

# High-Precision Vision Localization System for Autonomous Guided Vehicles in Dusty Industrial Environments

Xingjie Liu | Guolei Wang | Ken Chen

State Key Laboratory of Tribology,  
Tsinghua University, Beijing, China

## Correspondence

Guolei Wang, State Key Laboratory of  
Tribology, Tsinghua University, Beijing  
China.

Email: [wanguolei@mail.tsinghua.edu.cn](mailto:wanguolei@mail.tsinghua.edu.cn)

## Abstract

Ensuring a convenient yet accurate localization solution is an essential problem in wireless-denied industrial scenarios. Therefore, in this study, a vision localization system with light emitting diode (LED) array targets for autonomous guided vehicle (AGV) navigation is proposed. The visible targets are calibrated and the pose can be computed using a camera that views the LED target. A novel data filtering method that integrates the odometer data and inertial measurement unit (IMU) data with vision data is introduced to provide stable and accurate localization. The vision localization system was tested on a 5-m-long AGV and the results demonstrated that the proposed system obtained a static position accuracy at 6 mm, kinematic position accuracy at 10 mm, and angle accuracy at 0.052°, which is more precise than the other methods used in industrial AGV applications.

## Keywords

autonomous guided vehicle, dusty spraying environment, LED array target, vision localization

## 1 | INTRODUCTION

Autonomous guided vehicles (AGVs) are widely used in ports and warehouses during the ongoing Fourth Industrial Revolution (Industry 4.0). AGVs can transfer cargoes to their correct destinations or carry industrial robots to work from various directions. In cases like Arthur et al. (2008) and Hofacker (1993), engineers put a robot on an AGV to finish aircraft manufacturing instead of a fixed robotic system such as a gantry crane or guide rail on the ground. An AGV can avoid structural transformation and has high adaptability in various applications. However, highly accurate AGV localization is required for safe driving and precise robotic operations.

Several studies have investigated high-precision localization methods for AGVs since 1908. A classical method involves setting magnetic or electric cables underground for tracking. AGVs would follow the cable path by electromagnetic induction (Chen et al., 2004) using special sensors. This method is still widely used in modern factories. It can maintain a position accuracy of around 10 mm. However, it damages the original ground and can be easily affected by magnetic and electric fields.

Some studies used a monocular camera (Wu et al., 2019) to track colored stripes and quick response (QR) codes. The positioning error in those studies could be

reduced to 10 mm with high resolution pictures. These fixed path tracking methods are stable for industrial use. However, they limit the flexible mobility of AGVs and cannot be easily modified for complex tasks.

Various path-free navigation methods have been researched to meet increasing localization demands. Schulze and Wullner (2006) set up reflectors and calculated the position with a laser range tracker on a robot. According to Ronzoni et al. (2011), efficient laser tracking can be achieved with a small localization error within an accuracy of 3 mm. However, the cost of building such a system is significantly higher for most applications. Alternatively, lidar can be used to guide mobile robots. However, the present lidar localization methods are mainly used in large scale road environments. Zou compared different indoor lidar localization methods for autonomous vehicles, and the best one achieved a translation error of 10 cm (Zou et al., 2021), which was not accurate enough for indoor robot operations requiring mm-level positioning.

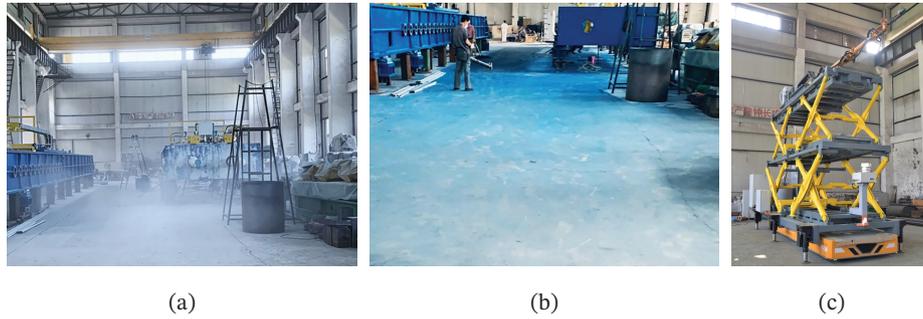
Substantial numbers of wireless localization methods have been introduced in recent years, including wireless fidelity (WiFi; Zhang et al., 2020), ZigBee (Bianchi et al., 2019), radio frequency identification (RFID; Tao et al., 2021), and ultra wide-band (UWB; Shule et al., 2020). These wireless tools are used in large-scale situations, and they provide an accuracy between 1 cm and 10 cm. Methods used by Ahmadiania et al. (2013, 2014) focus on energy consumption and improving the performance with data aggregation. It should be noted that even though wireless localization is convenient currently, it is not included in our project for safety concerns.

With the development of computer vision, there has been a rising trend of utilizing cameras for localization. Vision methods are of two types, namely the *non-markers method* and the *visual markers method*. Tang et al. (2021) built a vision odometer using convolutional neural networks for 2D-3D matching. Red-green-blue-depth (RGBD) cameras can perform better with extra depth information, which is popular in simultaneous localization and mapping (SLAM; Mur-Artal & Tardós, 2017). Karkus et al. (2021) proposed a differentiable SLAM network for indoor robot navigation with a mapping error of 20 cm. However, vision localization with only environmental features can be easily disturbed by image loss, and is not accurate enough. Artificial visual targets attached on a wall or ceiling can help in promoting localization precision (Ortiz-Fernandez et al., 2021) and stability. With well-calibrated targets, relative positioning error can be reduced to 4 mm.

Owing to the various advantages and disadvantages of single-sensor methods, multi-sensor localization methods have also been examined in the past (Li et al., 2019; Song et al., 2017). Nguyen et al. (2021) presented a tightly-coupled camera-IMU-UWB system focusing on drift reduction, and achieved an accuracy of 10 cm. The combination of different data provides a more robust and accurate position.

The Kalman filter (Kalman, 1960) method is a classical linear filtering method for data fusion. Several nonlinear filtering methods have been developed based on this filter. Extended Kalman filters (EKF; Jazwinski, 2007) use analytical approximation to approach nonlinear filtering, and unscented Kalman filters (UKF; Julier & Uhlmann, 1997) select state sampling. EKF and UKF are the most frequently used methods owing to their good performance and generalization ability.

In practical applications, the state covariance matrix might lose positive definiteness due to rounding in computer calculations. Methods presented by Holmes et al. (2008) used the square root for covariance matrix transmission to solve this problem. Li et al. (2020) proposed an improved maximum correntropy cubature Kalman filter to eliminate the measurement outliers in the cooperative localization of autonomous underwater vehicles (AUVs). Bai et al. (2020) utilized mixture variant Gaussian distributions to build a robust Kalman filter for AUVs.



**FIGURE 1** (a) The factory is always dusty during manufacturing; (b) The ground will be stained with paint after spraying; and (c) The mobile spraying system. An 8-m-long robotic arm was set on an 5-m-long AGV. It required high angle accuracy for spraying.

In our laboratory, a mobile spraying system was built for aircraft manufacturing, as shown in Figure 1(c). It included an 8-m-long spraying robot laying on an 5-m-long AGV. The workspace coordinate system was measured previously and the AGV stations were obtained from offline spraying plannings. A small angle error of  $0.2^\circ$  at the robotic base implied a 14 mm position error at the endpoint, which is dangerous for spraying path compensation. Thus, the AGV must arrive at the right station with a high angular accuracy. However, the spraying house is bad for several precise localization methods, as shown in Figure 1(a).

Wireless techniques can be considered in this condition, but they have been banned in our project due to safety concerns. Moreover, localization methods with laser trackers are significantly more expensive. The lidar method or non-marker vision odometer is sensitive to noise and not accurate enough. Therefore, the best alternative in our project was putting artificial markers (such as QR codes) to guide AGVs.

Common visual marks are susceptible to light, and the brightness in industrial environments is not stable enough. Some researchers attached printed targets on the ground (Tang et al., 2020), but they were easily stained with paint, as shown in Figure 1(b).

Factory ceilings are too high to set markers and walls are not allowed to be pasted with markers. Therefore, we made some movable targets (introduced in Section 2.1) with infrared light emitting diode (LED) lamp beads to build a vision localization system.

LED light can efficiently penetrate the paint mist and grinding dust. Therefore, this system can run in spraying environments. The camera positioning result is at low frequency and might get disturbed by obstacles occasionally, so we integrated an odometer and inertial measurement unit (IMU) to improve the publishing rate. Moreover, we optimized the vision results to reduce errors.

The main contributions of the introduced vision localization solution are summarized as follows:

1. A vision localization system using LED light targets was designed for wireless-denied aircraft manufacturing factories. The system is proven to have high adaptability and stability in dusty spraying environments in comparison to other localization methods.
2. A fast and convenient least-squares calibration method between the camera and AGV control center is introduced.
3. The characteristics of vision localization are analyzed geometrically, and its noise is modeled using Gaussian distribution.

