



# IRAL: Robust and versatile UAV localization using infrared vision and altitude sensor fusion

Yixian Li <sup>a</sup>, Qiang Wang <sup>a,\*</sup>, Zhonghu Hao <sup>a</sup>, Shengrong Hu <sup>a</sup>, Jiaxing Wu <sup>a</sup>, Linkang Dong <sup>b</sup>

<sup>a</sup> School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100081, China

<sup>b</sup> School of Computer Science & Technology, Beijing Institute of Technology, Beijing 100081, China

## ARTICLE INFO

### Keywords:

Localization  
Unmanned aerial vehicle  
Cooperative method  
Infrared vision  
Sensor fusion

## ABSTRACT

Supplementing or replacing the Global Satellite Navigation System (GNSS) for robust UAV localization remains a challenge. In this work, we propose an infrared vision and altitude sensor fusion method called IRAL, which mainly includes on-board near-infrared (NIR) beacons, off-board narrowband-pass vision sensors with the same wavelength as the beacon and strapdown high precision altitude sensors. The beacon with a high signal-to-noise ratio as a cooperative target provides robust features, thereby facilitating beacon recognition through the designed gradient-based sequential frame template matching (GSFTM) algorithm. The proposed method measures the altitude difference between the UAV and the vision sensor through the altitude sensor to accomplish depth estimation. After obtaining the beacon's pixel coordinates and depth, combined with the intrinsics and extrinsics of the vision sensor, the observation equation can be set up to solve the UAV's spatial position. Real-world experiments under various scenarios demonstrate that the proposed method stably achieves high accuracy.

## 1. Introduction

In recent years, unmanned aerial vehicles (UAVs) have had extensive applications in various scenarios due to their low cost and flexible use, such as border security [1], search and rescue (SAR) [2], fire-fighting [3], precision agriculture [4], mapping [5,6], structural health monitoring (SHM) [7,8], etc. In practical scenarios, UAVs generally require accurate and reliable localization to accomplish their missions effectively. At present, UAV localization still mainly relies on the Global Satellite Navigation System (GNSS). However, GNSS has the fatal drawback of being highly susceptible to dropout, jamming, interference and is extremely difficult to cover closed environments including indoors and urban canyons [9,10]. Therefore, it is critical to achieve reliable and accurate UAV localization in GNSS-denied environments.

Visual localization has become one of the main methods to solve this problem due to the advantages of vision sensors such as resistance to electromagnetic interference, low weight and small size [11]. It relies on the UAV's on-board vision sensor to capture feature points in the surrounding environment during flight, which is then combined with the sensor's intrinsics to calculate its position relative to the feature points. Since 2D images do not contain depth information, i.e., the distance of pixels relative to the sensor, it is not possible to solve the position based on a single frame, but requires the aid of other measures to estimate depth for localization. Thus, visual localization involves two

key issues: feature point acquisition and depth estimation. The former is mainly related to the acquisition of environmental features, and the latter depends largely on the type of sensor adopted. The accuracy and stability of these two aspects basically dictate the performance of a visual localization method.

Depending on whether the environmental features are acquired in advance, visual localization can be categorized as map-based or map-less. Map-based localization requires a large amount of data to be collected and processed in advance, leading to increased implementation costs as well as inconvenience [12]. Further, since the map is a priori, it lacks adaptability to environmental changes [13]. Map-less localization, such as simultaneous localization and map building (SLAM), requires real-time environmental perception and data processing. It is not only sensitive to the number and recognizability of feature points in the environment, but also places requirements on the already limited on-board computational resources [14,15]. It suffers from the problem in terms of accuracy degradation or even failure due to the lack of feature points that can be captured by the vision sensor [16,17]. In addition, general visual localization is based on visible light sensors, which depend on ideal illumination and thus are difficult to function in extreme illumination environments such as night [18].

In order to improve the robustness of feature point acquisition in visual localization, optical beacon-based cooperative methods are

\* Correspondence to: Beijing Institute of Technology, No. 5 Zhongguancun South Street, Haidian District, Beijing 100081, China.  
E-mail address: [wang\\_qiang@bit.edu.cn](mailto:wang_qiang@bit.edu.cn) (Q. Wang).

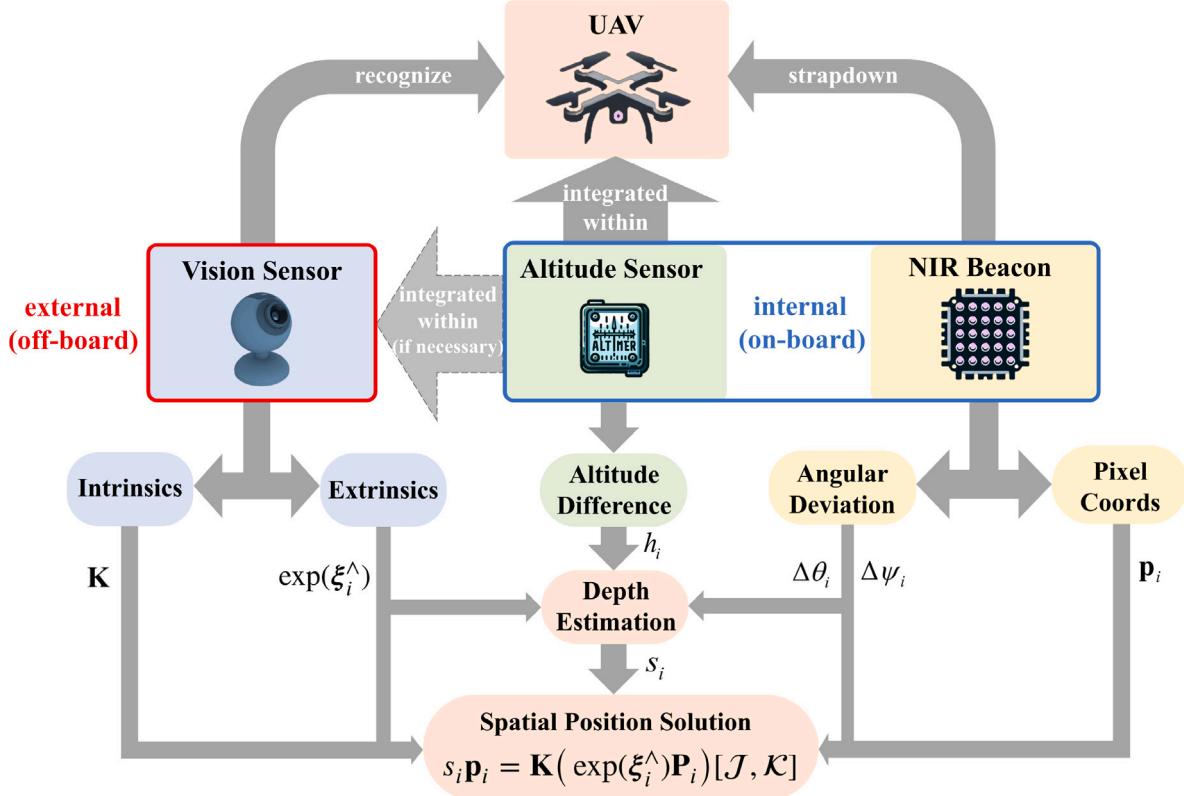


Fig. 1. Schematic of the proposed IRAL.

being increasingly deployed. The idea is to use patterns or optical markers with certain prior features as beacons, and then capture them by the matching sensor or algorithm. In this way the beacons form cooperative targets with respect to the visual sensors, which can be recognized as robust feature points during localization. Depending on whether the vision sensor is on-board or off-board, it can be categorized into internal methods and external methods. Internal methods require prior deployment of a sufficient number of off-board beacons in the environment expected to be covered by the UAV during its flight, which largely limits the range available for localization [19–22]. Meanwhile, by analogy with map-based methods, internal localization is comparable to manually constructing a prior map by adding feature points to the environment, thus also suffering from high implementation costs and inconvenience. External methods can be regarded as the deployment of on-board beacons that turn UAVs into cooperative targets for recognition by off-board vision sensors, typically optical motion capture systems. Some commercial products can achieve millimeter-level or even higher accuracy in localization. However, they require multiple expensive, specialized cameras to be set up in advance on a defined site, with a limited coverage area and high cost of use [23,24]. Some works with more flexible use have also been proposed in recent years, whereas they suffer from a lack of robustness [25,26] or still rely on high-cost sensors [25,27]. In addition, both internal and external methods have the problems of being interfered with by visible light in the environment and sensitivity to light conditions [28], as long as the beacons used operate in the visible light band.

As mentioned above, in addition to feature point capture, depth estimation is also a key component necessary for visual localization. For this aspect, the type of vision sensor used plays a major role. Since single-frame images do not contain depth information, visual localization using a monocular RGB camera needs to rely on the movement of static feature points in time-series frame images to solve for depth. To obtain depth with a simple low-cost monocular RGB vision sensor, other factors must be added. It is currently popular to

obtain depth based on fusion with LiDAR and monocular RGB vision sensors, whereas LiDAR-vision sensor fusion often requires complex calibrations before use and has a high implementation cost [29]. Other types of vision sensors such as binocular or RGB-D cameras can acquire depth directly from the image. However, the former lacks robustness to illumination changes and motion blur [30]. The latter has lower resolution and a shorter effective range [31].

