 1 to perform the EKF for localization fusion. In Figure 12(a), the blue, red and green curves represent the localization trajectories obtained using UWB, SLAM and the EKF fusion. The figure illustrates that, when the visual SLAM system does not localize successfully, the EKF will still utilize to the positioning result from the previous covariance matrix computation. As a result, large errors will present in the final fusion trajectory. In the implementation, we change the covariance matrix parameter to 999 (or a huge number) if the SLAM has a tracking failure detected. As the fusion trajectories indicated in Figure 12(b), the SLAM failure does not affect the result of EKF fusion when the proposed adjustment is used.

For the evaluation of SLAM algorithms with ground-truth data, most RGB-D datasets publicly available are collected

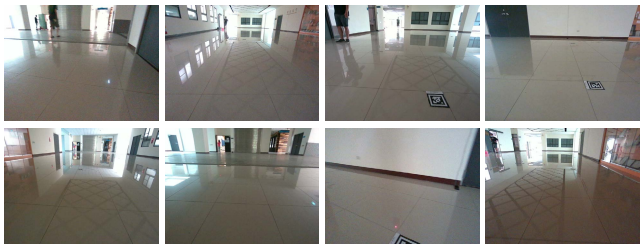


FIGURE 11. Some of the images captured and used for the visual SLAM computation.

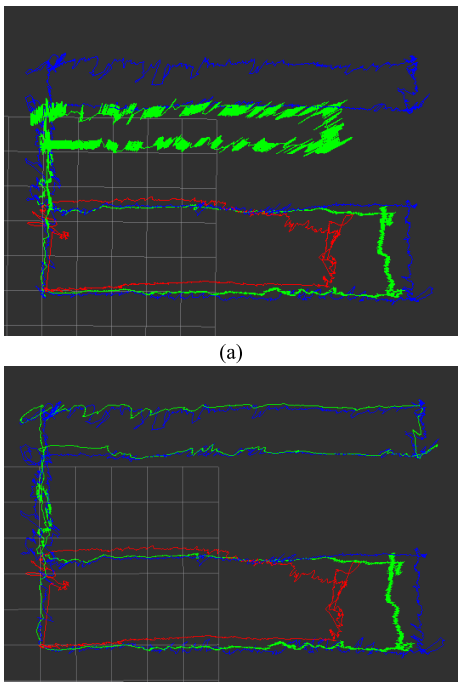


FIGURE 12. The blue, red and green curves represent the localization trajectories obtained using UWB, SLAM and the EKF fusion. (a) The trajectories are derived using the covariance set as 1. (b) The trajectories are derived using the covariance setting of 999.

with high-precision motion capture systems. In the experiments, the localization results obtained from AprilTags are used for accuracy comparison. We calculate the camera poses based on the coordinate origins of individual AprilTag settings. To evaluate the accuracy by absolute trajectory errors, a series of discrete samples are taken and used as reference ground-truth locations. The relative errors in the results of UWB, SLAM and EKF fusion are derived with N sample points using

$$\begin{aligned} x_{error} &= \frac{\sum_{i=1}^{N}\sqrt{(x_{est}(i)-x_{gt}(i))^2}}{N} \\ y_{error} &= \frac{\sum_{i=1}^{N}\sqrt{(y_{est}(i)-y_{gt}(i))^2}}{N} \\ z_{error} &= \frac{\sum_{i=1}^{N}\sqrt{(z_{est}(i)-z_{gt}(i))^2}}{N} \end{aligned} \tag{13}$$

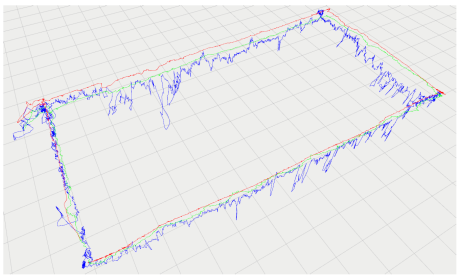


FIGURE 13. The experiment for a mobile robot navigating in a rectangular trajectory. Each block in the figure is 1 square meter, and horizontal and vertical axes are the x- and y- axis, respectively.

TABLE 1. The translation errors of the experiment when the robot moves in a rectangular path as shown in Figure 13.

	SLAM	SLAM'	UWB	EKF
x ave. error	$0.647m$	$0.182m$	$0.175m$	$0.146m$
y ave. error	$1.399m$	$0.387m$	$0.188m$	$0.245m$
z ave. error	$0.709m$	$0.528m$	$0.391m$	$0.081m$
# of samples	1372	1372	1372	1372

where x_{error}, y_{error} and z_{error} are the mean ATE (absolute trajectory error) along the three axes of the translation matrix T. The translation estimate and the ground-truth position at the sample point i are denoted by $(x_{est}(i), y_{est}(i), z_{est}(i))$ and $(x_{gt}(i), y_{gt}(i), z_{gt}(i))$, respectively. We conduct the experiments with several navigation trajectories in an indoor scene, and evaluate with UWB, SLAM and EKF fusion localization.