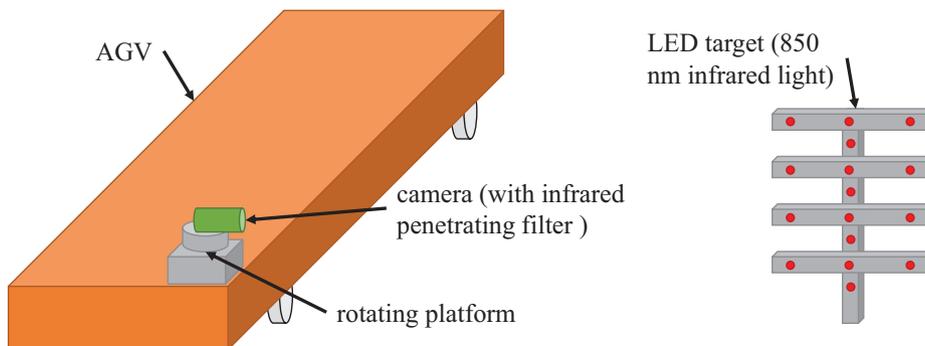
4. A novel multi-sensor fusion algorithm was created to estimate a more robust and precise pose at 10 Hz. The algorithm contains a vision localization pre-optimization technique called VLPO, and a square root central difference Kalman filter (SR-CDKF; Nørgaard et al., 2000). The system obtained position accuracy within 10 mm and angle accuracy within  $0.052^\circ$ , thereby exhibiting superior performance.

The remainder of this paper is organized as follows: The vision localization system and multi-sensor fusion operation are introduced in Section 2; Section 3 discusses the experiments conducted to verify the performance of our localization method; and Section 4 concludes this work.

## 2 | LOCALIZATION SYSTEM

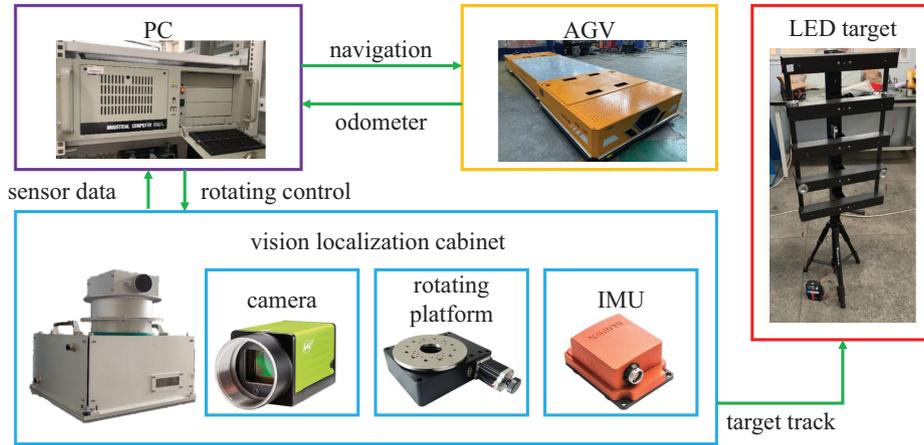
In this section, an overview of the proposed AGV navigation system and its localization algorithm is presented. The AGV navigation system is illustrated in Figure 2. The camera is fixed on a rotating platform that is placed in front of the AGV. The LED target was designed by Chen et al. (2019) from our laboratory. There are specially arranged infrared LED lights for vision positioning. The camera tracks the LED target when the system is running. The valid distance between the target and camera is 3.5~10 m. The pose of the AGV can be obtained using the algorithm detailed in the following subsections.

The concrete structure of the system is shown in Figure 3. We used a personal computer (PC) with 4 Intel Core i7-7700HK CPU at 3.6 GHz, 16 GB of RAM, and a solid state drive of 256 GB for processing the localization algorithm. The AGV used in this study was a four-wheel steering vehicle with a length and width of 5,660 mm and 1,835 mm, respectively. It has a Siemens S2000 programmable logic controller (PLC) to control itself and an ethernet cable to communicate with the navigation PC. The AGV sends four-wheel angles and speeds to the PC at 10 Hz, which is similar to a car odometer. The camera is a complementary metal oxide semiconductor (CMOS) sensor with a resolution of  $2560 \times 2048$ . Image processing to compute the vision localization pose requires 0.2 s. The rotating platform has an angle accuracy of  $0.01^\circ$ , and it communicates with the PC at 10 Hz. Additionally, an IMU operating at 60 Hz provides high precision yaw angle information for AGV in a short time.

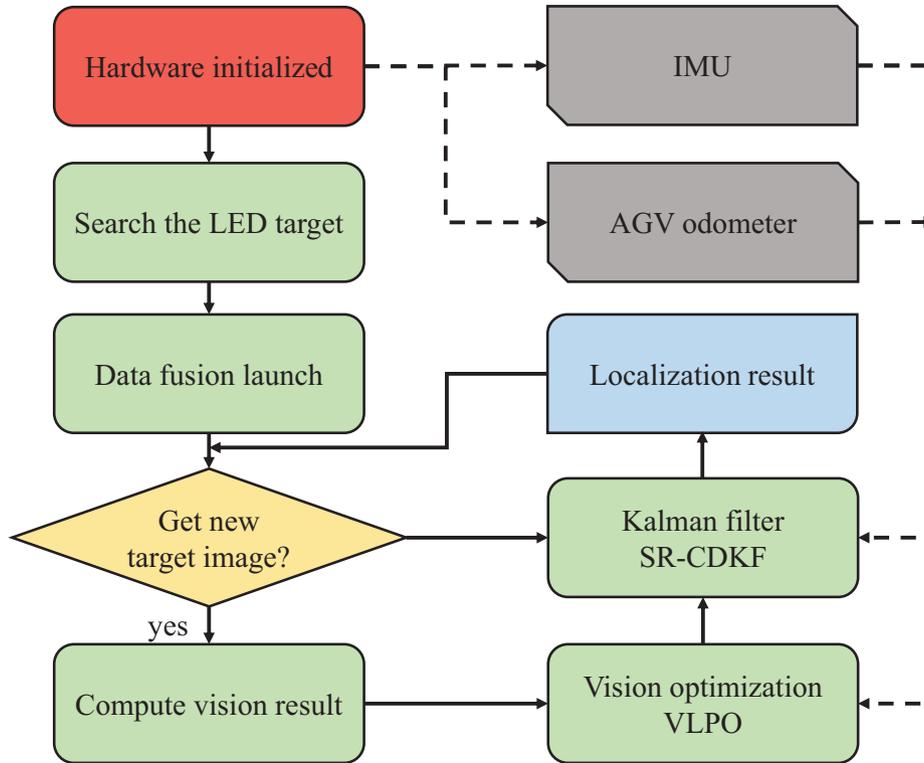


**FIGURE 2** Overview of the experimental AGV navigation system; the length and width of the AGV are 5 and 1.8 m, respectively. The camera located before the AGV and the LED target should be 3.5~10 m away from the camera. An infrared penetrating filter is mounted on the camera lens. Distributed LED lights are placed on the target; they emit infrared light of 850 nm. The camera will track the LED target through rotations. Thus, the pose of the AGV with full view of the LED target can be computed.

The camera, rotating platform, and IMU are put into a cabinet for explosion protection. They are connected with the navigation PC through USB ports. The LED target has a height and width of 560 mm and 320 mm, respectively. We attached the target to a tripod and adjusted it to the same height as the camera. The workings of the system are illustrated in Figure 4.



**FIGURE 3** Structure of localization system: A navigation personal computer (PC) was connected with sensors and AGV to run the positioning system. The rotating platform and camera formed a feedback servo tracking system that maintained the LED target in the middle of the camera view. The camera transferred the LED target image to the PC to compute the vision pose. An IMU was installed in the vision localization cabinet to improve vision results. The AGV then sends the odometer data to the PC and receives control instructions.



**FIGURE 4** Flowchart of the localization system: The camera searches the LED target after system initialization and continues to track the target. When the camera obtains a new valid image, including the full target, a vision computation thread is used to handle the image and optimize results. Then, the vision observation results are sent to the Kalman filter process with the IMU and odometer data. The filtering process estimates the final AGV pose for navigation. If there is no new image in the data fusion loops, the odometer and IMU are used to predict the pose.

When the system is initialized, the navigation PC controls the rotating platform and the camera captures the LED target. The PC computes a priori vision results. Then the localization filter progress launches with the data of the camera, IMU, and odometer. As the AGV moves, the camera tracks the target.

## 2.1 | Vision Positioning Using LED Target

The core of our localization method in the vision positioning system consists of a camera and LED target. The camera is set to a short exposure time of 300  $\mu$ s and low light gain. We installed an infrared penetrating filter on the camera lens. Thus, we could only see the 850 nm LED lights from the camera, as shown in Figure 5. This worked efficiently, and the image stayed almost unchanged during the test for the entire day in the factory. No brightness adjustment was made to the system, which is an advantage in comparison to conventional vision methods.