In this work, we propose a novel external cooperative method using InfraRed vision and ALtitude sensor fusion called IRAL for robust and versatile UAV localization in GNSS-denied environments, whose schematic is shown in Fig. 1. Wavelength-specific near-infrared (NIR) integrated LEDs as strapdown on-board beacons provide robust feature points for UAVs. For the off-board narrowband-pass vision sensor of the same wavelength as the beacon, the UAV becomes a cooperative target for a high signal-to-noise ratio. In this context, we design a gradient-based sequential frame template matching (GSFTM) algorithm for robust beacon recognition to obtain pixel coordinates with low computational resource requirements. For general external cooperative methods, multiple visual sensors are required for depth estimation. In contrast, the proposed IRAL relies on a single low-cost monocular RGB vision sensor fused with altitude sensors to achieve it. Specifically, the on-board strapdown high-precision altitude sensor measures the altitude of the UAV above the ground. Since the off-board vision sensor remains static during the localization process, its altitude above the ground is easily measured manually or also by an altitude sensor if necessary. In this way, the altitude difference between the UAV and the vision sensor is obtained, which is then combined with the angular deviation of the recognized beacon relative to the sensor optical center and the known extrinsics of the vision sensor to calculate the depth accurately and efficiently. Synthesizing the above, the pixel coordinates of the recognized beacon, combined with the estimated depth and the intrinsics & extrinsics of the vision sensor, are used to set up an observation equation to solve the spatial position of the UAV. In addition, since

NIR light is independent of the visible spectrum and the narrowband-pass ensures a high signal-to-noise ratio of the beacon, the proposed IRAL is immune to any visible light interference in the environment during the localization process. Meanwhile, it can still perform well under extreme light conditions due to its complete insensitivity to illumination.

In general, this work offers several key contributions, including:

1. A novel external cooperative visual localization model based on on-board NIR beacons, off-board narrow-band pass monocular RGB visual sensors, and strapdown altitude sensors is established.
2. The proposed method constitutes a solution to the two key issues of feature point acquisition and depth estimation for visual localization in a simple and efficient way with low-cost components.
3. A GSFTM algorithm with low computational resource requirements for efficient cooperative beacon recognition is designed.
4. The proposed method achieves centimeter-level accuracy and is insensitive to both environmental features and illumination, thus providing a robust and versatile solution for UAV localization in GNSS-denied environments.

The paper is organized as follows: Section 2 gives a brief introduction to the related work. Section 3 details the methodology of this work. Section 4 describes the design and recognition of the NIR beacon. Section 5 includes the performance of the proposed method in the experiments with the corresponding discussion. Section 6 concludes the paper.

## 2. Related work

This section presents related work on optical beacon-based internal and external cooperative localization methods for unmanned vehicles.

### 2.1. Internal methods

Most existing methods fall into this category. First introduced by [32], a structured beacon inspired by the barber-pole was applied to estimate the relative position and heading of two collaborative master-slave mobile robots. This work can essentially be seen as deploying a mobile beacon in the environment while using an on-board vision sensor to recognize it and solve for relative position. Some recent work [25,26] has drawn on this idea as well. However, this kind of method requires additional robots resulting in higher implementation costs and lacks robustness. Thus in contrast, the deployment of multiple fixed off-board beacons is more widely adopted. In [20–22], multiple LEDs are installed in a certain distribution on the ceiling for indoor positioning. Some other work uses triangulation algorithms to achieve beacon-based localization [33–35]. Whereas these methods require prior installation of the beacons at predefined geometric positions and the localization range is extremely limited. Deploying a number of beacons with specific features and recognizing them based on visible light vision sensors is a feasible solution as well [28,36]. For example, the ArUco marker-based localization is introduced in [36]. The ArUco marker is a square patch with a unique black and white barcode. It is applied by deploying a sufficient number of markers in advance over the desired flight area and recording their prior positions. Multiple markers are recognized by the on-board vision sensor and then the relative position of the UAV can be solved. While like ordinary SLAM, it is unable to overcome the sensitivity to illumination, which leads to failure in extreme light conditions. In addition, the pre-deployment of a large number of beacons leads to inconvenience and limits the localization range. [37,38] focuses on mounting infrared beacons of specific wavelengths on roadway landmarks such as street lights, traffic signals, etc., while placing narrowband-pass vision sensors of the same

wavelengths on vehicles to recognize the beacons for self-localization. This method can be used to facilitate autonomous driving, but also requires a large number of beacons to be pre-placed and does not apply to UAVs. In this work, beacon recognition mainly relies on the high signal-to-noise ratio from NIR narrowband-pass rather than relying on specific geometric features. In this way, not only robustness is ensured, but also failure due to visible light interference as well as insufficient environmental illumination can be avoided. Moreover, the proposed method does not require prior deployment of multiple beacons in the environment. This effectively increases the ease of implementation.

### 2.2. External methods

Up to now, some mature commercial motion capture systems are based on this idea, such as Optitrack [23], Vicon [24] and FZMotion. It has more applications than just unmanned vehicles, so it is generally referred to as marker-based optical tracking. These systems can achieve millimeter-level or even higher accuracy in localization. Nevertheless, they necessitate the deployment of numerous costly, specialized cameras in advance at designated locations, resulting in restricted coverage and a high operational expense. [39] proposes a low-cost method, whereas it is only applicable to estimating the range between unmanned vehicles in a swarm. By deploying stroboscopic infrared LEDs as coded beacons on the robot, [40] can achieve high-precision localization in indoor environments. However, at least six on-board beacons are required for this work and they must be captured by the vision sensor at the same time. Therefore, it only works at very close range and has limited robustness. [41] implements the use of Wii infrared vision sensors to recognize on-board infrared beacons for target localization. While it only verifies the viability of this method from the perspective of signal-to-noise ratio without establishing a complete localization model and applying it to unmanned vehicles. In contrast, the proposed method achieves high accuracy localization in both indoor and outdoor environments with cost-effectiveness. Moreover, it works at large ranges between the vision sensor and the on-board beacon without degradation. In particular, the proposed method includes a low-cost camera only and eliminates the requirement for expensive sensors like dynamic vision sensors (DVS) as compared to [27].

## 3. Methodology

### 3.1. Vision sensor intrinsics

In this work, the off-board vision sensor consists of a narrowband filter and a monocular RGB camera. We adopt the pin-hole model to describe the off-board vision sensor's intrinsics. The non-singular intrinsics matrix, which contains the internal vision sensor parameters, is denoted as follows

$$\mathbf{K} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}, \quad (1)$$

where  $(c_x, c_y)^T$  is the principal point in pixels. Let the image resolution, i.e., the total pixels that the vision sensor has be  $M \cdot N$ , then the pixel coordinates of the principal point are equal to half the resolution.  $f_x$  and  $f_y$  are the horizontal and vertical pixel focal lengths, respectively. They are usually not directly available but can be obtained by multiplying the lens focal length  $f$  with the scaling factors  $\alpha$  and  $\beta$  from the image sensor (CMOS or CCD) size to the pixel plane size. Let the sensor size be  $x_{\text{sens}} \cdot y_{\text{sens}}$ , and the pixel plane size is numerically equal to the image resolution  $M \cdot N$ , the scaling factors can be estimated by

$$\alpha = \frac{M}{x_{\text{sens}}}, \quad \beta = \frac{N}{y_{\text{sens}}}, \quad (2)$$

Then the pixel focal length in both directions can be obtained by  $f_x = \alpha f$  and  $f_y = \beta f$ . Note that the pixel focal length no longer has the measure of length and its unit is the pixel.

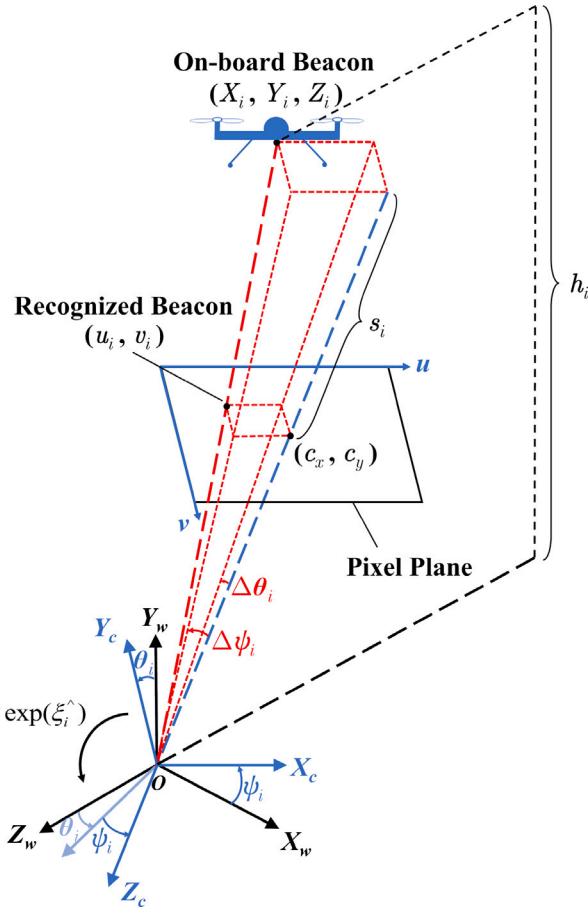


Fig. 2. Geometry for the proposed method.

### 3.2. Vision sensor extrinsics

For visual localization, the extrinsics represents the pose transformation of the vision sensor, including translation and rotation [42–44]. We first describe the rotation and translation of the vision sensor based on the geometry for the proposed method, and then elaborate on the construction of the extrinsics with the Lie group. As shown in Fig. 2,  $X_wY_wZ_w$  represents the world coordinate system used for localization. With the initial position of the vision sensor's optical center as the origin, the initial attitude parallel to the ground defines the three-axis orientation by the right-hand frame.  $X_cY_cZ_c$  represents the camera coordinate system defined by the current orientation of the vision sensor. During localization, the position of the vision sensor is stationary. While in orientation it has 2-DoF including pitch and yaw. As shown in Fig. 2, for the  $i$ th frame, the angle at which the sensor rotates around the  $x$ -axis is the pitch  $\theta_i$ , i.e., the angle between  $Y_w$  and  $Y_c$ . Correspondingly, the angle at which the sensor rotates around the  $y$ -axis is the yaw  $\psi_i$ , i.e., the angle between  $X_w$  and  $X_c$ . With this definition of orientation, there is also theoretically a roll defined by the angle at which the sensor rotates around the  $z$ -axis. However, in this work, the vision sensor has no DoF in terms of roll. In other words, it cannot perform rotations around the  $z$ -axis, thus roll is not considered in the orientation.