The experimental result of a mobile robot navigating in a rectangular path is shown in Figure 13, with the localization trajectories derived from UWB, SLAM and the EKF fusion marked in blue, red and green, respectively. Table 1 tabulates the translation errors from different localization methods, where SLAM' indicates the SLAM result with drifting correction. The table shows that the mean errors along the x and y axes are reduced from 0.647 meters to 0.182 meters and from 1.399 meters to 0.387 meters after the drift correction, respectively. The results illustrate that our proposed technique is able to provide great improvements on drifting errors. In the experiment, the EKF fusion can generally suppress the translation error. Table 1 also indicates that the UWB positioning has slight better performance in the y direction, despite the noisy trajectory as illustrated in Figure 13. The main reason is the evaluation carried out only on discrete sample locations, which results in the UWB localization approximate the ground-truth even if the complete path is rather noisy. Figure 14 shows the CDF (cumulative distribution function) of the positioning error per meter. The mean and standard deviation of the EKF fusion results are 0.616 and 0.045, respectively. Since the proposed technique is able to perform in real-time, we consider the comparison in terms of computation time is not significant.

The second experiment is performed on a more complicated navigation path to evaluate the stability of the proposed technique in a spacious indoor space. The mobile robot moves in straight lines, curves and irregular paths rather than travels

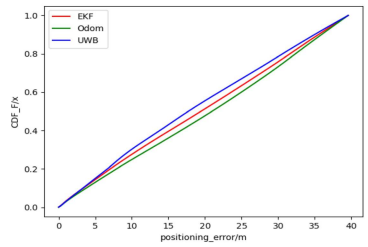


FIGURE 14. The cumulative distribution function (CDF) of the positioning error per meter.

FIGURE 15. Some of the images captured and used for the visual SLAM computation.

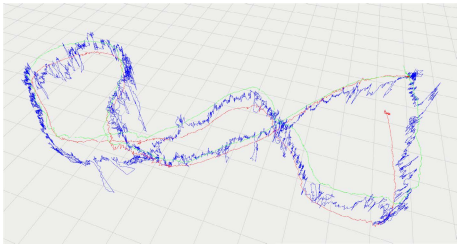


FIGURE 16. The experiment carried out on a more complicated navigation path to evaluate the stability of the proposed technique in a spacious indoor space. The resulting trajectory contains many harsh positioning conditions with few feature points, but our method can still provide good localization accuracy with less drifting errors.

TABLE 2. The translation errors of the experiment when the robot moves in an arbitrary path as shown in Figure 16.

	SLAM	SLAM'	UWB	EKF
x ave. error	$0.661m$	$0.183m$	$0.203m$	$0.183m$
y ave. error	$0.547m$	$0.346m$	$0.247m$	$0.258m$
z ave. error	$1.246m$	$0.746m$	$0.433m$	$0.103m$
# of samples	1503	1503	1503	1503

against the wall. We have performed several empirical tests and found the threshold of 0.5 m is most suitable for the restriction on drifting errors while preserving the short-term result of visual SLAM. As illustrated in Figure 16, the resulting trajectory contains many harsh positioning conditions with few feature points, but our method can still provide good localization accuracy with less drifting errors. The evaluation on the translation errors is tabulated in Table 2. As indicated in the table, although the drifting is more severe in the z-axis due to the lack of detected features for localization, we are able to suppress the errors in all directions as shown in SLAM'. In addition, the EKF fusion result not only successfully reduces the error, but also avoids the UWB positioning noise. This has demonstrated the feasibility of our approach for many challenging situations in occurred practical applications.

V. CONCLUSION

One major issue of current SLAM systems is the accumulation of drift errors for the long range navigation. In this work, we present an approach for indoor localization by integrating different positioning approaches. The basic idea is to mitigate the localization errors introduced by proprioceptive sensors through the integration with UWB technology. The localization trajectory in the global scale is split to a frame-by-frame basis to estimate the relative position of the robot movement. Since the error from UWB does not accumulate over time, the localization failure can be monitored continuously and the drifting errors of the local trajectory are reduced by the UWB positioning. In the experiments, various SLAM approaches are conducted in the real-world scenes for performance comparison. The evaluation has illustrated the robustness of the proposed drift-free visual SLAM by integrating the UWB technology for indoor localization. In the development, we consider the application for automated factory in a scope of about 20 squared meters. The limitation of the current system is mainly the assumption that the environment is relatively spacious without dynamic objects. In the future work, we will evaluate more complex scenes, and consider the applications on high precision manufacturing using mobile robots.