When the camera obtained a full sight of the LED lights on the target, it detected all 2-dimensional (2D) blobs and matched them with the real 3-dimensional (3D) points. The LED light blobs were recognized by a simple algorithm called simpleBlobDetector from OpenCV (Bradski, 2000) using the following steps:

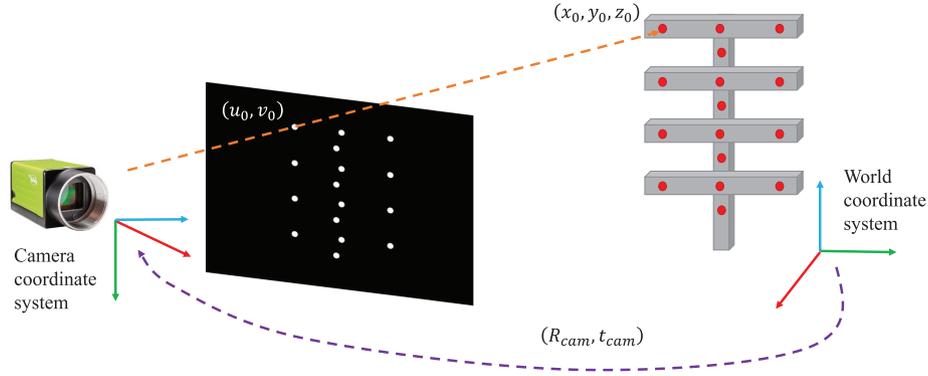
1. The source image was converted to binary images (values from 0 to 255) with several gray-scale thresholds. Here, the neighboring thresholds start from 50 to 140 with an interval of 15.
2. In each binary image, the connected white pixels were grouped together as white blobs.
3. The blob centers were computed and the blobs of all binary images located closer than 4 pixels were merged.
4. The blobs with less than 16 or more than 60 pixels were filtered.

The final centers of blobs were returned, and we sorted them by height in correspondence to the correct LED lights. SimpleBlobDetector is the fastest method in comparison to others such as HoughCirclesDetector. However, the detection time varied from 0.14~0.19 s, which effectively limits the vision publishing rate.

The carbon fiber target in Figure 5 was calibrated using a coordinate measuring machine (CMM). We controlled the CMM probe to touch the 16 LED lights and



**FIGURE 5** LED light target (left) and its image captured by the camera (right); the LED lights fixed on the bracket are non-coplanar. There are five ball probe seats on the target that can hold the laser tracker target ball while the camera is set to dark mode and has an infrared penetrating filter. Thus, we can only see the white LED light blobs.



**FIGURE 6** The camera captures the LED light target and recognizes the white blobs in its image. The blobs are sorted by height to match the real LED lights.  $(u_0, v_0)$  is the 2D point coordinate of the first blob in the image plane.  $(x_0, y_0, z_0)$  is the 3D point coordinate of the first LED light in the world coordinate system defined by the laser tracker. The camera pose in the world frame is the transformation  $(R_{cam}, t_{cam})$ , which can be computed through the ePnP algorithm.

recorded their coordinates  $P_{light}^C$  in the CMM coordinate system. We also measured the surface of the laser tracker target ball attached on the seats in Figure 5. The ball probe seat center coordinates  $P_{ball}^C$  were given through spherical center fitting. A Leica AT901 laser tracker was used in the project to define a world coordinate system for robotic spraying planning. The LED target can be transferred into the world coordinate system using simple operations. We attached the laser tracker ball on the ball probe seats of the LED target and obtained new coordinates  $P_{ball}^w$  in the world coordinate system. Rotation  $R_{cam}$  and translation vector  $t_{cam}$ , such that  $R_{cam}P_{ball}^C + t_{cam} = P_{ball}^w$ , can be estimated by singular value decomposition (SVD; (Sorkine-Hornung & Rabinovich, 2017)). Subsequently, the world coordinates of LED lights can be computed with  $P_{light}^w = R_{cam}P_{light}^C + t_{cam}$ .

Assume that the coordinates of the  $i$ -th 2D LED blob centers computed above are  $(u_i, v_i)$  in the image plane, and the related LED light is  $(x_i^w, y_i^w, z_i^w)$  in the world coordinate system. Equation (1) can be used for 3D-2D point transformation using the perspective projection model  $\Pi$  and the camera intrinsic parameter matrix  $A$ :

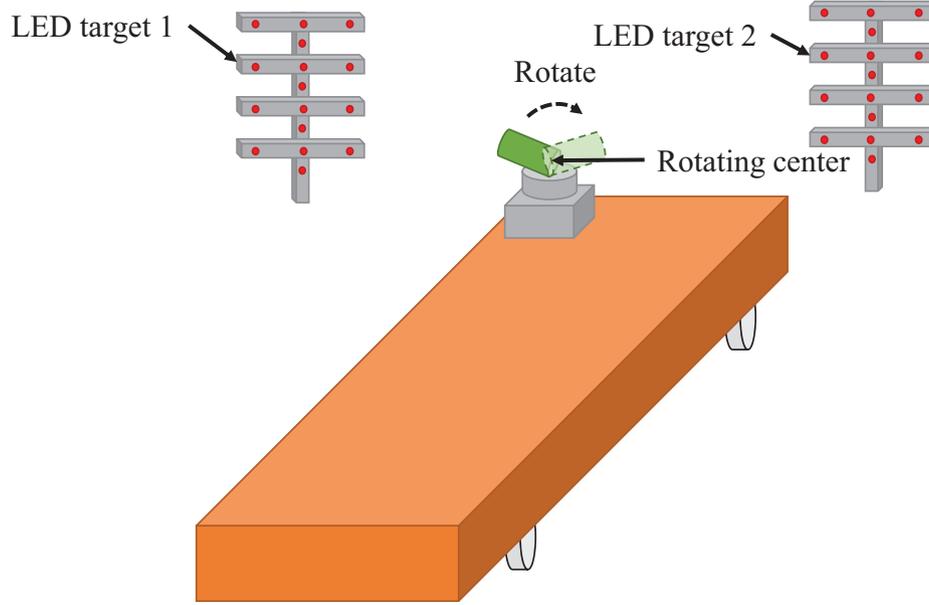
$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = A\Pi T_{cam} \begin{bmatrix} x_i^w \\ y_i^w \\ z_i^w \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_{cam} & t_{cam} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_i^w \\ y_i^w \\ z_i^w \\ 1 \end{bmatrix} \quad (1)$$

where  $T_{cam} = [R_{cam}, t_{cam}; 0, 1]$  is the desired pose of the camera in the world coordinate system in which there are 12 unknown parameters.

We used 16 LED lights for computation (i.e., we solved a 16-linear-equation problem). The camera intrinsic parameters  $(f_x, f_y, c_x, c_y)$  were known and camera distortion was considered in the program. The calculation was performed using an algorithm called Efficient Perspective-n-Point Camera Pose Estimation (ePnP; Lepetit et al., 2009). ePnP treats the control points (16 LED lights) as a weighted sum of four virtual control points. Therefore, the coordinates of the control points become the unknowns of the problem. The final pose of the camera can be obtained using these control points as seen in Figure 6. The ePnP code from OpenCV is efficient and costs less than 0.2 ms.

## 2.2 | Calibration Between the Camera and AGV

The camera has a horizontal sight angle of  $37^\circ$ . We installed it on a rotating platform to widen the measurement field, as shown in Figure 3. The naive vision result calculated by the ePnP method yielded the camera pose matrix  $(R_{cam}, t_{cam})$  in the world frame. It is necessary to transfer the camera pose to the rotating center. In Figure 7, we set two calibrated LED targets and fixed the AGV.



**FIGURE 7** This is an illustration of a calibration experiment that transfers the camera pose to the rotating center. Two calibrated LED targets are fixed on the ground while the AGV remains still in the middle. We rotate the platform to ensure that the camera views each target and computes the vision results.

The rotating center in the world coordinate with:

$$X_{cen}^w = [x_{cen}^w, y_{cen}^w, z_{cen}^w]^T \quad (2)$$

The rotating center in the camera coordinate is expressed as:

$$X_{cen}^{cam} = [x_{cen}^{cam}, y_{cen}^{cam}, z_{cen}^{cam}]^T \quad (3)$$

When the camera obtains full view of a target, the pose matrix  $(R_{cam}, t_{cam})$  can be obtained as:

$$R_{cam} \cdot X_{cen}^w + t_{cam} = X_{cen}^{cam} \quad (4)$$

where  $X_{cen}^{cam}$  is unchanged because the rotating center is fixed in the camera coordinate system. The rotating platform is rotated so that the camera can view each target and get  $(R_{cam1}, t_{cam1})$  and  $(R_{cam2}, t_{cam2})$  with ePnP. Then, we have:

$$X_{cen1}^w = R_{cam1}^{-1} \cdot (X_{cen}^{cam} - t_{cam1}) \quad (5)$$

and

$$X_{cen2}^w = R_{cam2}^{-1} \cdot (X_{cen}^{cam} - t_{cam2}) \quad (6)$$

The rotation center is stable in the world coordinate system. Thus,  $X_{cen1}^w$  should be equal to  $X_{cen2}^w$ . Next, the AGV was put at different places and a series of  $(R_{cam1\_i}, t_{cam1\_i}, R_{cam2\_i}, t_{cam2\_i})$  was recorded. Subsequently, the estimated rotating center  $\hat{X}_{cen}^{cam}$  can be calculated through:

$$X_{cen1\_i}^w = R_{cam1\_i}^{-1} \cdot (X_{cen}^{cam} - t_{cam1\_i}) \quad (7a)$$

$$X_{cen2\_i}^w = R_{cam2\_i}^{-1} \cdot (X_{cen}^{cam} - t_{cam2\_i}) \quad (7b)$$