In this context, the vision sensor extrinsics can be constructed based on Lie group SE(3). We first consider the Lie algebra  $\mathfrak{se}(3)$  corresponding to SE(3), which belongs to the  $\mathbb{R}^6$  space [43], as follows

$$\mathfrak{se}(3) = \left\{ \xi_i = \begin{pmatrix} \rho_i \\ \phi_i \end{pmatrix} \in \mathbb{R}^6, \rho_i \in \mathbb{R}^3, \phi_i \in \mathfrak{so}(3) \right\}, \quad (3)$$

where  $\xi_i$  is the extrinsics expressed as a vector at frame  $i$ . Its first three dimensions represent the translation, denoted as  $\rho_i$ . The second three dimensions represent the rotation, denoted as  $\phi_i$ . Now we need to transform the vectors in  $\mathfrak{se}(3)$  into matrices in SE(3). In Lie algebra  $\mathfrak{se}(3)$ , the  $\wedge$  operation maps vectors to matrices [43], then we have

$$\xi_i^\wedge = \begin{pmatrix} \phi_i^\wedge & \rho_i \\ \mathbf{0}_3^T & 0 \end{pmatrix} \in \mathbb{R}^{4 \times 4}, \quad (4)$$

where  $\mathbf{0}_3^T$  denotes the transpose of 3-dimensional zero vector. In  $\mathfrak{so}(3)$ , the  $\wedge$  operation maps vectors to antisymmetric matrices belonging to the  $\mathbb{R}^{3 \times 3}$  space, i.e.,  $\phi_i^\wedge \in \mathbb{R}^{3 \times 3}$  [43]. Here we are not concerned with the specific form of  $\phi_i^\wedge$ , but simply derive its exponential mapping from Rodrigues' formula [43,44] as

$$\exp(\phi_i^\wedge) = \sum_{n=0}^{\infty} \frac{1}{n!} (\phi_i^\wedge)^n = \mathbf{R}_i, \quad (5)$$

where  $\mathbf{R}_i$  is the Euler matrix of the  $i$ th frame

$$\mathbf{R}_i = \begin{pmatrix} \cos \psi_i & 0 & -\sin \psi_i \\ \sin \theta_i \sin \psi_i & \cos \theta_i & -\sin \theta_i \cos \psi_i \\ \cos \theta_i \sin \psi_i & \sin \theta_i & \cos \theta_i \cos \psi_i \end{pmatrix}, \quad (6)$$

Notice that  $\mathbf{R}_i \in \text{SO}(3)$ . It is also known that the transformation matrix in SE(3) is composed of the Euler matrix in SO(3) and the vector representing the translation [43]. Then similarly, we can transform  $\xi_i$  to SE(3) also by the exponential mapping [44], as follows

$$\exp(\xi_i^\wedge) = \begin{pmatrix} \sum_{n=0}^{\infty} \frac{1}{n!} (\phi_i^\wedge)^n & \sum_{n=0}^{\infty} \frac{1}{(n+1)!} (\phi_i^\wedge)^n \rho_i \\ \mathbf{0}_3^T & \exp(0) \end{pmatrix} = \begin{pmatrix} \mathbf{R}_i & \mathbf{t} \\ \mathbf{0}_3^T & 1 \end{pmatrix}. \quad (7)$$

As mentioned above, since the translation of the visual sensor is not considered in this work,  $\rho_i$  can be regarded as zero vector in the operation. Thus we have  $\mathbf{t} = \mathbf{0}_3$ . At this point, the extrinsics denoted by SE(3) is obtained.

### 3.3. Depth estimation

In this work, we measure the altitude difference between the UAV and the vision sensor to solve for depth. The on-board strapdown altitude sensor measures the height above ground of the UAV  $h_d$ . Since the position of the vision sensor is static during localization, the altitude above ground at its optical center  $h_c$  is also constant. Therefore, the altitude difference between the UAV and the vision sensor at frame  $i$  is  $h_i = h_d - h_c$ . The projected point of the beacon on the image plane and the principal point of the image plane each connected to the origin  $O$  form an inclination. It can be decomposed into horizontal and vertical angular deviations in terms of the direction of the pixel plane's two axes, which are classified into the notations  $\Delta\psi_i$  and  $\Delta\theta_i$  for the  $i$ th frame, respectively. Then according to the proposed localization geometry, the depth  $s$  can be calculated by

$$s_i = \frac{h_i}{\sin \theta_i + \cos \theta_i \tan \Delta\theta_i}, \quad (8)$$

where

$$\tan \Delta\theta_i = -\frac{v_i - c_y}{f_y}. \quad (9)$$

### 3.4. Position solution

After obtaining the beacon's pixel coordinates by recognition and depth by altimetry, combined with the intrinsics and extrinsics of the vision sensor, the observation equation can be set up to solve the UAV's spatial position. Note that the 3D spatial position of the UAV is 4-dimensional when expressed in homogeneous coordinates, while the homogeneous pixel coordinates obtained from beacon recognition

are 3-dimensional. Thus the corresponding processing is required in order to satisfy the requirements of matrix multiplication during the computation. In summary, the following observation equation can be set up as

$$s_i \mathbf{p}_i = \mathbf{K}(\exp(\xi_i^\wedge) \mathbf{P}_i)[\mathcal{J}, \mathcal{K}], \quad (10)$$

where  $\mathbf{p}_i = (u_i, v_i, 1)^T$  is the pixel coordinates of the  $i$ th frame, and  $\mathbf{P}_i = (X_i, Y_i, Z_i, 1)^T$  is the position of the UAV in the world coordinate system. Both coordinates are expressed in homogeneous form.  $(\exp(\xi_i^\wedge) \mathbf{P}_i)[\mathcal{J}, \mathcal{K}]$  denotes the submatrix taken without loss of generality [45] to satisfy the matrix multiplication requirement in the operation. Considering the size of  $(\exp(\xi_i^\wedge) \mathbf{P}_i)$  as  $m \cdot n$ , we have

$$\mathcal{J} = \bigcup_{j=1}^m \binom{[m]}{1}, \quad \mathcal{K} = \bigcup_{k=1}^n \binom{[n]}{1}. \quad (11)$$

#### 4. Beacon design & recognition

##### 4.1. Beacon design

The basic principle of beacon design is to increase the signal-to-noise ratio of the beacon as much as possible, i.e., to exclude interference from other light sources in the environment. Since the NIR light is independent of the visible spectrum, interference from visible light is naturally excluded. During the daytime, however, there is inevitably interference from the sun. The sun, as an idealized blackbody radiation source, emits light in the full wavelength spectrum at different intensities [46]. On average, the sunlight is most intense in the visible band and less intense in the infrared and ultraviolet bands. Most vision sensors operate effectively in the visible and infrared bands, so an infrared light source is most suitable as a beacon.

Since light transmission in the atmosphere will be attenuated by the scattering and absorption of atmospheric molecules, the intensity of sunlight at different wavelengths will have different degrees of decrease, and the greatest degree of decrease in the NIR band is 940 nm [37,46], as shown in Fig. 3. Therefore, we select the NIR LED of this wavelength as the beacon, and configure the vision sensor with a narrowband filter of the same wavelength. In addition, to sufficiently separate the beacon from the background and raise its signal-to-noise ratio, we set the vision sensor's exposure time to a low level of 10–100 μs.

##### 4.2. Beacon recognition

Although the intensity of sunlight at 940 nm is low, there is still interference during the day. As shown in Fig. 4, in the absence of solar interference, the background is basically filtered out to appear black, while only the beacon is clearly visible in the image. Since most of the background is filtered out after narrowband-pass, the signal-to-noise ratio of the beacon in the image is already high, so the recognition can theoretically be accomplished by binary segmentation. Fixed-threshold-based methods require thresholds to be determined in advance and are unable to be changed during use, resulting in inconvenience and sensitivity to illumination changes that may occur during UAV flight. Adaptive segmentation algorithms such as the widely used Otsu method [47], may be able to address these issues to some extent. However, since the area occupied by the beacon in the image is too small resulting in an extremely low proportion of its histogram component, adaptive segmentation fails to effectively separate the beacon. Methods with high computational resource requirements, such as those based on neural networks, have extreme difficulty in ensuring real-time so they are unnecessary and unsuitable for the purposes of this work. In addition, on a clear day, if the sun is unavoidably present in the field of view, there would be strong interference, resulting in the most extreme challenge of the beacon recognition.

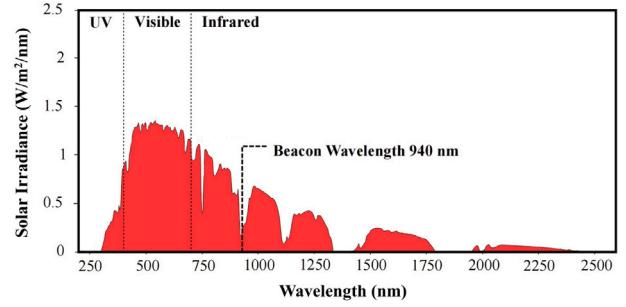


Fig. 3. The solar radiation spectrum.