REFERENCES

[1] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1309–1332, Dec. 2016.
[2] C. Debeunne and D. Vivet, "A review of visual-LiDAR fusion based simultaneous localization and mapping," *Sensors*, vol. 20, no. 7, p. 2068, Apr. 2020.
[3] G.-S. Cai, H.-Y. Lin, and S.-F. Kao, "Mobile robot localization using GPS, IMU and visual odometry," in *Proc. Int. Autom. Control Conf. (CACS)*, Nov. 2019, pp. 1–6.
[4] G. Jiang, L. Yin, S. Jin, C. Tian, X. Ma, and Y. Ou, "A simultaneous localization and mapping (SLAM) framework for 2.5 D map building based on low-cost LiDAR and vision fusion," *Appl. Sci.*, vol. 9, no. 10, p. 2105, May 2019.
[5] W. Hess, D. Kohler, H. Rapp, and D. Andor, "Real-time loop closure in 2D LIDAR SLAM," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016, pp. 1271–1278.
[6] F. Liu, J. Liu, Y. Yin, W. Wang, D. Hu, P. Chen, and Q. Niu, "Survey on WiFi-based indoor positioning techniques," *IET Commun.*, vol. 14, no. 9, pp. 1372–1383, Jun. 2020.
[7] H. Xu, M. Wu, P. Li, F. Zhu, and R. Wang, "An RFID indoor positioning algorithm based on support vector regression," *Sensors*, vol. 18, no. 5, p. 1504, May 2018.
[8] F. Mazhar, M. G. Khan, and B. Sällberg, "Precise indoor positioning using UWB: A review of methods, algorithms and implementations," *Wireless Pers. Commun.*, vol. 97, no. 3, pp. 4467–4491, Dec. 2017.
[9] X. Li, J. Wang, and C. Liu, "A Bluetooth/PDR integration algorithm for an indoor positioning system," *Sensors*, vol. 15, no. 10, pp. 24862–24885, 2015.

