Robust Visual Positioning of the UAV for the Under Bridge Inspection With a Ground Guided Vehicle

Zhaoying Wang, Sensen Liu, Gang Chen, *Graduate Student Member, IEEE*, and Wei Dong, *Member, IEEE*

***Abstract*— Regular defect inspection of the bridge's bottom is necessary for the maintenance of the bridge. Usually, conducting such an inspection with traditional under bridge inspection vehicles (UBIVs) is high-cost and laborious. To improve the inspection efficiency, the newly developed unmanned aerial vehicle (UAV) technique may provide a promising alternative solution. As the global navigation satellite system (GNSS) is not available under the bridge, the robust positioning of the UAV is still a challenge during fully autonomous inspections. Although accumulated works have attempted to utilize visual odometry to localize the UAV, their performance may easily deteriorate under simultaneously existed varying illumination and intense light noises. To cope with this issue, we design a ground–air mobile system and a dual-source positioning algorithm to enhance the robustness of the UAV's positioning. Specifically, the ground part of the mobile system is a ground vehicle (GV) equipped with infrared markers, which provides the referential fiducials for the relative positioning of the UAV. To well identify the infrared markers, the spatial relationship between the UAV and the GV is first optimized by an observation model. Then, to guarantee the robustness of marker detection, both color image and infrared image are simultaneously captured, and a dual-source algorithm is proposed accordingly. To implement the algorithm, a candidate region containing the infrared markers is first identified by a deep convolutional neural network. Subsequently, this candidate region projects to the infrared image, and a searching–tracking–aiming algorithm robustly detects those infrared markers. Following the marker detection, the position of the UAV can be finally estimated by a perspective-3-point algorithm of the marker and an inertial measurement unit (IMU) compensation. To verify the performance of the developed positioning approach, we conduct various real-world experiments in challenging light conditions under the bridge. The results demonstrate that our ground–air system and dual-source algorithm can provide robust positioning for the UAV in intense light noises and varying illumination during the bridge inspection.**

***Index Terms*— Bridge inspection, ground–air system, visual positioning.**

I. Introduction

REGULAR defect inspection of a bridge is crucial for its maintenance and administration [1]–[4]. The inspection of the bridge mainly depends on under bridge inspection vehicles (UBIVs) at present. These tasks usually require lane closures and traffic guidance schemes [5], which are usually high-cost and time-consuming. To tackle this problem, employing unmanned aerial vehicles (UAVs) in those tasks would be promising [6]–[10]. Currently, UAV inspections are commonly semiautonomous and require pilot assistance [11]–[14]. To further improve the efficiency of inspection and reduce human costs, a fully autonomous solution is demanding [15]. However, without global navigation satellite system (GNSS) under the bridge, the robust positioning of the UAV is still a challenge [16], [17].

This challenge attracts many investigations. For example, Kakillioglu *et al.* [18] proposed a positioning method depending on matching depth images from a 3-D camera with the 3-D map of the bridge structure. The position of the UAV is estimated by backprojections of matching regions. Tomiczek *et al.* [19] utilized optical flow with the texture of the bridge to assist localization under the bridge. Loquercio *et al.* [20] utilized the T265 device to execute the visual simultaneous localization and mapping (V-SLAM) algorithm to localize the UAV without GNSS. Systems presented above all depend on visual features of the bridge to estimate the position of the UAV. However, it is well known that image processing could easily deteriorate with varying illumination [21]. Apart from this, some specific bridge inspections only perform at night to avoid disruption to civil commuting in the daytime. In such a case, the visual features of the bridge are actually not sufficient in a low illumination environment. To cope with this problem, utilizing artificial features or markers is a promising approach [22]. As placing stationary markers along the whole bridge is laborious, a mobile ground–air system, where the guided vehicle is equipped with specially designed features, could provide a flexible positioning for the UAV.