---

#### Algorithm 1 GSFTM algorithm

```

Input: narrowband-pass image sequence  $\mathcal{F} = \{f_i(x, y)\}$  containing  $l$  frames, where  $x \in [1, M]$ ,  $y \in [1, N]$   $\triangleright$  Input images are gray-scale.
Output: segmented image sequence  $\mathcal{G} = \{g_i(x, y)\}$ 
Output: beacon pixel coordinates sequence  $\mathcal{P} = \{(u_i, v_i)\}$ 

1: Calculate the initial frame's gradient vector  $\nabla f_1(x, y) = \left( \frac{\partial f_1(x, y)}{\partial x}, \frac{\partial f_1(x, y)}{\partial y} \right)$ 
2: Calculate gradient magnitude from Euclidean paradigm of gradient vector  $M_1(x, y) = \|\nabla f_1(x, y)\|$ 
3: if the  $M_1(x, y)$  is an outlier then
4:    $temp \leftarrow (x, y)$ 
5: end if
6: if  $(x, y) \in temp$  then
7:    $g_1(x, y) \leftarrow 1$ 
8: else
9:    $g_1(x, y) \leftarrow 0$ 
10: end if
11: for  $i \geq 2$  do
12:    $NSD_i(x, y) \leftarrow \frac{\sum_{(a,b) \in temp} (f_{i-1}(a,b) - f_i(x+a,y+b))^2}{\sqrt{\sum_{(a,b) \in temp} f_{i-1}^2(a,b) \cdot \sum_{(a,b) \in temp} f_i^2(x+a,y+b)}}$ 
13:   if the  $NSD_i(x, y)$  is an outlier then
14:      $g_i(x, y) \leftarrow 1$ 
15:      $temp \leftarrow (x, y)$ 
16:   else
17:      $g_i(x, y) \leftarrow 0$ 
18:   end if
19: end for
20: Calculate the expectation of  $x$  and  $y$  from each set's non-zero points in  $\mathcal{G} = \{g_i(x, y)\}$  yields  $\mathcal{P} = \{(u_i, v_i)\}$ 

```

---

In this context, the GSFTM algorithm with low computational resource requirements is designed to accomplish beacon recognition efficiently. For sequences of images captured by the vision sensor in localization, gradient vectors are first calculated for each point of the initial frame. Then the magnitude is obtained from the vector to get the edge information. In this way the area of the beacon in the image can be segmented and thus the pixel coordinates can be obtained. Since the placement of the vision sensor is pre-determined in practice, it is always possible to manually ensure that there are no solar interferences in the initial frame. Thus after the processing, the initial template is obtained. This allows the template to be matched to the next frame based on the normalized squared difference (NSD) to segment the beacons and obtain the pixel coordinates. In the subsequent processing, the segmented image obtained in each frame is used as the template for the next frame. Thus the beacon recognition can be accomplished by sequential frame template matching. As shown in Fig. 4, even though there is solar and reflective interference in the sequential frames, the beacons can still be recognized accurately. To minimize computational

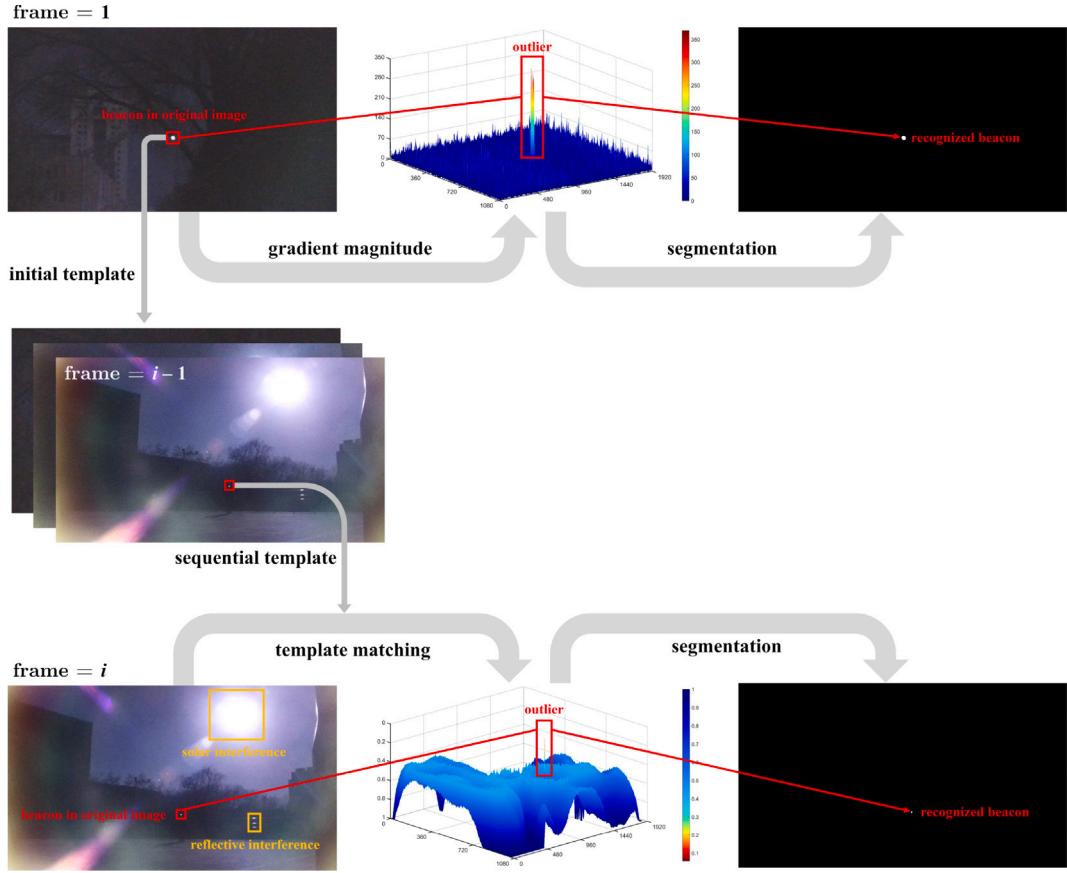


Fig. 4. The procedure of the designed GSFTM algorithm and the processing results.

costs, it is sufficient to keep only one template at runtime. Meanwhile, the color features of beacons can be ignored due to their high signal-to-noise ratio in narrowband-pass images. Therefore it is reasonable and efficient to convert the multi-channel image captured by the vision sensor to single-channel grayscale form before processing. The complete process is shown in Algorithm 1.

## 5. Experiments

### 5.1. Experimental setup

In order to fully validate the performance of the proposed method, we conducted experiments in various real-world scenarios. We set up typical scenarios including indoor and outdoor based on environmental features and light conditions to demonstrate the versatility of the proposed method. Fig. 5 shows the experimental scenarios under favorable illumination conditions. Note that we also conducted experiments in these four typical environments under extreme illumination conditions. Direct demonstrations of dark environments are not shown here as poor lighting can make it difficult to discern. Experimental results for all real-world scenarios and related discussions are detailed below.

For the indoor experiments, in order to measure the accuracy of the proposed method, we used the FZMotion motion capture system with millimeter-level accuracy to obtain the ground truth. We deployed the calibration rods used to determine the FZMotion system in close proximity to the vision sensor and oriented them in the same initial orientation. By measuring and correcting the deviation from the origin we can overlap their right-hand coordinate systems to compare the trajectories they solve. Since the motion capture camera of the FZMotion system adopts 850 nm light source, there will be no interference with the NIR vision of the proposed method. In the outdoor experiments, we take the localization results of a common consumer-grade

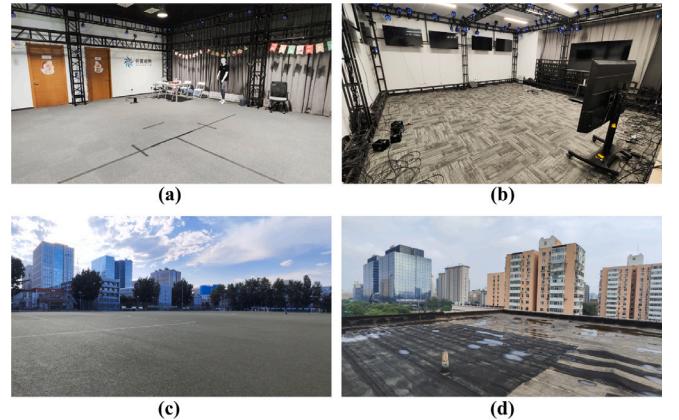
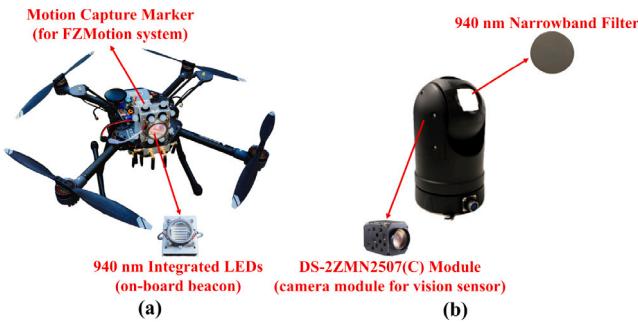


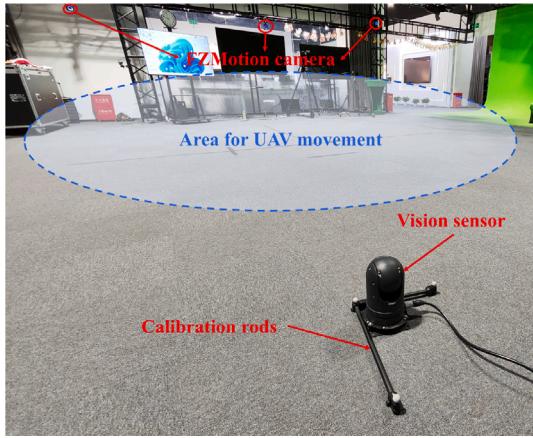
Fig. 5. Experimental scenarios under favorable illumination conditions. (a) Feature-rich indoor environment (Scenario 1). (b) Feature-poor indoor environment (Scenario 3). (c) Open outdoor environment (Scenario 5). (d) Outdoor environment of urban buildings (Scenario 7).

Global Positioning System (GPS) module as a reference to compare the performance of the proposed method with GNSS.