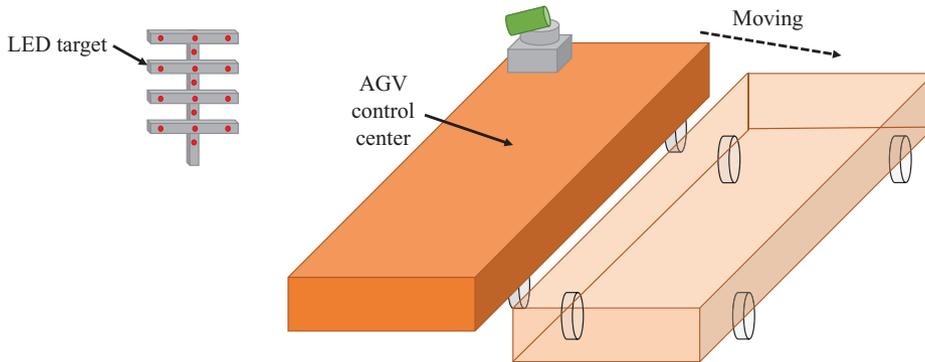
$$\hat{X}_{cen}^{cam} = \arg \min_{X_{cen}^{cam}} \sum_{i=1}^n \|X_{cen1\_i}^w - X_{cen2\_i}^w\| \quad (7c)$$

Real-time  $X_{cen}^w$  can be estimated with the current vision result  $(R_{cam}, t_{cam})$ :

$$X_{cen}^w = R_{cam}^{-1} \cdot (\hat{X}_{cen}^{cam} - t_{cam}) \quad (8)$$

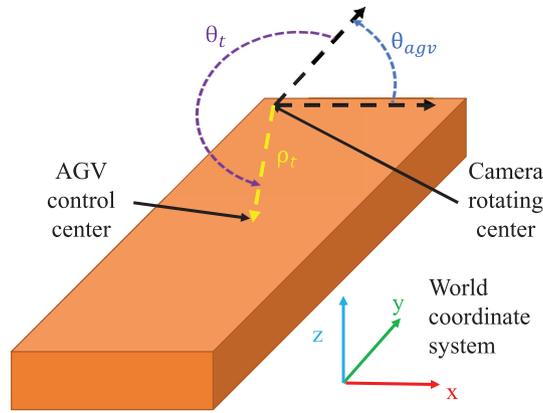
The navigation system focuses on the 2D plane. Thus, only  $x_{cen}^w$  and  $y_{cen}^w$  are useful. The camera angle  $\theta_{cam} = atan2(-y_{cen}^w, -x_{cen}^w)$  minus the rotating platform angle provides the yaw angle  $\theta_{agv}$  of the AGV.

For AGV navigation, the pose at the AGV control center (center of gravity) is required. In Figure 8, the AGV was driven to several destinations. The positions of the camera rotating center  $X_{cen\_i}^w = [x_{cen\_i}^w, y_{cen\_i}^w, z_{cen\_i}^w]^T$  and yaw angle of AGV  $\theta_{agv\_i}$  were recorded. We also used a laser tracker to measure the true AGV control point positions  $X_{agv\_i}^w = [x_{agv\_i}^w, y_{agv\_i}^w]^T$ .



**FIGURE 8** This is a depiction of a calibration experiment that transfers the camera rotating center to the AGV control center. Calibrated LED targets are fixed to the ground and the AGV stays at different stations.

The relative position transformation between the camera rotating center and AGV control center remained unchanged.  $\rho_t$  and  $\theta_t$  denote the position bias and angle bias in the polar coordinate system as shown in Figure 9.



**FIGURE 9** Position transformation between the camera rotating center and AGV control center

We can estimate the optimal bias  $\hat{\rho}_t$  and  $\hat{\theta}_t$  with:

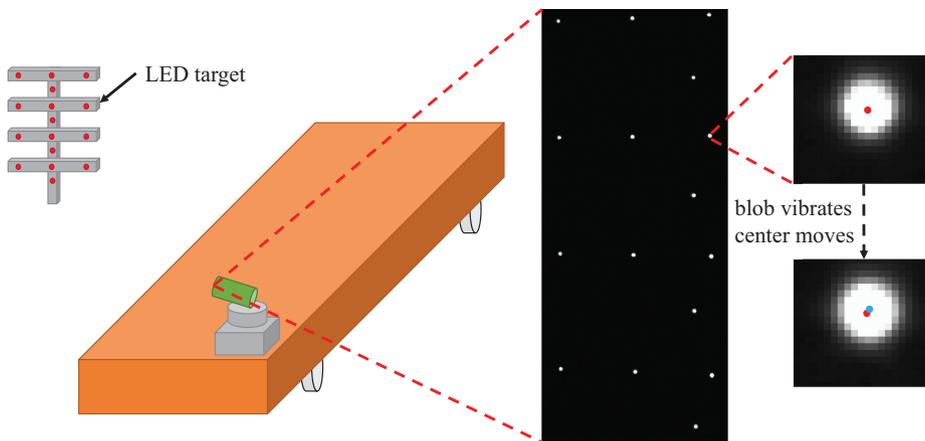
$$x_{bias\_i} = x_{cen\_i}^w + \rho_t \cos(\theta_{agv\_i} + \theta_t) - x_{agv\_i}^w \quad (9a)$$

$$y_{bias\_i} = y_{cen\_i}^w + \rho_t \sin(\theta_{agv\_i} + \theta_t) - y_{agv\_i}^w \quad (9b)$$

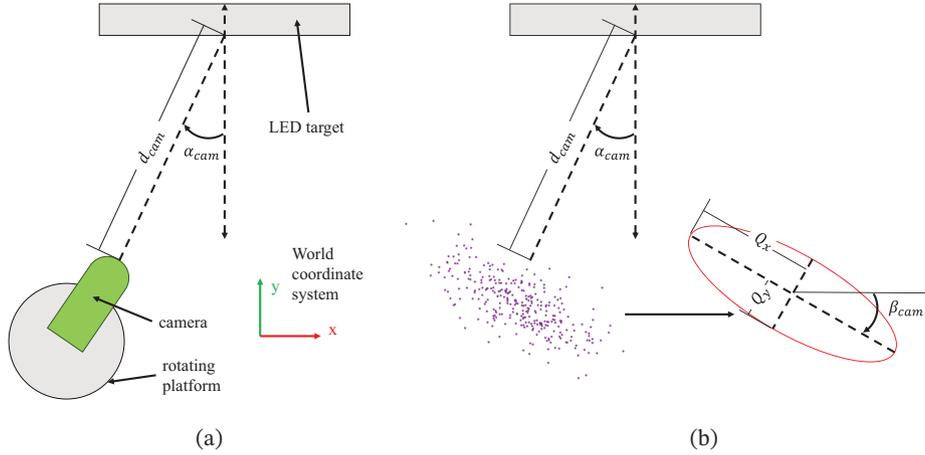
$$(\hat{\rho}_t, \hat{\theta}_t) = \arg \min_{\rho_t, \theta_t} \sum_{i=1}^n \|x_{bias\_i}^2 + y_{bias\_i}^2\| \quad (9c)$$

### 2.3 | Error Distribution of Static Vision Localization Result

The LED blob in the camera image is fitted as a circle to compute the center. The AGV vibrates when the electric and mechanical systems are running. Thus, the white light blob will jitter in camera imaging, as demonstrated in Figure 10. The blob center that recognized from the image will move among adjacent pixels. These vibrations make changes in  $u_i$  and  $v_i$ , as demonstrated in Equation (1). Correspondingly, there was noise in  $(R_{cam}, t_{cam})$ , which in turn produced errors for the vision results. To improve the accuracy of localization, the error distribution



**FIGURE 10** LED blob image when the stationary camera is watching the target; the LED blob has pixel-level vibrations, which generate vision noise at a position and yaw angle of 2 mm and  $0.02^\circ$ , respectively.



**FIGURE 11** (a) Top view of the camera and LED target in which  $d_{cam}$  is the distance between them and  $\alpha_{cam}$  is their intersection angle; (b) Purple dots denote the vision pose results of the camera rotating center. The dots approximately obey the Gaussian distribution and can be simplified as the red ellipse.  $Q_x$  is the semi-major,  $Q_y$  is the semi-minor, and  $\beta_{cam}$  is the rotating angle.

had to be determined. The angle error was approximately  $0.02^\circ$ , and we determined that the angle was precise enough. The position error was approximately 2 mm, and it had to be reduced.

It is difficult to parse out the error directly from the computations used in the localization process. However, we could analyze the image vibrations geometrically. The vertical bias of the blob center in the image affects the height of the vision result. However, the navigation is on the 2D-plane without height so that the vertical vibration can be ignored.

The horizontal bias introduced a vision pose result variation on the navigation plane. In Figure 11(a), the camera was fixed and heading toward the target. We recorded the vision result of the camera rotating center and obtained the purple dots in Figure 11(b).