Accumulated works attempt to address the UAV positioning problem with designed features in ground–air systems. Patruno *et al.* [23] proposed a vision-based helipad detection algorithm to estimate the pose of the UAV. Cantieri *et al.* [24] deployed an augmented reality tag (ARTag) on the ground vehicle (GV) to provide positioning for the UAV in power-line inspection. As summarized in the benchmark of current fiducial markers [25], ARTag, AprilTag, ArUco, and helipad all acquire the anchor position by detecting corners of the contour. However, the corner feature is not clear in case of low illumination. In order to make the anchor points of the marker to be clearly detected, Rudol deploys a cube of LEDs serving as the anchor points of the marker for ground–air cooperative

Manuscript received September 4, 2021; revised November 3, 2021; accepted November 21, 2021. Date of publication December 14, 2021; date of current version February 21, 2022. This work was supported by the Ministry of Science and Technology of the People's Republic of China and in part by the National Natural Science Foundation of China. The Associate Editor coordinating the review process was Dr. Shovan Barma. *(Corresponding author: Wei Dong.)*

The authors are with the State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: wangzhaoying@sjtu.edu.cn; sensenliu@sjtu.edu.cn; chg947089399@sjtu.edu.cn; dr.dongwei@sjtu.edu.cn).

Digital Object Identifier 10.1109/TIM.2021.3135544

indoor exploration [26]. With the extracted anchor points, the UAV could robustly estimate a relative position through a perspective-n-point (PnP) algorithm. The LEDs are distinctive in the dim environment but would deteriorate under high intensity of illumination. To tackle this problem, Li *et al.* [27] utilized an infrared camera and four reflective markers to make the marker distinct from the high intensity of illumination of the environment. However, the simple infrared feature is sensitive to light noise in the background.

In summary, the previous marker design and detection algorithm can be divided into two types. The first type depends on corner detection for positioning, including ARTag, AprilTag, and ArUco. The advantage is robustness to light noises because their coded pattern provides a candidate detecting region. The drawback is corner detection failure under low illumination. On the contrary, the second type utilizes independent and distinct marker points, including LED lights, reflective markers, tennis balls [28], and infrared boards [29]. The advantage is that the marker is distinct under various illumination intensities. The drawback is sensitivity to light noises. Due to no candidate searching region provided, the light noise could easily mix with these independent markers, such as the noises from sunlight or the car headlight. Thus, it can be concluded that the current markers can perform well in one-sided condition, i.e., varying illumination or complex light noises. However, in the practical bridge inspection at night, the light condition includes both various illumination intensities and complex light noises. The combination of lights in the environment makes the visual positioning of the UAV extremely difficult.

To well handle the problem, in this article, we propose a dual-source positioning method to provide robust marker detection and visual positioning for the UAV under the bridge. First, we design an observation model to optimize the spatial position of the GV and the UAV. Second, a marker detection algorithm from coarse to fine filtering based on the dual-source images, i.e., color image and infrared image of the downward camera on the UAV, is proposed. Third, the pose of the UAV is estimated by the visual position and inertial measurement unit (IMU) compensation. Finally, feedback control of the UAV is enhanced by a complementary filter (CF) of the velocity during the flight under the bridge. To inspect the defect area on the bridge bottom, an upward camera on the UAV continuously records the bottom surface when the car leads the UAV. To demonstrate the effectiveness of the proposed system, we perform real-world bridge inspection experiments under different bridges with our ground–air mobile system. Various experiments are also conducted to verify the robustness of our positioning method under varying illumination and complex light noises. In addition, the precision of positioning is evaluated with ground truth, which is provided by a global positioning system with real-time kinematic (GPS-RTK). We summarize our contributions as follows.

1) We first design a ground–air system for bridge inspection. In the GNSS-denied environment, the GV can easily lead the UAV to cover the inspection area under the bridge, where the UAV can robustly stabilize itself by observing the visual fiducials on the GV and capture the defect image of the bridge bottom surface.
2) We propose a robust visual positioning method utilizing dual-source images for the UAV to estimate its relative position from the GV. Our method can resist varying illumination and complex light noises.
3) Various real-world experiments are carried out to validate the precision and robustness of our approach. We compare with the other state-of-the-art positioning methods, including V-SLAM and popular artificial markers to demonstrate the superiority of our dual-source positioning method.