The main experimental devices consisted of a UAV with required loads and an external narrowband-pass vision sensor, as shown in Fig. 6. 940 nm integrated LEDs with heat sink were installed on the UAV as an on-board beacon. TFmini-S laser rangefinder and SPL06 barometer are used as the strapdown altitude sensors with high accuracy for indoor and outdoor experiments, respectively. In order to eliminate the unfavorable effects of ground pressure variations on the altitude measurements as well as to offset some of the systematic errors,



**Fig. 6.** Experimental devices. (a) Experimental UAV with required loads mainly include 940 nm integrated LEDs used as the on-board NIR beacon and motion capture markers for FZMotion system used to get ground truth in indoor experiments. (b) External vision sensor is mainly composed of DS-2ZMN2507(C) camera module with a 940 nm narrowband filter.



**Fig. 7.** Sensor placement and UAV movement area in real-world indoor experiments.

**Table 1**

Indoor experimental settings and performance of the proposed method.

Setting	Environment	Illumination	ATE RMSE (m)	Peak error (m)
Scenario 1	Feature-rich	Bright	0.0684	0.1245
Scenario 2	Feature-rich	Dark	0.0643	0.1157
Scenario 3	Feature-poor	Bright	0.0537	0.1003
Scenario 4	Feature-poor	Dark	0.0559	0.1087

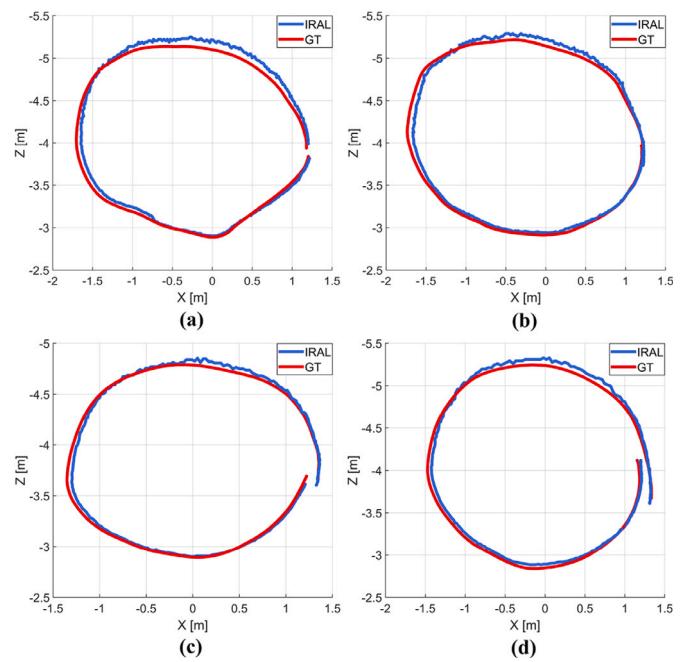
the same type of barometers are integrated into both the UAV and the vision sensor in the outdoor experiments. Hikvision's DS-2ZMN2507(C) camera module with a 940 nm narrowband filter form the off-board narrowband-pass vision sensor, which also contains a set of 2-DoF servo motors with an angular resolution of 0.01° in the pitch and yaw directions to derive accurate extrinsics.

## 5.2. Indoor experiments

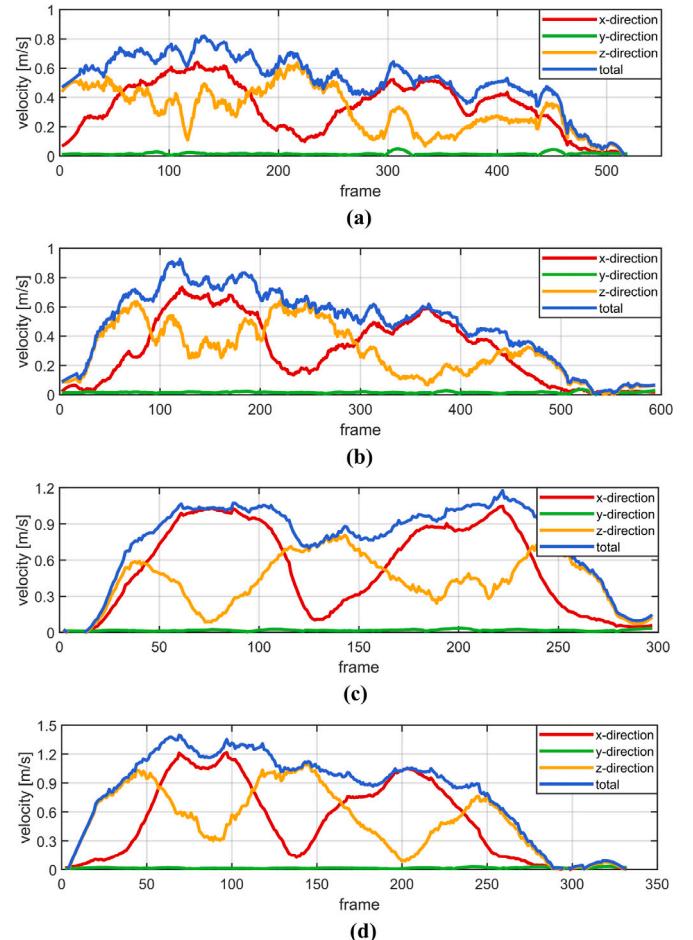
### 5.2.1. Evaluation metrics

In order to quantitatively assess the performance of the proposed method, the absolute trajectory error (ATE) for test sequences of length  $l$  is employed. ATE reflects the deviation between the measured values derived from the proposed method and the ground truth. In order to harmonize the magnitude of the error with the experimental data and to avoid the influence of the length of the test sequence on the results, the ATE is defined based on the root mean square error (RMSE), as follows

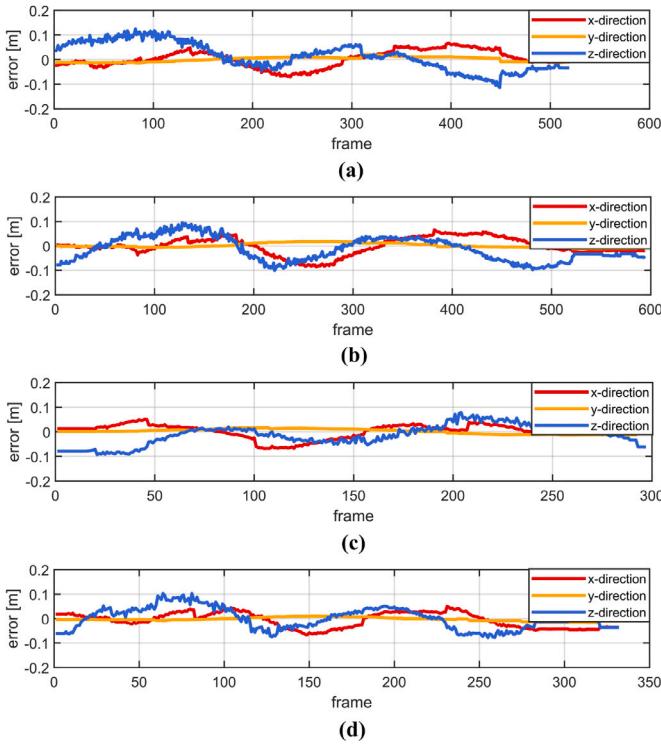
$$\text{ATE} = \left( \frac{1}{l} \sum_{i=1}^l \|P_i - \hat{P}_i\|^2 \right)^{\frac{1}{2}}, \quad (12)$$



**Fig. 8.** Trajectories under indoor experimental scenarios. (a) Trajectories for Scenario 1. (b) Trajectories for Scenario 2. (c) Trajectories for Scenario 3. (d) Trajectories for Scenario 4.



**Fig. 9.** UAV velocities in indoor experiments. (a) Velocity for Scenario 1. (b) Velocity for Scenario 2. (c) Velocity for Scenario 3. (d) Velocity for Scenario 4.



**Fig. 10.** The error of the proposed method in x, y, z directions for each experimental scenario. (a) Error for Scenario 1. (b) Error for Scenario 2. (c) Error for Scenario 3. (d) Error for Scenario 4.

where  $P_i$  and  $\hat{P}_i$  denote the ground truth and the measured value of the  $i$ th frame in the test sequence, respectively. It is represented in the Table 1 as ATE RMSE. In order to examine whether there is a situation where the proposed method deviates excessively from the ground truth at a certain point in the test sequence, the peak error, i.e.,  $\max \|P_i - \hat{P}_i\|$  is calculated as well.

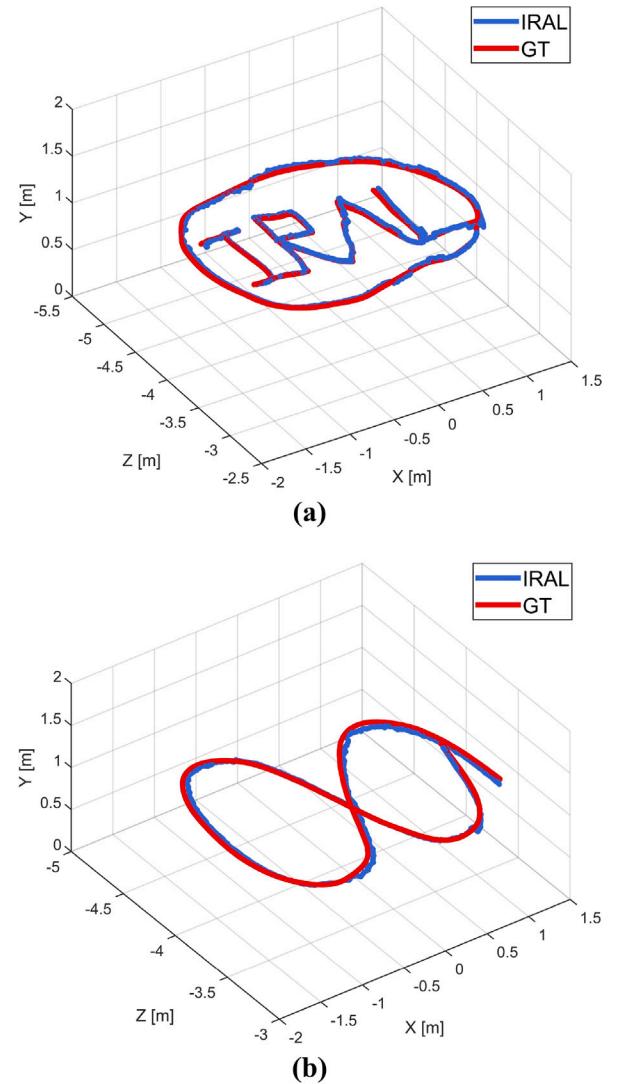
In addition, we also analyzed the error of the proposed method in three directions for each experimental scenario separately. The results are also expressed as the RMSE between the measured coordinate and the ground truth of the corresponding direction, as shown in Fig. 10.