Similar operations were performed at different places in the LED valid region. We found that the vision results could be plotted as an ellipse as shown in Figure 11(b), which is suitable for a Gaussian distribution approximation. The shape of the ellipse is relative to the distance  $d_{cam}$  and  $\alpha_{cam}$ . The covariance matrix of the vision results  $\Sigma_{vo}$  can be calculated using Equation (10).

$$\Sigma_{vo} = R_{\beta_{cam}} \cdot S_{vo} \cdot R_{\beta_{cam}}^T = R_{\beta_{cam}} \cdot \begin{bmatrix} Q_x & 0 \\ 0 & Q_y \end{bmatrix} \cdot R_{\beta_{cam}}^T \quad (10)$$

$$\text{where } R_{\beta_{cam}} = \begin{bmatrix} \cos\beta_{cam} & -\sin\beta_{cam} \\ \sin\beta_{cam} & \cos\beta_{cam} \end{bmatrix}$$

The angle of ellipse  $\beta_{cam}$  is roughly equal to angle  $\alpha_{cam}$  as shown in Figure 11(b). The numerical model of the ellipse axis fitted by the MATLAB optimization tool is presented in Equation (11). We can obtain  $\alpha_{cam}$  and  $d_{cam}$  with simple math computations to ensure that  $\Sigma_{vo}$  is available for later vision optimization. Although  $\Sigma_{vo}$  is computed in static situations, the error distribution is still helpful in kinematic localization.

$$Q_x = a_x + b_x \cdot e^{d_{cam} - c_x} \cdot \left( \left| 10\cos(\alpha_{cam}) \right| + \left| \sin(\alpha_{cam}) \right| \right) \quad (11a)$$

$$Q_y = a_y + b_y \cdot e^{d_{cam} - c_y} \cdot \left( \left| \cos(\alpha_{cam}) \right| + \left| 10\sin(\alpha_{cam}) \right| \right) \quad (11b)$$

where

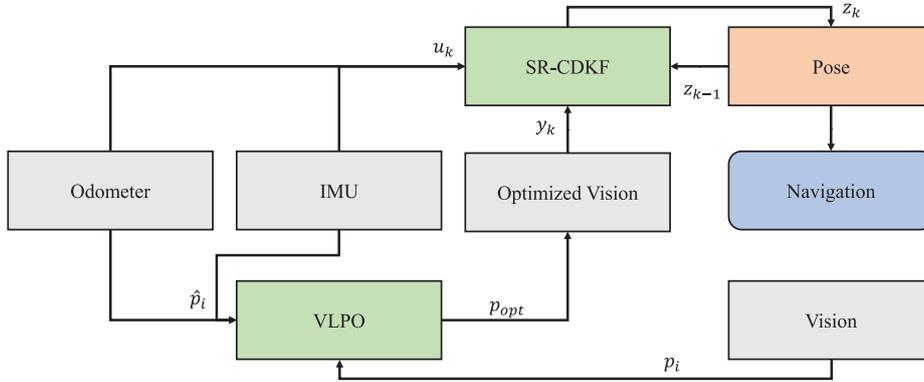
$$a_x = -0.05373, b_x = 18.2, c_x = 12.29 \quad (11c)$$

$$a_y = -0.01638, b_y = 1.095, c_y = 9.598 \quad (11d)$$

## 2.4 | Data Fusion

AGVs in this aircraft painting application require at least 10-Hz pose messages for navigation. The vision localization data were obtained at a low frequency so they are sensitive to environmental vibrations. The camera cannot work if its vision is obstructed. Therefore, it is necessary to use other sensors to improve performance.

The odometer and IMU are reliable for short-time measurements. When the camera data are missing, the odometer and IMU can still predict the pose for a short duration until the target is found again. If the vision element works accurately, the odometer and IMU contribute to the Kalman filter. Filtering can reduce the noise such that the result is maintained in a reasonable range. Additionally, a high-frequency Kalman filter loop can provide a stable localization solution. The fusion process is illustrated in Figure 12.



**FIGURE 12** This is the structure of data fusion. The vision result of the camera rotating center  $p_i = [x_i, y_i, \theta_i]$  is sent to the optimization process, VLPO. The odometer and IMU provide the transformation constraint  $\hat{p} = [\mathcal{L}x_i, \mathcal{L}y_i, \Delta\theta_i]$  to improve  $p_i$  to obtain a more accurate result  $p_{opt}$ . The square root central difference Kalman filter (SR-CDKF; Nørgaard et al., 2000) runs at 10 Hz to estimate the current AGV pose  $z_k$ . The SR-CDKF receives the measurement vector  $y_k$  (AGV control center pose), control input vector  $u_k = [v, w]$ , and pose  $z_{k-1}$  in the last frame. The pose will be sent to navigation block for AGV trajectory tracking control.

### 2.4.1 | Vision Localization Pre-Optimization

The Kalman filter is based on the Markov model. Therefore, estimating the state of the current frame is related only to the last frame and the constraint between the two frames. Hence, other recent earlier frames are wasted.

In this study, the nearest frames and constraints were collected to optimize the sum of the re-projection error, called VLPO. Vision localization provides the camera rotating center pose  $[p_i, \theta_i]^T = [x_i, y_i, \theta_i]^T$  and its covariance  $S_{vo,i}$  mentioned in Section 2.3 for the  $i$ -th frame. The constraint  $\hat{p}_i$  from the odometer and IMU indicates the movement between two poses in the  $i$ -th frame. Because vision localization includes efficient angle measurement, the cost function is built only with the position errors. Equation (12) uses the inverse square root of the covariance matrix to multiply the residual to give the cost function  $F(p)$  of  $k$ -frames:

$$f(p_i) = S_{v_{0,i}}^{-1/2} [R_i^T (p_{i+1} - p_i) - \hat{p}_i] \quad (12a)$$

$$F(p) = [f(p_1), f(p_2), \dots, f(p_k)]^T \quad (12b)$$

where  $R_i = \begin{bmatrix} \cos \theta_i & -\sin \theta_i \\ \sin \theta_i & \cos \theta_i \end{bmatrix}$

We are solving the least squares problem to get the optimal position vector  $p_{opt}$ :

$$p_{opt} = \underset{p}{\operatorname{argmin}} \frac{1}{2} \|F(p)\|^2 \quad (13)$$

The Jacobian matrices  $J(p)$  are given by  $J_i = \frac{\partial}{\partial x_i} F(p)$ . Linearization  $F(p + \Delta p) \approx F(p) + J(p)\Delta p$  is utilized to solve a sequence of approximations. The initial vector  $p$  with prior variable  $\mathcal{C}$  is calculated, and a new least squares problem is obtained, as follows:

$$\Delta p_{opt} = \underset{\Delta p}{\operatorname{argmin}} \frac{1}{2} \|F(p) + J(p)\Delta p\|^2 \quad (14)$$

The problem can be expressed in another way (i.e., solving an equation):

$$H(p)\Delta p = -g(p) \quad (15)$$

where  $H(p) = J(p)^T J(p)$  is the Hessian matrix and  $g(p) = \nabla \frac{1}{2} \|F(p)\|^2 = J(p)^T F(p)$  is the gradient vector.

The problem is sparse when referring to several frames. The most efficient strategy is the decomposition of  $H(p) = R^T(p)R(p)$  with Cholesky factorization, where  $R$  is an upper triangular matrix. The correction is calculated in Equation (16):

$$\Delta \hat{p} = -R^{-1}(p)R^T(p)g(p) \quad (16)$$

$\Delta \hat{p}$  multiplies a scale factor  $\alpha$  to limit the step and  $p_{opt} = p + \mu \Delta \hat{p}$ . Then, the aforementioned processes are iterated until the convergence is realized. The problem was solved using the Ceres solver (Agarwal et al., n.d.). After pre-optimization (i.e., VLPO), a better localization sequence was obtained. The latest optimal frame was transferred to the AGV control center for the following filtering process.

### 2.4.2 | Filtering Loop

The optimized vision localization technique yields 5 Hz which is not enough for AGV navigation with the 10 Hz mentioned above. Moreover, the latest frame is not the current frame due to the computational time cost. The fusion algorithm uses a square root central difference Kalman filter to provide better localization. A central difference Kalman filter (CDKF) requires only one parameter, which is simpler than UKF but as accurate as UKF. Vision localization is treated as an observation.

The odometer and IMU data offer the input message in the state equation following the motion model. The state estimation problem for the system is described in Equation (17). The state variable vector is  $z_k = [x, y, \theta]^T$ . Optimized vision localization directly gives the pose. Therefore, the measurement vector is  $y_k = [x, y, \theta]^T$ .

The input vector  $u_k = [v, w]^T$  consists of linear and angular velocities.  $w_k$  and  $n_k$  are the process and measurement zero mean noise vectors, respectively.  $R_w = \operatorname{diag}\{0.002m, 0.002m, 0.002rad/s\}$  and  $R_n = \operatorname{diag}\{0.002m, 0.002m, 0.002rad/s\}$  are their corresponding covariance matrices.

$$z_k = f(z_{k-1}, u_{k-1}, w_{k-1}) \quad (17a)$$

$$y_k = h(z_k, n_k) \quad (17b)$$

The filter is initialized with a state mean vector  $\hat{z}_0$  and the square root of a covariance:

$$\hat{z}_0 = E[z_0] \quad (18a)$$

$$S_0 = chol\{E[(z_0 - \hat{z}_0)(z_0 - \hat{z}_0)^T]\} \quad (18b)$$

where  $E[\cdot]$  calculates the expectation and  $chol\{\cdot\}$  calculates the Cholesky factorization.