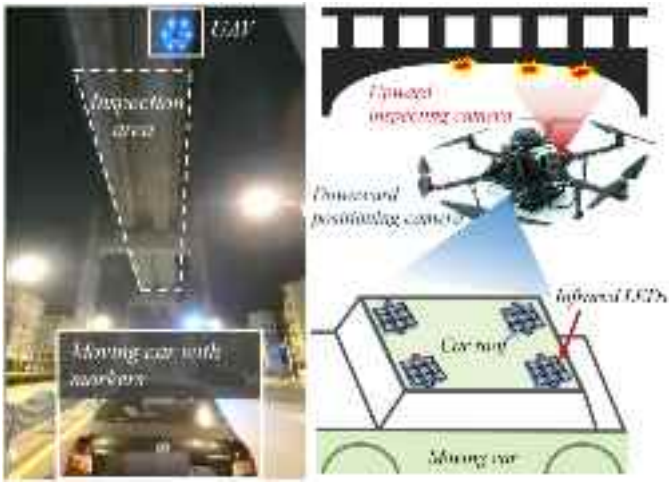

Fig. 1. Architecture of our ground–air system for inspection of the bottom of the bridge.

II. System Architecture

The architecture of our ground–air system is shown in Fig. 1. The UAV can acquire its position by observing the infrared markers attached to the roof of the GV. Meanwhile, the upward camera of the UAV can capture the bottom of the bridge to execute the visual inspection. A car is used as the GV in our experiment. Two inspection modes are designed. The first mode is area coverage, where the car leads the UAV flying at a speed of 5 m/s under the bridge. The UAV continuously records the bottom surface of the bridge for offline analysis. The second mode is specific point inspection. The car is stationary at a specific point. The UAV can hover and record the area of interest. The detailed hardware descriptions are as follows.

A. Unmanned Aerial Vehicle

The UAV is a custom hexacopter with an axle base of 550 mm. An Intel RealSense D455 camera is fixed in front of the UAV with a diagonally down 45° view for visual positioning. This camera can synchronously acquire infrared and color video streams at 30 Hz. A 4k USB camera is installed on the top of the UAV to capture images of the bottom surface of the bridge. All vision process and motion planning are running with robot operating system (ROS) on the Up Xtreme computing board with Intel i7-8665UE CPU

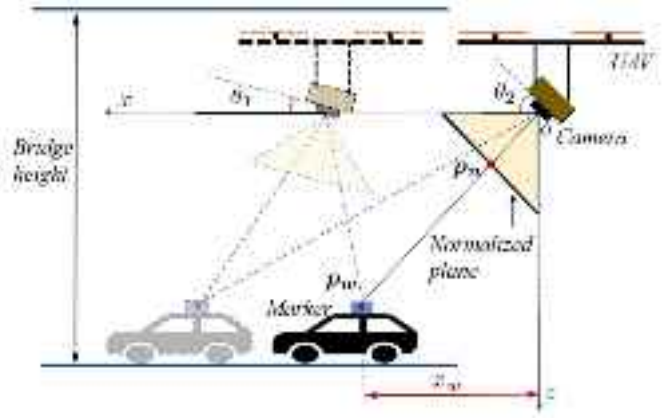

Fig. 2. Spatial relationship of the camera and the UAV as well as the marker.

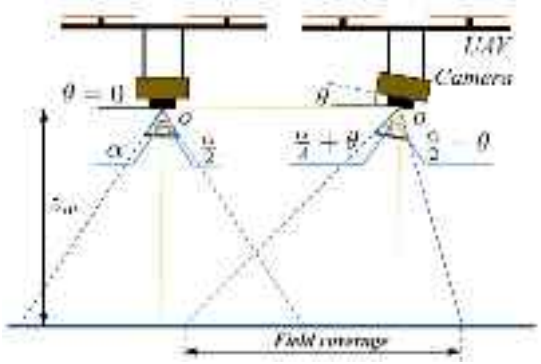

Fig. 3. Field coverage of the observation with different installation angles.

and Intel Movidius Myriad X 2485. The flight is controlled by Pixhawk 4 powered by PX4 autopilot.