### 5.2.2. Experimental results

We manipulate the UAV moving along approximately circular trajectories and simultaneously measure the spatial trajectory using the proposed IRAL and the FZMotion system under each experimental scenario, as shown in Fig. 7. The measured trajectories are shown in Fig. 8 and the performance is shown in Table 1. From the results, it can be seen that the proposed method is able to achieve centimeter-level accuracy stably regardless of both environmental features and illumination conditions. The peak errors in all experimental scenarios are also around 0.1 m, indicating that the proposed method is not subject to large jumps during localization. Additionally, in conjunction with the velocity curves demonstrated in Fig. 9, it is evident that the speed of the UAV does not affect the performance of the proposed method.

According to Fig. 10, the proposed method performs high accuracy in each direction as well, where the errors in the z-direction are slightly larger. This is due to the fact that this error is mainly generated by the vertical pixel error of the recognized beacon in the image plane. Since the vision sensor's field of view is a rectangle that is horizontally wide and vertically narrow, the vertical pixel resolution is lower resulting in larger errors.

In addition, experiments under complex preset trajectories are also conducted. We manipulate the UAV to move along the track of the



**Fig. 11.** Results under complex trajectories. (a) Trajectory of the letters that make up the name of the proposed method. (b) Trajectory of eight-pattern.

letters IRAL which make up the name of the proposed method and the path of eight-pattern. The former is more complex compared to the latter. The measured trajectories are shown in Fig. 11. According to the results, the ATE RMSE of the proposed method is 0.0783 m and 0.0697 m under two complex trajectories, respectively. It demonstrates that the proposed method is able to maintain centimeter-level accuracy even though the motion trajectory of the UAV is more complicated. Thus, its robust performance is further demonstrated.

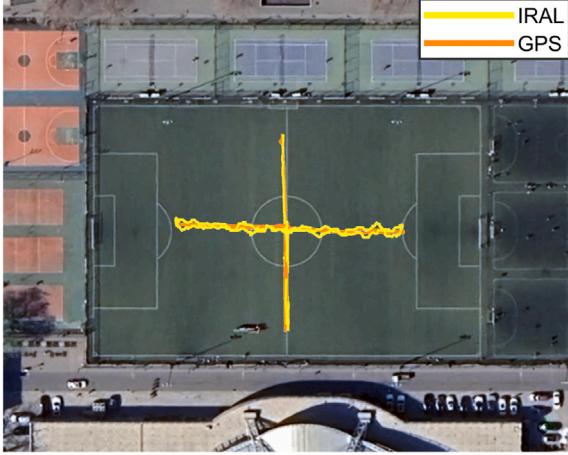
### 5.3. Outdoor experiments

The outdoor experiments are conducted in two venues with the detailed settings shown in Table 2. The experiments in an open football field are conducted to evaluate the performance of the proposed method under long trajectories. In the result shown in Fig. 12, the lengths of the two trajectories measured by the proposed method at daytime and night are 212.68 m and 232.93 m, respectively. During UAV flight, a common consumer-grade GPS module performs simultaneous localization for evaluation. It should be noted that the absolute position measurements of consumer-grade GPS modules on their own typically have large drifts. Thus for GNSS-based localization of UAVs in

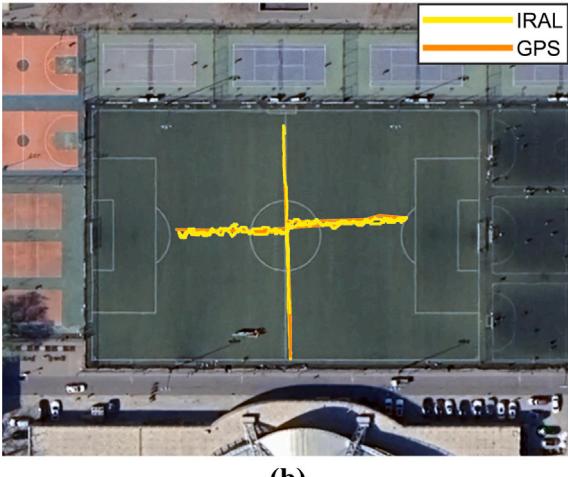
**Table 2**

Outdoor experimental settings of the proposed method.

Setting	Environment	Illumination	Aver. depth (m)
Scenario 5	Open football field	Daytime	39.30
Scenario 6	Open football field	Night	42.33
Scenario 7	Rainy urban building	Daytime	125.43
Scenario 8	Rainy urban building	Night	125.30



(a)

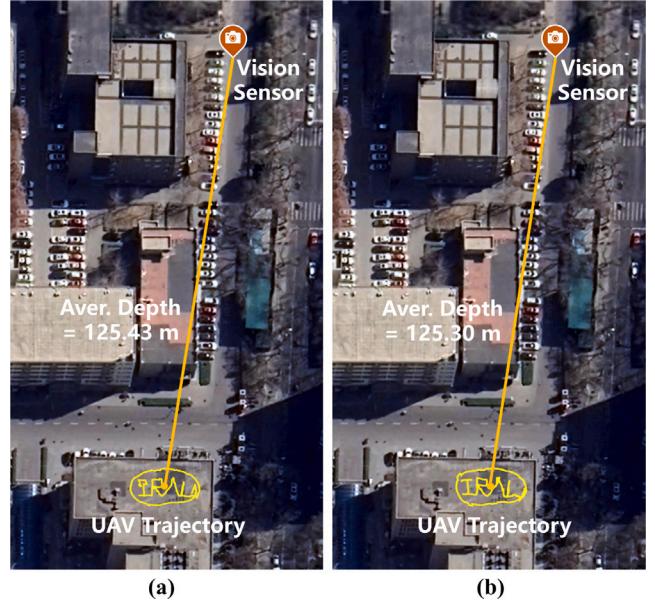


(b)

**Fig. 12.** Trajectories for outdoor experiments in an open football field. (a) Trajectories for Scenario 5. (b) Trajectories for Scenario 6.

real-world scenarios, the consumer-grade GPS module is always used in an integrated navigation system with inertial measurement units to ensure relative drift-free during motion. Specifically, the relative accuracy can be maintained at the centimeter level with good satellite signals. And this accuracy of the module used in the experiment is tested to be about 4 cm RMSE. Thus it is feasible to calculate the deviation between the two trajectories provided they are aligned to the same coordinate system and start points to evaluate the drift of the proposed method. As seen in Fig. 12, the results of the two localization systems are very close and the proposed method achieves drift-free during long trajectory localization. As with the ATE calculation, the RMSE is still used to quantify this deviation. The results of Scenario 5 and Scenario 6 are 0.279 m and 0.326 m, respectively. This indicates that the proposed IRAL has the ability to replace or assist GNSS in sustained flight in outdoor environments.

The main purpose of the outdoor experiments on an urban building is to validate the localization performance of the proposed method at

**Fig. 13.** Experimental results at large depths in outdoor environments of an urban building. (a) Results for Scenario 7. (b) Results for Scenario 8.

large depths, i.e., ranges between the beacon and the vision sensor. Meanwhile, in order to evaluate the influence of severe weather on the proposed method, we organized experiments especially on a rainy day. However, it should be noted that since the electronic devices included in this work are not waterproof, we chose to conduct the experiment during the time period when the rain had just stopped. Consistent with the outdoor experiments at the previous venue, to verify the effect of illumination conditions, we tested at both daytime and night, with average depths of 125.43 m and 125.30 m, respectively. In this context, in order to investigate the robustness under complex trajectories, we take the same path as in the indoor experiment, which is manipulating the UAV to move along the track of the letters IRAL. As can be seen from the results shown in Fig. 13, the proposed method achieves drift-free even at large depths, and the shape of the complex trajectories is generally well preserved. This also demonstrates the capability of the proposed method to be practically deployed in urban environments with severe weather.

#### 5.4. Discussion

**Table 3** lists recent cooperative visual localization methods. In comparison, the proposed IRAL not only achieves the highest level of accuracy with a minimum number of required vision sensors and beacons, but also has significant advantages in terms of environmental adaptability as well as illumination sensitivity. The proposed method works both indoors and outdoors and maintains high accuracy at large depths exceeding a hundred meters. From a cost-effectiveness perspective, the vision sensor adopted in the proposed method is essentially composed of a low-cost RGB camera without relying on expensive special sensors such as DVS. Therefore, the proposed method achieves the optimal comprehensive performance among similar works and can be used as a robust and versatile solution for UAV localization in GNSS-denied environments.

Naturally, there are some limitations to the proposed method accordingly. In this work, the prerequisite for ensuring localization accuracy and stability is that the on-board beacon is always recognized by the external vision sensor during the flight. For outdoor applications, although real-world experiments have demonstrated that the proposed method is assured to be drift-free at large depths and long trajectories,

**Table 3**

Comparison of the proposed method with recent optical cooperative localization works.