Then, for  $k = 1, 2, \dots$ , the SR-CDKF obtains the sigma-point with:

$$\hat{z}_{k-1}^{a_w} = [\hat{z}_{k-1}; \bar{w}] \quad (19a)$$

$$S_{k-1}^{a_w} = diag\{S_{k-1} \ chol\{R_w\}\} \quad (19b)$$

$$\psi_{k-1} = \left[ \hat{z}_{k-1}^{a_w} \ \hat{z}_{k-1}^{a_w} + hS_{k-1}^{a_w} \ \hat{z}_{k-1}^{a_w} - hS_{k-1}^{a_w} \right] \quad (19c)$$

where the optimal parameter  $h$  is always  $\sqrt{3}$  and  $\bar{w} = [0, 0, 0]^T$ ,  $diag\{\cdot\}$  indicates the diagonal matrix.

The current attitude is predicted according to each sigma-point as  $\psi_{k/k-1} = f(\psi_{k-1}, u_{k-1}, 0)$  and the state mean and square-root covariance are estimated, as shown in Equation (20):

$$\hat{z}_k^- = \sum_{i=0}^{2L} W_i^m \psi_{i,k/k-1} \quad (20a)$$

$$S_{kc1}^- = \sqrt{W^{c_1}} (\psi_{1:L,k/k-1} - \psi_{L+1:2L,k/k-1}) \quad (20b)$$

$$S_{kc2}^- = \sqrt{W^{c_2}} (\psi_{1:L,k/k-1} + \psi_{L+1:2L,k/k-1} - 2\psi_{0,k/k-1}) \quad (20c)$$

$$S_k^- = qr\{(S_{kc1}^- + S_{kc2}^-)\} \quad (20d)$$

where  $L$  is the dimension of  $z_k$ ,  $qr\{\cdot\}$  calculates QR matrix decomposition,  $\psi_{i,k/k-1}$  is the  $i$ -th item of  $\psi_{1k/k-1}$ , and:

$$W_{i=0}^m = (h^2 - L) / h^2, W_{i>0}^m = \frac{1}{2h^2}, W^{c_1} = \frac{1}{4h^2}, W^{c_2} = \frac{h^2 - 1}{4h^2} \quad (21)$$

Next, the sigma-point for measurement update is calculated as:

$$\hat{z}_{k-1}^{a_n} = [\hat{z}_k^- \ \bar{n}] \quad (22a)$$

$$S_{k-1}^{a_n} = diag\{S_k^- \ chol\{R_n\}\} \quad (22b)$$

$$\psi_{k|k-1} = \left[ \hat{z}_{k-1}^{a_n} \ \hat{z}_{k-1}^{a_n} + hS_{k-1}^{a_n} \ \hat{z}_{k-1}^{a_n} - hS_{k-1}^{a_n} \right] \quad (22c)$$

where  $\bar{n} = [0, 0, 0]^T$ .

Then, the optimization predicts the measurements  $Y_{k|k-1} = h(\psi_{k|k-1})$  and innovates the measurement:

$$\hat{y}_k^- = \sum_{i=0}^{2L} W_i^m Y_{i,k|k-1} \quad (23a)$$

$$S_{\bar{y}_k c1} = \sqrt{W^{c1}} \left( Y_{1:L,k|k-1} - Y_{L+1:2L,k|k-1} \right) \quad (23b)$$

$$S_{\bar{y}_k c2} = \sqrt{W^{c2}} \left( Y_{1:L,k|k-1} + Y_{L+1:2L,k|k-1} - 2Y_{0,k|k-1} \right) \quad (23c)$$

$$S_{\bar{y}_k} = qr \left\{ \left( S_{\bar{y}_k c1} + S_{\bar{y}_k c2} \right) \right\} \quad (23d)$$

The Kalman gain matrix  $K_k$  is calculated as follows:

$$P_{z_k y_k} = W^{c1} S_k^- \left[ Y_{1:L,k|k-1} - Y_{L+1:2L,k|k-1} \right]^T \quad (24a)$$

$$K_k = (P_{z_k y_k} / S_{\bar{y}_k}^T) / S_{\bar{y}_k} \quad (24b)$$

Subsequently, the state mean and covariance of the latest frame are updated:

$$\hat{z}_k = \hat{z}_k^- + K_k (y_k - \hat{y}_k^-) \quad (25a)$$

$$S_k = cholupdate \left\{ S_k^-, K_k S_{\bar{y}_k}, -1 \right\} \quad (25b)$$

where  $cholupdate\{\}$  is a Cholesky factor updating method.

During the vision data intervals, the system only executes Equation (19) with  $\hat{z}_k$  to estimate the current  $\psi_{k+1/k}$  as an AGV pose. Thus, the filtering process can estimate  $\hat{z}_k$  at 10 Hz with optimal vision observation  $y_k$  at 5 Hz. When the LED target is lost, the fusion can still operate for 5 s until the error exceeds 10 mm.

### 3 | EXPERIMENT

To evaluate the validity of the vision localization system and performance of the data fusion method, experiments were performed. The world coordinate system was defined using the laser tracker. As shown in Figure 13, we installed two ball probe seats on the AGV to represent the AGV pose. Owing to the equipment limitations, we could only measure one target ball at a time. Consequently, we obtained the ground truth pose only when the AGV stopped. The real-time angle cannot be measured via laser tracker during driving.

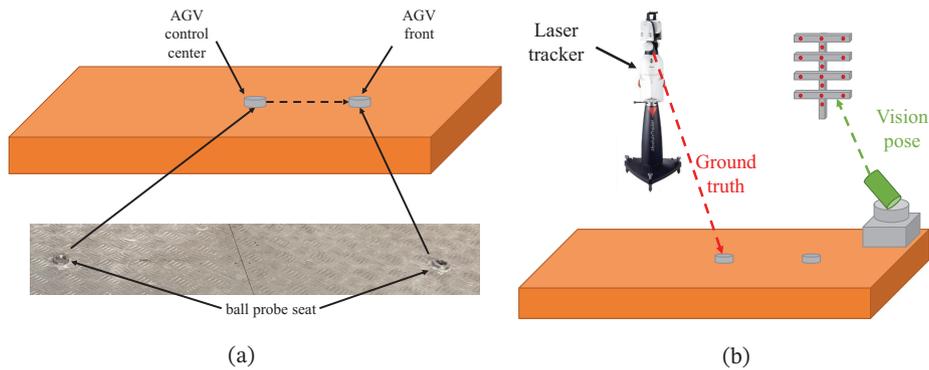
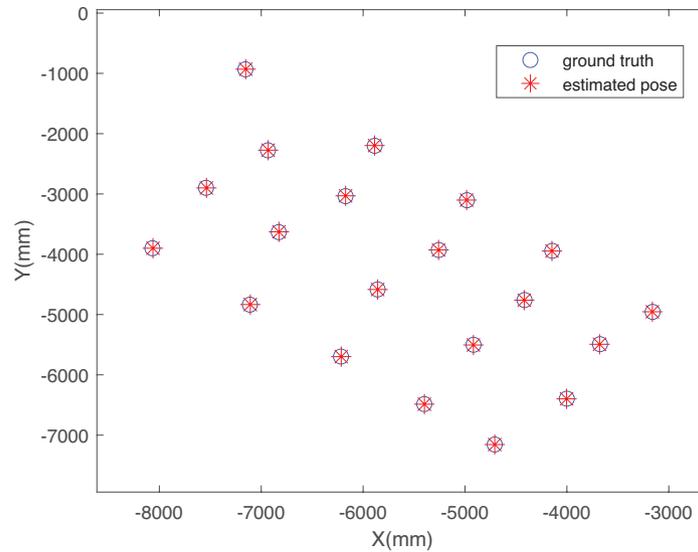


FIGURE 13 (a) These ball probe seats are fixed on the AGV for pose measurement; one is at the AGV control center and the other is at the front. Two seats form a direction that is the same as the AGV yaw angle. (b) depicts the localization experiment. The laser tracker measures the target ball attached on the ball probe seat to give the ground truth.

The AGV was driven to 20 different places to test the vision system. The positioning results were recorded after a while at each station, and the mean value was taken as the estimated AGV pose. The ground truth of the AGV pose was measured using the two ball probe seats by the laser tracker. Figure 14 shows the visual results of the experiment. Table 1 lists the position and angle



**FIGURE 14** This is the position of static vision localization. Red asterisks denote the vision results of the AGV pose. Blue circles denote the ground truth poses measured by the laser tracker.