B. Ground Platform

Our ground platform consists of a car and infrared markers. The car can provide a 500 mm × 500 mm flat roof for marker installation. The infrared markers are four independent infrared LEDs with an 850-nm wavelength. The size of each marker is of 60 mm × 60 mm, which is convenient to be placed on the car. In addition, our system has no strict requirement of the car color. Different colors of car roof are tested, including blue, brown, black, and white. The extensibility is another advantage. Due to the flexibility of our marker, it can easily be transplanted to other ground, air, and water vehicles to provide the visual fiducials for the UAV in various scenarios.

III. Methodology

Our ground–air system will be introduced in three parts. The first part analyzes the spatial relationship of the GV and the UAV and proposes an observation model to optimize their spatial position and observation angle. With the optimized spatial arrangement, the second part introduces the visual positioning process. We propose a dual-source positioning algorithm that can robustly acquire the relative position between the GV and the UAV. With the position feedback, the third part presents the state fusion estimation of the UAV and the system control.

A. Spatial Relationship

The UAV obtains its position by observing the marker attached to the GV. Thus, designing a spatial relationship to achieve a robust and accurate observation is crucial. In our model, the camera is fixed on the UAV and the field of view (FOV) of the camera is limited. The installation angle of the camera as well as the relative position between the marker and the UAV are the variables to be optimized. First, the installation angle of the camera is expected to be set to cover a larger area so that the camera can contain the marker in the view robustly. Second, the relative position between the marker and the UAV should coordinate the installation angle to achieve a high observation accuracy. Different installation angles of the camera and relative positions are shown in Fig. 2.

We build an observation model to optimize the spatial relationship in the ground–air system. The model contains three parts, which considers the field coverage of the camera, the observation accuracy, and the attitude disturbance of the UAV.

First, we discuss the field coverage of the camera. θ is the installation angle of the camera relative to the UAV. The FOV is depicted as a yellow triangle area. In our case, the angle of FOV is α, which is constant. z_w is the height of the camera from the ground plane, which depends on bridge height. For a specific bridge inspection mission, the bridge height is known and normally constant. The field coverage F_c can be calculated as (1). The illustration is shown in Fig. 3

$$F_c = (\mathrm{z}_w(\tan(\alpha/2-\theta)+\tan(\alpha/2+\theta))). \tag{1}$$

In order to keep the marker in the FOV of observation, the FOV is expected as large as possible. However, when $\alpha/2+\theta$ reaches $\pi/2$, F_c will reach positive infinity. We should avoid this condition, because in such a case, the remote location in FOV is actually not valid. F_c cannot reflect the effective observation field. Only the close location in FOV can be effectively observed in the real experiment. Thus, we add a log function on F_c to suppress the assessment of field coverage. The value function v_c of field coverage is designed as

$$v_c = \log(\mathrm{z}_w(\tan(\alpha/2-\theta)+\tan(\alpha/2+\theta))). \tag{2}$$

Next, we analyze the observation accuracy of the camera. First of all, we define variation sensitivity v_s. The variation of the marker along the x-axis in world coordinate is defined as Δx_w. The corresponding variation projected in the normalized plane of the camera is Δx_n. v_s can be expressed by the following equation.

$$v_s = \Delta x_n / \Delta x_w. \tag{3}$$

The variation sensitivity v_s is used to measure the observation accuracy. The higher v_s represents the higher observation accuracy. Then, we further infer v_s from the camera model. $\boldsymbol{p}_w = [\mathrm{x}_w, y_w, z_w]$ represents the position of the marker in world coordinate. $\boldsymbol{p}_c = [x_c, y_c, z_c]$ represents the position of the marker in the camera coordinate. θ represents the rotation of the camera in the xoz plane. We get the conversion between $\boldsymbol{p}_w$ and $\boldsymbol{p}_c$ as

$$\boldsymbol{p}_c = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ 0 & 1 \end{bmatrix} \boldsymbol{p}_w. \tag{4}$$