[10] D. François, K. Jaffres-Runser, A. van den Bossche, and T. Val, "Accurate and platform-agnostic time-of-flight estimation in ultra-wide band," in *Proc. IEEE 27th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2016, pp. 1–7.
[11] Z. Yin, X. Jiang, Z. Yang, N. Zhao, and Y. Chen, "WUB-IP: A high-precision UWB positioning scheme for indoor multiuser applications," *IEEE Syst. J.*, vol. 13, no. 1, pp. 279–288, Mar. 2019.
[12] C. H. Do and H. Y. Lin, "Incorporating neuro-fuzzy with extended Kalman filter for simultaneous localization and mapping," *Int. J. Adv. Robot. Syst.*, vol. 16, no. 5, 2019, Art. no. 1729881419874645.
[13] K. Mohan, S. B. Achalla, and A. Jain, "A comprehensive review of SLAM techniques," *Cyber-Phys., IoT, Auton. Syst. Ind.*, vol. 2021, pp. 341–369, Dec. 2021.
[14] J. Aulinas, Y. Petillot, J. Salvi, and X. Lladó, "The SLAM problem: A survey," *Artif. Intell. Res. Develop.*, pp. 363–371, Oct. 2008.
[15] H.-Y. Lin and J.-L. Hsu, "A sparse visual odometry technique based on pose adjustment with keyframe matching," *IEEE Sensors J.*, vol. 21, no. 10, pp. 11810–11821, May 2021.
[16] M. L. Wang and H. Y. Lin, "An extended-HCT semantic description for visual place recognition," *Int. J. Robot. Res.*, vol. 30, no. 11, pp. 1403–1420, 2011.
[17] C. Yu, Z. Liu, X.-J. Liu, F. Xie, Y. Yang, Q. Wei, and Q. Fei, "DS-SLAM: A semantic visual SLAM towards dynamic environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 1168–1174.
[18] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, Oct. 2017.
[19] T. Pire, T. Fischer, G. Castro, P. De Cristóforis, J. Civera, and J. J. Berlles, "S-PTAM: Stereo parallel tracking and mapping," *Robot. Auto. Syst.*, vol. 93, pp. 27–42, Jul. 2017.
[20] T. Schneider, M. Dymczyk, M. Fehr, K. Egger, S. Lynen, I. Gilitschenski, and R. Siegwart, "MAPLAB: An open framework for research in visual-inertial mapping and localization," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 1418–1425, Jul. 2018.
[21] Y. Lin, F. Gao, T. Qin, W. Gao, T. Liu, W. Wu, Z. Yang, and S. Shen, "Autonomous aerial navigation using monocular visual-inertial fusion," *J. Field Robot.*, vol. 35, no. 1, pp. 23–51, 2018.
[22] 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," *J. Field Robot.*, vol. 36, no. 2, pp. 416–446, 2019.
[23] K. Witrisal, S. Hinteregger, J. Kulmer, E. Leitinger, and P. Meissner, "High-accuracy positioning for indoor applications: RFID, UWB, 5G, and beyond," in *Proc. IEEE Int. Conf. RFID (RFID)*, May 2016, pp. 1–7.
[24] K. Huang, K. He, and X. Du, "A hybrid method to improve the BLE-based indoor positioning in a dense Bluetooth environment," *Sensors*, vol. 19, no. 2, p. 424, Jan. 2019.
[25] C.-Y. Yao and W.-C. Hsia, "An indoor positioning system based on the dual-channel passive RFID technology," *IEEE Sensors J.*, vol. 18, no. 11, pp. 4654–4663, Jun. 2018.
[26] S. Yang, Q. Sun, X. Dong, and J. Yuan, "A novel SLAM method using Wi-Fi signal strength and RGB-D images," in *Proc. IEEE 8th Annu. Int. Conf. CYBER Technol. Autom., Control, Intell. Syst. (CYBER)*, Jul. 2018, pp. 540–545.
[27] T. Kudo and J. Miura, "Utilizing WiFi signals for improving SLAM and person localization," in *Proc. IEEE/SICE Int. Symp. Syst. Integr. (SII)*, Dec. 2017, pp. 487–493.
[28] 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 *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 3495–3502.
[29] Y. Song, M. Guan, W. P. Tay, C. L. Law, and C. Wen, "UWB/LiDAR fusion for cooperative range-only SLAM," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 6568–6574.
[30] C. Wang, H. Zhang, T.-M. Nguyen, and L. Xie, "Ultra-wideband aided fast localization and mapping system," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 1602–1609.
[31] V. Magnago, P. Corbalan, G. P. Picco, L. Palopoli, and D. Fontanelli, "Robot localization via odometry-assisted ultra-wideband ranging with stochastic guarantees," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Nov. 2019, pp. 1607–1613.
[32] T. Moore and D. Stouch, "A generalized extended Kalman filter implementation for the robot operating system," in *Intelligent Autonomous Systems 13*. Cham, Switzerland: Springer, 2016, pp. 335–348.
[33] M. Labbé and F. Michaud, "Appearance-based loop closure detection for online large-scale and long-term operation," *IEEE Trans. Robot.*, vol. 29, no. 3, pp. 734–745, Jun. 2013.
[34] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EP*n*P: An accurate $O(n)$ solution to the PNP problem," *Int. J. Comput. Vis.*, vol. 81, no. 2, p. 155, 2009.
[35] O. R. Chum and J. Matas, "Optimal randomized RANSAC," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 8, pp. 1472–1482, Aug. 2008.
[36] X. Hu, Z. Luo, and W. Jiang, "AGV localization system based on ultra-wideband and vision guidance," *Electronics*, vol. 9, no. 3, p. 448, Mar. 2020.
[37] C. K. Chui and G. Chen, *Kalman Filtering*. Berlin, Germany: Springer, 2017.
[38] J. Wang and E. Olson, "AprilTag 2: Efficient and robust fiducial detection," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 4193–4198.

HUEI-YUNG LIN (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the State University of New York at Stony Brook, Stony Brook, NY, USA. In 2002, he joined the National Chung Cheng University, Taiwan, as an Assistant Professor, and was promoted to a Full Professor, in 2013. He was the Director of the Research Liaison Division, from 2009 to 2013, and the Academic Development Division, from 2012 to 2014, at the Office of Research and Development, National Chung Cheng University. He is currently a Professor at the Department of Computer Science and Information Engineering, National Taipei University of Technology. He has also a joint appointment with the Department of Electrical Engineering, National Chung Cheng University. He is an author of over 180 international conference and journal papers, and five book chapters. He holds 11 U.S. and nine Taiwan invention patents. His research interests include machine learning, computer vision, robotics, and mechatronics. He is a fellow of IET and a Senior Member of Optica. He was a recipient of the Excellent Research Award from the National Chung Cheng University, the Outstanding Academic-Industry Cooperation Award from the Taiwan Association of System and Science and Engineering (TASSE), and the Outstanding Robotics Engineer Award from the Robotics Society of Taiwan (RST). He serves as an organizing committee member and a program committee member for over 50 international conferences.

MING-CHI YEH received the B.S. degree in computer science and information engineering from the National Changhua University of Education, in 2019, and the M.S. degree in electrical engineering from the National Chung Cheng University, Taiwan, in 2021. He is currently with Zyxel Networks Corporation, Hsinchu, as a Computer Engineer. His research interests include autonomous mobile robot, computer vision, machine learning, image processing, and computer graphics.