Method	Vision sensor type <sup>a</sup>	Category	Required beacons	Environmental adaptability	Illumination sensitivity	Accuracy	Maximum depth (m)
Proposed IRAL	NIR band-pass RGB	External	1	Indoor & outdoor	Inensitive	Centimeter-level	126.8
SLO-VLP [20]	RGB	Internal	3	Indoor only	Sensitive	Centimeter-level	2.7
LED lamp-based [22]	RGB	Internal	1	Indoor only	Sensitive	Centimeter-level	3
VI-RPE [26]	PF-MPE Camera	External	3	Indoor & outdoor	Sensitive	Decimeter-level	15
UV/DAR-based [25]	Ultraviolet Camera	External	3	Outdoor only	Sensitive	Meter-level	15
DVS-based [27]	UV/IR cut DVS	External	1	Indoor only	Insensitive	Centimeter-level	6
Planar marker-based [28]	RGB	Internal	12	Indoor & outdoor	Sensitive	Decimeter-level	1.5
Infrastructure-based [37]	NIR band-pass RGB	Internal	3	Outdoor only	Insensitive	Meter-level	150
CoBe [40]	IR band-pass RGB	External	6	Indoor only	Insensitive	Centimeter-level	2

<sup>a</sup> All sensors listed are monocular.

the vision sensor needs to operate at long focal lengths in order to maintain the robustness of the beacon recognition. For a varifocal camera, elongation of the focal length implies a reduction in the field of view, which increases the possibility that the UAV flies out of visible reach resulting in the beacon capture failure. This problem can currently be solved by manually manipulating the vision sensor orientation. However, this would still negatively affect the convenience of the proposed method in practical situations where UAVs frequently maneuver at high speeds. Thus the current outdoor applications of the proposed method focus on scenarios where the UAV hovers or cruises continuously in a certain area to perform its functions, such as precision agriculture, firefighting, SAR, SHM, etc. Approaches that are expected to greatly improve this problem will be explored in future work, which is described in the final section.

### 5.5. Applications

The proposed method can be used as a cost-effective and efficient solution for UAV localization in GNSS-denied environments due to its advantages in robustness and versatility. In particular, in the following application scenarios, the proposed method not only replaces GNSS to accomplish localization, but also well bridges the gap where visual localization in general cannot be used.

#### 5.5.1. Feature-poor closed environments

Although many visual localization methods, such as SLAM, can perform well in closed environments such as indoors and urban canyons, they are susceptible to degradation or even failure if there is a lack of features in the environment. While for this work, the on-board NIR beacon provides robust features as a cooperative target, which makes the performance of the proposed IRAL in practical use completely independent of the environmental features.

#### 5.5.2. Open outdoor environments

When UAVs fly in open outdoor environments, such as plateaus, grasslands, and other wildernesses, they are often tens of meters above the ground. This results in environmental features being located beneath and at a distance from the UAV. Even if using an internal cooperative method, i.e., deploying beacons on the ground while using an on-board vision sensor to recognize and then achieve localization, the same difficulty of deficient robustness of the features is faced. In contrast, The proposed method not only effectively overcomes the difficulties in feature robustness through NIR narrowband-pass vision, but also is well suited for scenarios where the UAV is working in a certain area that is open and unobstructed. A typical case is the application of tethered UAVs which are now gaining increasing attention [48,49]. It is an emerging type of UAV that operates for long periods of time by connecting itself to a ground power source with a high-voltage wire for continuous electricity. Whereas the constant wired power supply limits the flight range of tethered UAVs as well. Thus in practice they mostly function by hovering or cruising over a limited area. The proposed method is well suited for these cases, so that it has a promising application to tethered UAVs.

#### 5.5.3. GNSS-denied environments in extreme illumination conditions

Since most visual localizations, including cooperative methods, rely on visible light, they are highly sensitive to illumination conditions. In this work, the on-board beacon operates in the NIR band, and the vision sensor also performs narrowband-pass at the same wavelength as the beacon, so that extreme illumination conditions can be disregarded. Thus the proposed IRAL overcomes the limitations of many visible light methods that cannot be used in extreme environments.

### 6. Conclusion and future work

In this work, we propose a robust and versatile UAV localization method using infrared vision and altitude sensor fusion. The proposed method stably performs high accuracy regardless of illumination conditions with low-cost components, which achieves optimal performance among similar works. The proposed method can thus serve as a simple yet efficient solution for UAV localization in GNSS-denied environments.

Although the proposed IRAL has already reached the maturity of real-world deployment, there are still some issues that can be addressed for future research. During the flight of UAVs in real-world environments, especially for outdoor applications, there is a possibility of flying out of the field of view or being obscured by interfering objects. Whereas the current scheme of manually manipulating the vision sensor orientation brings inconvenience. One effective solution to this issue is to automate the sensor orientation manipulation, i.e., achieve automatic adaptation of the field of view to beacon motion. Or deploy it on a mobile platform. This would require more sophisticated models to be constructed in future work. Another issue is that the proposed method currently only solves the spatial position of the UAV and cannot produce the attitude. Therefore, in future work we would design a coupling method to integrate IRAL with an inertial navigation system for 6-DoF localization.

### CRediT authorship contribution statement

**Xixian Li:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Qiang Wang:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Zhonghu Hao:** Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Shengrong Hu:** Validation, Software, Investigation, Data curation. **Jiaxing Wu:** Validation, Software, Data curation. **Linkang Dong:** Validation, Investigation.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The authors would like to express their gratitude to Xiangyang Han, Shuai Yan and Yiming Sun at Beijing YUAN Mate Co., Ltd., the developer of the FZMotion motion capture system, for their help in the indoor experiments.

## Data availability

The authors are unable or have chosen not to specify which data has been used.