$\boldsymbol{R}$ and $\boldsymbol{t}$ describe the Euclidean transformation of the camera coordinate in the world coordinate. Specifically, $\boldsymbol{R}$ is the rotation matrix and $\boldsymbol{t}$ is the translation vector. To analyze the movement of projection of $\boldsymbol{p}_w$ in the image plane, we use $\boldsymbol{p}_n$ to represent the projection of the marker in the normalized plane of the camera, which has the following conversion:

$$\boldsymbol{p}_n = [x_n, y_n, 1] = [x_c/z_c, y_c/z_c, 1]. \tag{5}$$

Then, we get the conversion from $\boldsymbol{p}_w$ to $\boldsymbol{p}_n$ in the x-axis as

$$x_n = \frac{x_w \cos\theta - z_w \sin\theta}{x_w \sin\theta + z_w \cos\theta}. \tag{6}$$

The variation sensitivity v_s can be formulated as

$$v_s = \frac{\Delta x_n}{\Delta x_w} = \frac{z_w}{(x_w \sin\theta + z_w \cos\theta)^2}. \tag{7}$$

Since z_w is constant as mentioned before, x_w and θ jointly decide v_s.

The third part analyzes the observation performance influenced by the attitude disturbance of the UAV. Since the camera is fixed on the UAV and the attitude of the UAV continuously adjusts during the flight, the variation of θ will cause the position of $\boldsymbol{p}_n$ fluctuating. Since large fluctuation would lead to the marker leaving the field of vision, the fluctuation sensitivity v_e is expected to be small. The detailed form is presented as follows:

$$v_e = \frac{\Delta x_n}{\Delta \theta} = -\frac{x_w^2 + z_w^2}{(x_w \cos\theta + z_w \sin\theta)^2}. \tag{8}$$

Considering the three parts above, the final value function can be expressed as

$$v = \lambda_1 v_c + \lambda_2 v_s + \lambda_3 v_e. \tag{9}$$

We choose $[\lambda_1 : \lambda_2 : \lambda_3] = [0.3, 0.2, 0.5]$ as weight coefficient.

The value function of the observation model is plotted in Fig. 4. In our case, z_w is set as 4 m. The maximum of the value function is achieved at $x_w = 4$ and $\theta = \pi/4$.

Actually, z_w is chosen according to the different heights of the bridge. The observation model can assist in designing other parameters x_w and θ to construct an optimized spatial relationship in a ground–air system.

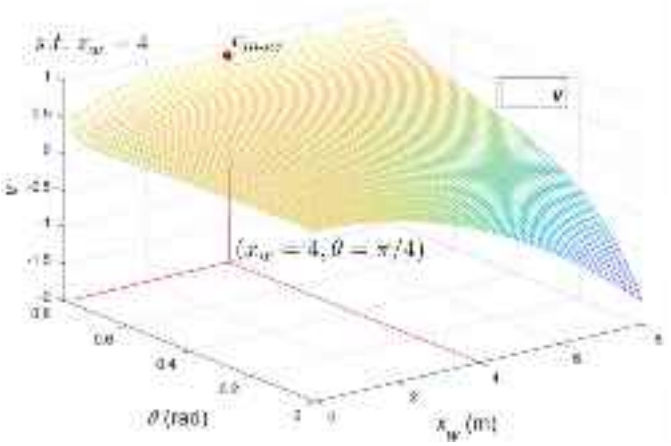

Fig. 4. Value function of the observation model.

B. Visual Processing

Our infrared markers are attached to the roof of the GV. To robustly estimate the position of the UAV by observing the markers, a dual-source positioning model is introduced in our vision positioning method. The model simultaneously processes dual image sources, i.e., the color image and the infrared image. We first explain the motivation of using dual image sources. Because we adopt a coarse-to-fine strategy to search the marker points, we divide our algorithm into two steps. The first step is to find the car roof, and the second step is to further search the infrared markers. Since the color image can provide rich features of color and structures while the infrared image is usually featureless due to its sensitive spectral range, the color and shape features of the car roof are easy to be detected in the color image. However, further searching the marker point cannot depend on the color image because varying illumination makes the corner and brightness feature not obvious to be robustly detected. On the contrary, infrared lights are always obvious in the infrared image under varying illumination. Thus, we adopt the color image to search candidate marker area and then the infrared image to further detect the infrared marker for pose estimation. It is worth noting that light noises, such as taillights, headlights, and floodlights, are easy to confuse with the infrared marker. Directly detecting infrared markers is not robust. A candidate region given by color image is indispensable. The dual-source positioning algorithm can be divided into two parts; the region proposal and the marker detection. The diagram of the model can be viewed in Fig. 5.