**TABLE 1**  
Accuracy of static vision localization results

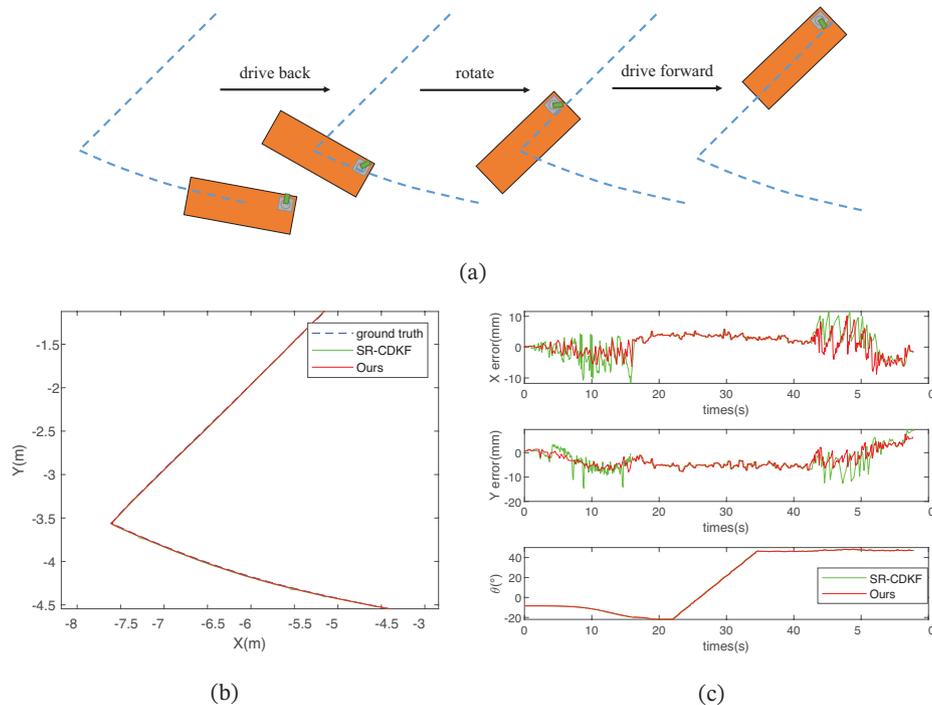
Number	Position(mm)	Angle(°)
1	4.136867620	-0.035228399
2	3.960575702	-0.048955928
3	3.876337586	-0.027618013
4	4.687936237	-0.009338273
5	3.801217524	0.026416594
6	4.366653928	0.013342928
7	3.633090001	0.035763361
8	1.891245422	0.003734939
9	5.450662837	-0.036851006
10	3.202877925	-0.051873036
11	4.801864309	-0.044247884
12	5.756986571	-0.019618353
13	1.190111912	0.012031608
14	4.202759910	0.031727095
15	4.126162090	0.013327649
16	0.797818209	-0.008425105
17	2.461203542	0.026433754
18	2.939329873	0.027528722
19	4.234428989	0.008643642
20	5.452405839	-0.021162262

errors in our vision results. The system reached a precise accuracy of 6 mm and  $0.052^\circ$  for static localization, which is equivalent to the state-of-the-art in AGV positioning.

The kinematic localization experiment was performed as shown in Figure 15(a). The AGV moved along the path in various modes. The laser tracker tracked the target ball attached to the AGV control center to obtain the ground truth of the AGV pose. Our data fusion technique focused on introducing pre-optimization during the vision process instead of improving filtering. Thus, we compared our data fusion method with pure SR-CDKF to observe the proposed method.

As shown in Figure 15(c), the AGV is driving back during 0~20 s, rotating during 20~40 s, and driving forward during the remaining time. The positioning results are almost the same during rotation because the angles of AGV wheels remained unchanged in this mode. The positioning errors are smaller than 8 mm, similar to static situations. The wheels will turn to follow the path when driving backward or forward, which leads to a bigger localization error. The traditional SR-CDKF has a position accuracy of approximately 15 mm. The fusion of pre-optimization and SR-CDKF improves the accuracy to 10 mm. The yaw angle errors cannot be obtained during moving; however, it can be observed that the angle results are quite smooth in both methods.

The computation cost of SR-CDKF is approximately 0.018 s, while that of our additional pre-optimization method is less than 0.006 s. However, the optimization process runs on a separate thread on multi-core CPUs, implying that our method does not increase the computing time of data fusion.



**FIGURE 15** (a) represents the path taken by the AGV in the localization experiment. The AGV starts with a backing process following a curve, and then rotates to change its direction. The remaining path is a simple forward line. (b) is a comparison of the real trajectory of AGV with the SR-CDKF estimated method and ours (VLPO + SR-CDKF). (c) depicts the attitude errors of the AGV, including the  $x$ -coordinate and  $y$ -coordinate. The ground truth of yaw angle  $\theta$  cannot be measured in our experiment; thus, only  $\theta$  is plotted. When AGV is driving back from 0~18 s,  $d_{cam}$  becomes larger and the error becomes larger. When AGV is driving forward from 40~55 s,  $d_{cam}$  becomes smaller and the error becomes smaller.

Some localization systems use the visible light position (VLP; Zhuang et al., 2019) signal strength. These low-cost methods have an average position accuracy of 10 cm for indoor AGVs. Wireless positioning methods are now common in AGV localization, including GPS, cellular communication, Zigbee, RFID, antennae (An et al., 2020), and UWB (He et al., 2020).

These methods are suitable for dusty spraying factories. However, their position accuracy is always over 2 cm, and they are banned in our project for safety reasons. High precision laser sensors can provide positions with 10 mm errors (Cho et al., 2018); however, the building cost of such systems is significant. Visual methods can achieve the best accuracy with general cost for AGV navigation.

The most accurate sample was obtained by Tang et al. (2020), which aimed to guide the aircraft transport AGV. This method placed QR codes on the floor for dual cameras to read and encode. Its distance error and angle were within 10 mm and  $0.15^\circ$ , respectively. QR codes were considered in our system owing to their convenience and performance. However, normal image recognition is considerably affected by the environmental brightness and the printed QR codes stain easily in spraying works.

Boutteau et al. (2020) used distributed LED lights attached to a robot for localization, which is the opposite of our method. However, they obtained a localization accuracy at 30 mm and  $1^\circ$  on a small robot. This may be because their camera was not in dark mode, and they only used four LED lights for computation. The experiment suggested that our vision localization system would have a precise positioning error at 10 mm and  $0.052^\circ$ . To the best of our knowledge, the accuracy obtained by the proposed method is higher than that of other methods.

Our vision localization system includes a camera, rotating platform, IMU, and LED targets, which is affordable for industrial AGVs. An expensive laser tracker was used to transfer the LED target to the global coordinate system for aircraft spraying. A laser tracker has been already used in the project to measure the work-piece, and it will not increase the cost of localization.

Irrespective of the method used, the coordinate system must be transferred by the laser tracker. The positioning method used in this study ensures high accuracy and flexibility for industrial use. LED targets can be conveniently placed on the ground and other devices can be installed on the AGV. Therefore, no structural reforms are required in factories. The camera works in the dark mode and can only see the LED blobs as mentioned in Section 2.1.

The good penetration ability of infrared LED light ensures that the system can work in dusty spraying factories. Consequently, our system exhibits strong adaptability in terrible industrial environments. Practical tests proved that localization can be performed all day during aircraft manufacturing with high accuracy. Table 2 compares the latest well-performed localization methods that can be applied to industrial AGVs. Considering the general cost and high adaptability, the proposed method obtained the best angle accuracy, which plays an important role in large mobile spraying systems.

It should be noted that our method still requires several improvements; specifically, a fast and stable LED blob detection mechanism. We are developing a blob tracking algorithm to improve detection performance. The visual process cannot continue if an obstacle is present between camera and target. This system used an IMU and odometer to ensure that localization is performed for a short duration. We are also investigating an auto-fill algorithm with partial LED blobs that can obtain

**TABLE 2**  
Comparison between AGV localization methods

Methods	Position Accuracy	Angle Accuracy	Cost	Adaptability for Spraying
VLP (Zhuang et al., 2019)	11 cm		Low	General
UWB (An et al., 2020)	20 cm	1.8°	High	High
RFID (Tao et al., 2021)	30 mm		General	General
IMU-UWB (He et al., 2020)	20 mm		High	High
Laser SLAM (Cho et al., 2018)	10 mm	1°	High	General
Visual Marker (Lee et al., 2013)	36 mm	2°	General	Low
Vision with UWB (Hu et al., 2020)	15 mm		High	Low
QR Code, Dual Camera (Tang et al., 2020)	10 mm	0.15°	General	Low
LED lights on robot (Boutteau et al., 2020)	30 mm	1°	General	High
Ours	10 mm	0.052°	General	High

similarly accurate results. Additionally, we plan to install two cameras in the front and back of the AGV. The fusion of double vision localization is another important topic for future work.

## 4 | CONCLUSION

In this paper, we presented a highly accurate vision localization system, which was implemented for industrial applications. The system obtained the pose by processing the LED target image captured by a camera installed on a rotating platform. To provide high frequency and stable positioning, we introduced a Kalman filter loop that integrated the IMU and odometer data.

Our method can provide an affordable localization method for AGVs in wireless-banned industrial environments. The proposed system could achieve a position and angle accuracy of 10 mm and 0.052°, respectively, which denote state-of-the-art results in AGV navigation among the known methods. The system was tested in a real spraying experiment, and it satisfied the localization requirements at high precision and adaptability. We will continue to improve the accuracy and efficiency of the proposed system in future work.

## ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China under Grant No. 51975308 and No. 61403226, and the Ministry of Industry and Information Technology of China under Grant No. MJ-2018-G-54. The authors would like to thank Wiley for their English language editing service.

## CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

## REFERENCES

- Agarwal, S., Mierle, K., & Others. (n.d.). *Ceres solver*. <http://ceres-solver.org>
- Ahmadinia, M., Alinejad-Rokny, H., & Ahangarikiasari, H. (2014). Data aggregation in wireless sensor networks based on environmental similarity: A learning automata approach. *Journal of Networks*, 9(10), 2567. <https://doi.org/10.4304/jnw.9.10.2567-2573>
- Ahmadinia, M., Meybodi, M. R., Esnaashari, M., & Alinejad-Rokny, H. (2013). Energy-efficient and multi-stage clustering algorithm in wireless sensor networks using cellular learning automata. *IETE Journal of Research*, 59(6), 774–782. <https://doi.org/10.4103/0377-2063.126958>
- An, X., Zhao, S., Cui, X., Shi, Q., & Lu, M. (2020). Distributed multi-antenna positioning for automatic-guided vehicle. *Sensors*, 20(4), 1155. <https://doi.org/10.3390/s20041155>
- Arthur, J., Bowman, R., & Straw, R. (2008). *Robotic laser coating removal system* (Tech. Rep.). Air Force Research Lab Wright-Patterson AFB OH. <https://apps.dtic.mil/sti/citations/ADA608206>
- Bai, M., Huang, Y., Chen, B., Yang, L., & Zhang, Y. (2020). A novel mixture distributions-based robust Kalman filter for cooperative localization. *IEEE Sensors Journal*, 20(24), 14994–15006. <https://doi.org/10.1109/JSEN.2020.3012153>
- Bianchi, V., Ciampolini, P., & De Munari, I. (2019). RSSI-based indoor localization and identification for ZigBee wireless sensor networks in smart homes. *IEEE Transactions on Instrumentation and Measurement*, 68(2), 566–575. <https://doi.org/10.1109/TIM.2018.2851675>
- Boutteau, R., Rossi, R., Qin, L., Merriaux, P., & Savatier, X. (2020). A vision-based system for robot localization in large industrial environments. *Journal of Intelligent & Robotic Systems*, 99(2), 359–370. <https://doi.org/10.1007/s10846-019-01114-x>
- Bradski, G. (2000). The OpenCV library. *Dr. Dobb's Journal of Software Tools*.
- Chen, C., Wang, B., & Ye, Q.-T. (2004). Application of automated guided vehicle (AGV) based on inductive guidance for newsprint rolls transportation system. *Journal of Donghua University*, 21(2), 88–92. [www.doi.org/10.3969/j.issn.1672-5220.2004.02.017](http://www.doi.org/10.3969/j.issn.1672-5220.2004.02.017)
- Chen, Z., Wang, G., & Hua, X. (2019). High-precise monocular positioning with infrared LED visual target. *2019 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, Irkutsk, Russia. <https://doi.org/10.1109/RCAR47638.2019.9044142>
- Cho, H., Kim, E. K., & Kim, S. (2018). Indoor SLAM application using geometric and ICP matching methods based on line features. *Robotics and Autonomous Systems*, 100, 206–224. <https://doi.org/10.1016/j.robot.2017.11.011>
- He, C., Tang, C., & Yu, C. (2020). A federated derivative cubature Kalman filter for IMU-UWB indoor positioning. *Sensors*, 20(12), 3514. <https://doi.org/10.3390/s20123514>
- Hofacker, S. A. (1993). Large aircraft robotic paint stripping (LARPS) system. *SAE Transactions*, 102, 11–21. <http://www.jstor.org/stable/44739952>
- Holmes, S., Klein, G., & Murray, D. W. (2008). A square root unscented Kalman filter for visual monoSLAM. *2008 IEEE International Conference on Robotics and Automation*, Pasadena, CA. <https://doi.org/10.1109/ROBOT.2008.4543780>
- Hu, X., Luo, Z., & Jiang, W. (2020). AGV localization system based on ultra-wideband and vision guidance. *Electronics*, 9(3), 448. <https://doi.org/10.3390/electronics9030448>
- Jazwinski, A. (2007). *Stochastic processes and filtering theory*. Dover Publications.
- Julier, S. J., & Uhlmann, J. K. (1997). New extension of the Kalman filter to nonlinear systems. In I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition VI* (Vol. 3068, pp. 182–193). <https://doi.org/10.1117/12.280797>
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1), 35–45. <https://doi.org/10.1115/1.3662552>
- Karkus, P., Cai, S., & Hsu, D. (2021). Differentiable SLAM-net: Learning particle SLAM for visual navigation. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2815–2825.
- Lee, J., Hyun, C.-H., & Park, M. (2013). A vision-based automated guided vehicle system with marker recognition for indoor use. *Sensors*, 13(8), 10052–10073. <https://doi.org/10.3390/s130810052>
- Lepetit, V., Moreno-Noguer, F., & Fua, P. (2009). EPnP: An accurate  $O(n)$  solution to the PnP problem. *International Journal of Computer Vision*, 81. <https://doi.org/10.1007/s11263-008-0152-6>
- Li, B., Hao, Z., & Dang, X. (2019). An indoor location algorithm based on Kalman filter fusion of ultra-wide band and inertial measurement unit. *AIP Advances*, 9(8). <https://doi.org/10.1063/1.5117341>
- Li, S., Xu, B., Wang, L., & Razzaqi, A. A. (2020). Improved maximum correntropy cubature Kalman filter for cooperative localization. *IEEE Sensors Journal*, 20(22), 13585–13595. <https://doi.org/10.1109/JSEN.2020.3006026>
- Mur-Artal, R., & Tardós, J. D. (2017). ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. *IEEE Transactions on Robotics*, 33(5), 1255–1262. <https://doi.org/10.1109/TRO.2017.2705103>

- Nguyen, T. H., Nguyen, T.-M., & Xie, L. (2021). Range-focused fusion of camera-IMU-UWB for accurate and drift-reduced localization. *IEEE Robotics and Automation Letters*, 6(2), 1678–1685. <https://doi.org/10.1109/LRA.2021.3057838>
- Nørgaard, M., Poulsen, N. K., & Ravn, O. (2000). New developments in state estimation for nonlinear systems. *Automatica*, 36(11), 1627–1638. [https://doi.org/10.1016/S0005-1098\(00\)00089-3](https://doi.org/10.1016/S0005-1098(00)00089-3)
- Ortiz-Fernandez, L. E., Cabrera-Avila, E. V., da Silva, B. M., & Gonçalves, L. M. (2021). Smart artificial markers for accurate visual mapping and localization. *Sensors*, 21(2), 625. <https://doi.org/10.3390/s21020625>
- Ronzoni, D., Olmi, R., Secchi, C., & Fantuzzi, C. (2011). AGV global localization using indistinguishable artificial landmarks. *2011 IEEE International Conference on Robotics and Automation*, Shanghai, China. <https://doi.org/10.1109/ICRA.2011.5979759>
- Schulze, L., & Wullner, A. (2006). The approach of automated guided vehicle systems. *2006 IEEE International Conference on Service Operations and Logistics, and Informatics*, Shanghai, China. <https://doi.org/10.1109/SOLI.2006.328941>
- Shule, W., Almansa, C. M., Queralta, J. P., Zou, Z., & Westerlund, T. (2020). UWB-based localization for multi-UAV systems and collaborative heterogeneous multi-robot systems. *Procedia Computer Science*, 175, 357–364. <https://doi.org/10.1016/j.procs.2020.07.051>
- Song, Y., Nuske, S., & Scherer, S. (2017). A multi-sensor fusion MAV state estimation from long-range stereo, IMU, GPS and barometric sensors. *Sensors*, 17(1), 11. <https://doi.org/10.3390/s17010011>
- Sorkine-Hornung, O., & Rabinovich, M. (2017). Least squares rigid motion using SVD.
- Tang, J., Zhu, W., & Bi, Y. (2020). A computer vision-based navigation and localization method for station-moving aircraft transport platform with dual cameras. *Sensors*, 20(1), 279. <https://doi.org/10.3390/s20010279>
- Tang, S., Tang, C., Huang, R., Zhu, S., & Tan, P. (2021). Learning camera localization via dense scene matching. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1831–1841.
- Tao, B., Wu, H., Gong, Z., Yin, Z., & Ding, H. (2021). An RFID-based mobile robot localization method combining phase difference and readability. *IEEE Transactions on Automation Science and Engineering*, 18(3), 1406–1416. <https://doi.org/10.1109/TASE.2020.3006724>
- Wu, X., Sun, C., Zou, T., Xiao, H., Wang, L., & Zhai, J. (2019). Intelligent path recognition against image noises for vision guidance of automated guided vehicles in a complex workspace. *Applied Sciences*, 9(19), 4108. <https://doi.org/10.3390/app9194108>
- Zhang, L., Chen, Z., Cui, W., Li, B., Chen, C., Cao, Z., & Gao, K. (2020). WiFi-based indoor robot positioning using deep fuzzy forests. *IEEE Internet of Things Journal*, 7(11), 10773–10781. <https://doi.org/10.1109/JIOT.2020.2986685>
- Zhuang, Y., Wang, Q., Shi, M., Cao, P., Qi, L., & Yang, J. (2019). Low-power centimeter-level localization for indoor mobile robots based on ensemble Kalman smoother using received signal strength. *IEEE Internet of Things Journal*, 6(4), 6513–6522. <https://doi.org/10.1109/JIOT.2019.2907707>
- Zou, Q., Sun, Q., Chen, L., Nie, B., & Li, Q. (2021). A comparative analysis of lidar SLAM-based indoor navigation for autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 1–15. <https://doi.org/10.1109/TITS.2021.3063477>

**How to cite this article:** Liu, X., Wang, G., & Chen, K. (2022) High-precision vision localization system for autonomous guided vehicles in dusty industrial environments. *NAVIGATION*, 69(1). <https://doi.org/10.33012/navi.502>