## References

- [1] X. Lei, X. Xu, G. Wang, H. Luo, A multi-UAV deployment method for border patrolling based on Stackelberg game, *J. Syst. Eng. Electron.* 34 (1) (2023) 99–116, <http://dx.doi.org/10.23919/JSEE.2023.000022>.
- [2] M. Lyu, Y. Zhao, C. Huang, H. Huang, Unmanned aerial vehicles for search and rescue: A survey, *Remote Sens.* 15 (13) (2023) 3266, <http://dx.doi.org/10.3390/rs15133266>.
- [3] D. Perez-Saura, M. Fernandez-Cortizas, R. Perez-Segui, P. Arias-Perez, P. Campoy, Urban firefighting drones: Precise throwing from UAV, *J. Intell. Robot. Syst.* 108 (4) (2023) 66, <http://dx.doi.org/10.1007/s10846-023-01883-6>.
- [4] P. Zhang, X. Sun, D. Zhang, Y. Yang, Z. Wang, Lightweight deep learning models for high-precision rice seedling segmentation from UAV-based multispectral images, *Plant Phenomics.* 5 (2023) 0123, <http://dx.doi.org/10.34133/plantphenomics.0123>.
- [5] K. Strzbała, P. Ćwiąkała, W. Gruszczyński, E. Puniach, W. Matwij, Determining changes in building tilts based on UAV photogrammetry, *Measurement* 202 (2023) 111772, <http://dx.doi.org/10.1016/j.measurement.2022.111772>.
- [6] J.A. Gonçalves, R. Henriques, UAV photogrammetry for topographic monitoring of coastal areas, *ISPRS-J. Photogramm. Remote Sens.* 104 (2015) 101–111, <http://dx.doi.org/10.1016/j.isprsjprs.2015.02.009>.
- [7] D. Kang, Y. Cha, Autonomous UAVs for structural health monitoring using deep learning and an ultrasonic beacon system with geo-tagging, *Comput. Aided Civ. Infrastruct. Eng.* 33 (10) (2018) 885–902, <http://dx.doi.org/10.1111/mice.12375>.
- [8] R. Ali, D. Kang, G. Sah, Y. Cha, Real-time multiple damage mapping using autonomous UAV and deep faster region-based neural networks for GPS-denied structures, *Autom. Constr.* 130 (2021) 103831, <http://dx.doi.org/10.1016/j.autcon.2021.103831>.
- [9] A. Couturier, M.A. Akhloufi, A review on absolute visual localization for UAV, *Robot. Auton. Syst.* 135 (2021) 103666, <http://dx.doi.org/10.1016/j.robot.2020.103666>.
- [10] N. Zhu, J. Marais, D. Bétaillie, M. Berbineau, GNSS position integrity in urban environments: A review of literature, *IEEE Trans. Intell. Transp. Syst.* 19 (9) (2018) 2762–2778, <http://dx.doi.org/10.1016/10.1109/TITS.2017.2766768>.
- [11] Y. Zhuang, X. Sun, Y. Li, J. Huai, L. Hua, X. Yang, X. Cao, P. Zhang, Y. Cao, L. Qi, Multi-sensor integrated navigation/positioning systems using data fusion: From analytics-based to learning-based approaches, *Inf. Fusion.* 95 (2023) 62–90, <http://dx.doi.org/10.1016/j.inffus.2023.01.025>.
- [12] B. Zhou, J. Pan, F. Gao, S. Shen, RAPTOR: Robust and perception-aware trajectory replanning for quadrotor fast flight, *IEEE Trans. Robot.* 37 (6) (2021) 1992–2009, <http://dx.doi.org/10.1109/TRO.2021.3071527>.
- [13] J. Zhang, R. Liu, K. Yin, Z. Wang, M. Gui, S. Chen, Intelligent collaborative localization among air-ground robots for industrial environment perception, *IEEE Trans. Ind. Electron.* 66 (12) (2019) 9673–9681, <http://dx.doi.org/10.1109/TIE.2018.2880727>.
- [14] T. Qin, P. Li, S. Shen, VINS-mono: A robust and versatile monocular visual-inertial state estimator, *IEEE Trans. Robot.* 34 (4) (2018) 1004–1020, <http://dx.doi.org/10.1109/TRO.2018.2853729>.
- [15] T. Qin, J. Pan, S. Cao, S. Shen, A general optimization-based framework for global pose estimation with multiple sensors, 2019, arXiv preprint [arXiv:1901.0364](https://arxiv.org/abs/1901.0364).
- [16] X. Yang, S. Zheng, X. Lin, F. Zhu, Improving RGB-D SLAM accuracy in dynamic environments based on semantic and geometric constraints, *Measurement* 217 (2023) 113084, <http://dx.doi.org/10.1016/j.measurement.2023.113084>.
- [17] M. Blöesch, S. Weiss, D. Scaramuzza, R. Siegwart, Vision based mav navigation in unknown and unstructured environments, in: Proc. IEEE Int. Conf. Robot. Automat. ICRA, 2010, pp. 21–28, <http://dx.doi.org/10.1109/ROBOT.2010.5509920>.
- [18] D. Kong, Y. Zhang, W. Dai, Direct near-infrared-depth visual SLAM with active lighting, *IEEE Robot. Automat. Lett.* 6 (4) (2021) 7057–7064, <http://dx.doi.org/10.1109/LRA.2021.3096741>.
- [19] E. Brassart, C. Pegard, M. Mouaddib, Localization using infrared beacons, *Robotica* 18 (2000) 153–161, <http://dx.doi.org/10.1017/S0263574799001927>.
- [20] W. Guan, L. Huang, B. Hussain, C. Yue, Robust robotic localization using visible light positioning and inertial fusion, *IEEE Sens. J.* 22 (6) (2022) 4882–4892, <http://dx.doi.org/10.1109/JSEN.2021.3053342>.
- [21] J. Krejsa, S. Vechet, Infrared beacons based localization of mobile robot, *Elektron. Elektrotech.* 117 (1) (2012) 17–22, <http://dx.doi.org/10.5755/j01.eee.117.1.1046>.
- [22] H. Li, H. Huang, Y. Xu, Z. Wei, S. Yuan, P. Lin, H. Wu, W. Lei, J. Fang, Z. Chen, A fast and high-accuracy real-time visible light positioning system based on single LED lamp with a beacon, *IEEE Photon. J.* 12 (6) (2020) 1–12, <http://dx.doi.org/10.1109/JPHOT.2020.3032448>.
- [23] J. Guo, Q. Zhang, H. Chai, Y. Li, Obtaining lower-body Euler angle time series in an accurate way using depth vision sensor relying on optimized kinect CNN, *Measurement* 188 (2022) 110461, <http://dx.doi.org/10.1016/j.measurement.2021.110461>.
- [24] N. Goldfarb, A. Lewis, A. Tacescu, G.S. Fischer, Open source vicon toolkit for motion capture and gait analysis, *Comput. Methods Programs Biomed.* 212 (2021) 1–13, <http://dx.doi.org/10.1016/j.cmpb.2021.106414>.
- [25] V. Walter, N. Staub, A. Franchi, M. Saska, UVDAR system for visual relative localization with application to leader-follower formations of multirotor UAVs, *IEEE Robot. Automat. Lett.* 4 (3) (2019) 2637–2644, <http://dx.doi.org/10.1109/LRA.2019.2901683>.
- [26] L. Teixeira, F. Maffra, M. Moos, M. Chil, VI-RPE: Visual-inertial relative pose estimation for aerial vehicles, *IEEE Robot. Automat. Lett.* 3 (4) (2018) 2770–2777, <http://dx.doi.org/10.1109/LRA.2018.2837687>.
- [27] H. Stuckey, A. Al-Radaideh, L. Sun, W. Tang, A spatial localization and attitude estimation system for unmanned aerial vehicles using a single dynamic vision sensor, *IEEE Sens. J.* 22 (15) (2022) 15497–15507, <http://dx.doi.org/10.1109/JSEN.2022.3187423>.
- [28] Z. Wang, Z. Zhang, W. Zhu, X. Hu, H. Deng, G. He, X. Kang, A robust planar marker-based visual SLAM, *Sensors* 23 (2) (2023) 1–13, <http://dx.doi.org/10.3390/s23020917>.
- [29] O. Ni, H. Cai, J. Yang, J. Wang, Targetless extrinsic calibration of vision sensor and low-resolution 3-D LiDAR, *IEEE Sens. J.* 23 (10) (2023) 10889–10899, <http://dx.doi.org/10.1109/JSEN.2023.3263833>.
- [30] D. Qiu, S. Li, T. Wang, Q. Ye, R. Li, K. Ding, H. Xu, A high-precision calibration approach for vision sensor-IMU pose parameters with adaptive constraint of multiple error equations, *Measurement* 153 (2020) 107402, <http://dx.doi.org/10.1016/j.measurement.2019.107402>.
- [31] M. Krawez, T. Caselitz, J. Sundram, M. Van Loock, W. Burgard, Real-time outdoor illumination estimation for vision sensor tracking in indoor environments, *IEEE Robot. Automat. Lett.* 6 (4) (2021) 6084–6091, <http://dx.doi.org/10.1109/LRA.2021.3090455>.
- [32] P. Bison, G. Chemello, C. Sossai, G. Trainito, Using a structured beacon for cooperative position estimation, *Robot. Auton. Syst.* 29 (1) (1999) 33–40, [http://dx.doi.org/10.1016/S0921-8890\(99\)00036-6](http://dx.doi.org/10.1016/S0921-8890(99)00036-6).
- [33] M. Cięzkowski, A prototype of static IR beacon-receiver positioning system based on triangulation method, *Measurement* 128 (2018) 149–159, <http://dx.doi.org/10.1016/j.measurement.2018.06.039>.
- [34] V. Pierlot, M. Van Droogenbroeck, Beams: A beacon-based angle measurement sensor for mobile robot positioning, *IEEE Trans. Robot.* 30 (3) (2014) 533–549, <http://dx.doi.org/10.1109/TRO.2013.2293834>.
- [35] V. Pierlot, M. Van Droogenbroeck, A new three object triangulation algorithm for mobile robot positioning, *IEEE Trans. Robot.* 30 (3) (2014) 566–577, <http://dx.doi.org/10.1109/TRO.2013.2294061>.
- [36] W. Ali, Y. Cha, Deep learning-based obstacle-avoiding autonomous UAVs with fiducial marker-based localization for structural health monitoring, *Struct. Health Monit.* 23 (2) (2023) 971–990, <http://dx.doi.org/10.1177/14759217231177314>.
- [37] A. Kampmann, M. Lamberti, N. Petrovic, S. Kowalewski, B. Alrifaei, Investigating outdoor recognition performance of infrared beacons for infrastructure-based localization, in: Proc. IEEE Intell. Veh. Symp. IV, 2022, pp. 1107–1113, <http://dx.doi.org/10.1109/IV51971.2022.9827288>.
- [38] L. Hugh Sing, P. Grantham, Positioning beacon system using digital vision sensor and LEDs, *IEEE Trans. Veh. Technol.* 52 (2) (2003) 406–419, <http://dx.doi.org/10.1109/TVT.2003.808800>.
- [39] J. Faigl, T. Krajnik, J. Chudoba, L. Preucil, M. Saska, Low-cost embedded system for relative localization in robotic swarms, in: Proc. IEEE Int. Conf. Robot. Automat. ICRA, 2013, pp. 993–998, <http://dx.doi.org/10.1109/ICRA.2013.6630694>.
- [40] R. Rabinovich, I. Jubran, A. Wetzel, R. Kimmel, Cobe - coded beacons for localization, object tracking, and SLAM augmentation, in: Proc. IEEE Int. Conf. Autom. Sci. Eng. CASE, 2020, pp. 1367–1374, <http://dx.doi.org/10.1109/CASE48305.2020.9216779>.
- [41] M. Lin, K. Lee, Outdoor target positioning using wii remote IR vision sensor and signal modulation, *Sensors* 20 (8) (2020) 2163, <http://dx.doi.org/10.3390/s20082163>.
- [42] V. Lippiello, J. Cacace, Robust visual localization of a UAV over a pipe-rack based on the Lie group se(3), *IEEE Robot. Automat. Lett.* 7 (1) (2022) 295–302, <http://dx.doi.org/10.1109/LRA.2021.3125039>.
- [43] J.M. Selig, Lie groups and Lie algebras in robotics, in: Proc. Comput. Noncommut. Algebra Appl., 2005, pp. 101–125, [http://dx.doi.org/10.1007/1-4020-2307-3\\_5](http://dx.doi.org/10.1007/1-4020-2307-3_5).

- [44] L. Quan, Z. Lan, The vectorial parameterization of rotation, *Nonlinear Dynam.* 32 (1) (2003) 71–92, <http://dx.doi.org/10.1023/A:1024265401576>.
- [45] A. Frieze, R. Kannan, Quick approximation to matrices and applications, *Combinatorica* 19 (2) (1999) 175–220, <http://dx.doi.org/10.1007/s004930050052>.
- [46] M. Iqbal, *An Introduction To Solar Radiation*, Academic Press, 1983.
- [47] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Trans. Syst. Man Cybern.* 9 (1) (1979) 62–66, <http://dx.doi.org/10.1109/TSMC.1979.4310076>.
- [48] M.A. Kishk, A. Bader, M. Alouini, On the 3-D placement of airborne base stations using tethered UAVs, *IEEE Trans. Commun.* 68 (8) (2020) 5202–5215, <http://dx.doi.org/10.1109/TCOMM.2020.2993885>.
- [49] O.M. Bushnaq, M.A. Kishk, A. Celik, M. Alouini, T.Y. Al-Naffouri, Optimal deployment of tethered drones for maximum cellular coverage in user clusters, *IEEE Trans. Wirel. Commun.* 20 (3) (2021) 2092–2108, <http://dx.doi.org/10.1109/TWC.2020.3039013>.