1) Region Proposal: First, we will expound on the region proposal module. As mentioned before, color and shape are the main features that can be used to detect the car roof. However, the color segmentation or contours extraction of shape cannot resolve this issue under challenging illumination. For example, the red or blue car looks similar to the black car under low illumination. The contours of the car roof are also vague under dim lights. As deep convolution neural networks could extract more features with sufficient datasets [30], we choose the object detection method to detect the candidate marker area. Popular object detection approaches can be divided into one-stage mode and two-stage mode. The advantage of the one-stage model is fast, while that of the two-stage model is accurate. Since real-time performance is more concerned in our algorithm, the one-stage object detection method is preferred. Through our test, YOLOv3 [31] outperforms other methods in mAP and inference time. We finally choose YOLOv3 to train our model with collected image data. The training data include 1000 pictures of the car roof at different altitudes and different light conditions. With 480 × 480 input size and 5000 iterations, a trained YOLOv3-tiny weight is finally used to inference the car roof, which is the candidate region of markers. The detailed process is as follows.

The region proposal module acquires synchronous color and infrared video stream as input. A pretrained YOLOv3 network is loaded to infer the candidate image region of the car roof in the color image. The proposed region of the color image is

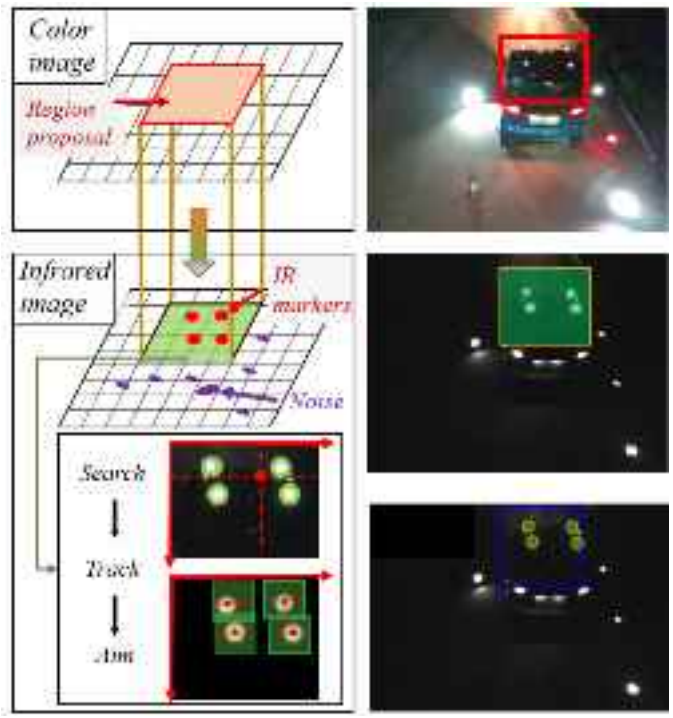

Fig. 5. Diagram of the dual-source positioning model with the corresponding image process.

defined as the color region S_C. The outside of the proposed region is defined as $\bar{S}_C$. Most light noises, such as car lights and street lamp lights, are located in $\bar{S}_C$. The proposed region S_C can effectively filter out the light noise. Since the later marker point detection process is performed in the infrared image, the last step of the region proposal is processing the dual image source. Specifically, S_C is projected to the aligned infrared image to generate a candidate region S_I at the same location

$$S_I = \text{Proj}_I\, S_C. \tag{10}$$

It is worth noting that the region proposal by the YOLOv3 algorithm is easily extensible for other platforms with sufficient training data.

2) Marker Detection: The marker detection module aims to accurately extract the infrared LED points. Three strategies, including searching target $\mathcal{ST}$, tracking target $\mathcal{TT}$, and aiming target $\mathcal{AT}$, are applied in different stages.

Searching target $\mathcal{ST}$ is deployed when the marker detection module initializes. The goal of $\mathcal{ST}$ is to search the pixel position of markers in S_I and to label these pixels with serial numbers. Since the wavelength of infrared markers matches the infrared-pass filter of the camera, markers are bright in the infrared image. Thus, a threshold function is applied to search markers.

Yan *et al.* [29] adopted direct binarization of full image with a constant threshold. To handle a more challenging illumination in our situation, infrared markers are detected by a dynamic thresholding function in the proposed region S_I. T represents the threshold in the binarization process.

The desired number of contours in the binary image is four. However, due to varying illumination, some light noises may mix in the desired contours. T is adjusted according to the actual number of the contours C_n

$$T = \begin{cases} T + \Delta T, & C_n > 4 \\ T - \Delta T, & C_n < 4 \end{cases}. \tag{11}$$

$\tilde{T}$ is the initial value, which is adjusted by ΔT. In our case, $\tilde{T} = 140$ and $\Delta T = 5$. Next, a virtual geometry center of four markers is calculated. Four quadrants are generated according to the geometry center. Finally, the pixel of each marker is labeled as $p_i, i = 1, 2, 3,$ and 4 according to its location in the quadrantal diagram.

The tracking target $\mathcal{TT}$ process follows the searching target $\mathcal{ST}$, which is deployed to track the marker in the next frame. We suppose that the brightness of p_i keeps constant between the consecutive infrared frames, which satisfies the brightness constancy constraint

$$I_x u + I_y v + I_t = 0. \tag{12}$$

We could minimize the sum of the squared differences (SSD) to find pixel displacement as follows:

$$E_{\text{SSD}}(u) = \sum_i [I_1(x_i + u) - I_0(x_i)]^2. \tag{13}$$

The common method in this SSD function is gradient descent, which is proposed by Lucas and Kanade [32]. Take the Taylor series approximation of image function

$$I(x, y, t) + I_x u + I_y v + I_t = I(x + dx, y + dy, t + dt). \tag{14}$$

Then, u and v can be calculated by the normal equation

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum I_x I_t \\ -\sum I_y I_t \end{bmatrix}. \tag{15}$$

Finally, p_i in the current frame will be tracked to a new p_i in the next frame. The aiming target $\mathcal{AT}$ is used after the tracking target $\mathcal{TT}$. Since the residual error $e_i = I_1(x_i + u) - I_0(x_i)$ in the least square fit of L-K optical flow cannot minimize to zero, p_i produced by optical flow is not exactly the center of the marker point. Specifically, each marker projected on the image is a small bright block, and each block contains $n \times n$ pixels. p_i would be included in the block but normally not exact the center of the block. Thus, the calculation of the centroid of the tracked bright block is necessary. In a binary image, all the pixels around p_i are defined as $M_{n\times n}$. The pixel with a gray value of 255 in $M_{n\times n}$ should be summed up to recalculate the center of the bright block. The new center is defined as p_i'

$$p_i' = \frac{1}{N} \sum (u, v) \tag{16}$$

where $I(u, v) = 255$, $(u, v) \subseteq M_{n\times n}$, and N is the number of the pixels that satisfy the above condition.

Finally, with the knowledge of the structure of the infrared marker points, we use the P3P algorithm to calculate the relative position between the onboard camera and the markers

$$\boldsymbol{p}_c = F_{\text{P3P}}(\boldsymbol{p}_w, p_i). \tag{17}$$

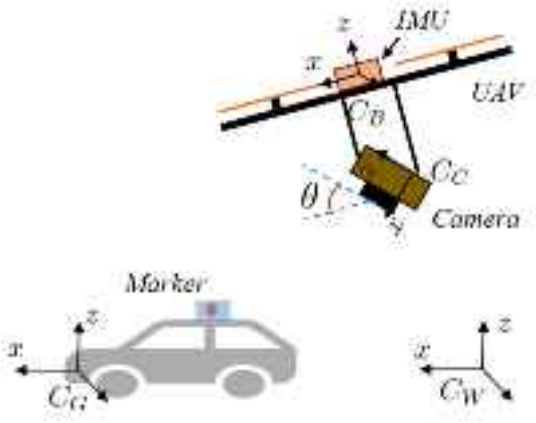

Fig. 6. Coordinates of the ground–air system.

C. State Fusion and Control

This section mainly discusses the state fusion of the UAV using visual position and onboard IMU. The system control policy is also briefly described.

First, the position estimation of the UAV with an IMU assistant will be discussed. Since our visual positioning camera is fixed on the UAV, the attitude fluctuation of the UAV will lead to the displacement of the pixel position of the marker in the image. The relative position estimated by the P3P algorithm will also change. Actually, the position of the UAV estimation should not be influenced by the attitude of the UAV. Thus, the pose measured by the onboard IMU needs to be fused for attitude compensation.

Several coordinates are presented in Fig. 6 at first. We define the coordinate of the GV as coordinate C_G. Since markers are fixed on the roof of the GV, we regard them as a rigid body. The camera coordinate is defined as C_C, with a θ installation angle between the UAV coordinate. The IMU within the flight controller (FC) is parallel to the body of the UAV. Thus, both the IMU and the UAV body are defined as coordinate C_B. Also, world coordinate is defined as coordinate C_W. All coordinates follow the east–north–up (ENU) orientation rule.

$\boldsymbol{T}_{GW}$ is the transform matrix from the GV coordinate C_G to the world coordinate C_W. $\boldsymbol{T}_{BW}$ is the transform matrix from the UAV coordinate C_B to the world coordinate C_W. $\boldsymbol{T}_{CB}$ is the transform matrix from the camera coordinate C_C to UAV coordinate C_B. Since the installation angle θ is fixed, the transform matrix $\boldsymbol{T}_{CB}$ is constant. Onboard IMU can measure the pose and provide a real-time $\boldsymbol{T}_{BW}$ transform matrix. The relative position obtained by the camera is defined as $\boldsymbol{p}_c$. $\boldsymbol{p}_g$ is the position of the GV. $\boldsymbol{p}_b$ is the position of body center of the UAV. The estimation of relative position between the GV and the UAV is defined as follows:

$$\boldsymbol{p}_r = \boldsymbol{p}_g - \boldsymbol{p}_b = {\boldsymbol{T}_{GW}}^{-1} \cdot \boldsymbol{T}_{BW} \cdot \boldsymbol{T}_{CB} \cdot \boldsymbol{p}_c. \tag{18}$$

Second, the estimation of velocity of the UAV is accomplished by a CF as

$$\dot{\boldsymbol{p}}_r(k+1) = w_1 \cdot \frac{d\,\boldsymbol{p}_r(k)}{dt} + w_2 \cdot \left(\dot{\boldsymbol{p}}_r(k) + \boldsymbol{a} \cdot dt\right). \tag{19}$$

w_1 and w_2 are set to adjust weight between position $\boldsymbol{p}_r$ differentiation and acceleration a integration for best velocity estimation. In our experiment, $w_1 : w_2 = 1 : 1$ achieves both smooth and low-latency estimation.

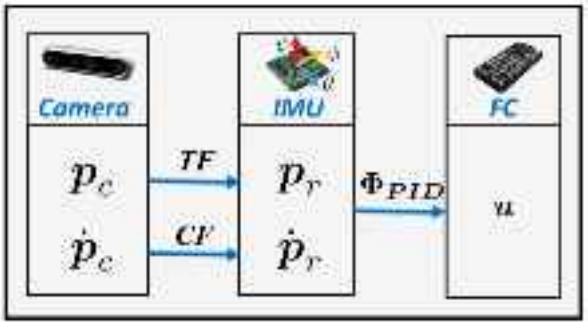

Fig. 7. Flowchart of the state fusion and control.

Fig. 8. UAV executes an inspection mission under the highway bridge. (a) Color and infrared images captured by the D455 camera. (b) UAV equipped with two types of cameras. (c) Captured image containing the defect point on the bridge.

In the end, we discuss the control policy in the bridge inspection mission. The UAV needs to keep a constant relative position from the GV, which can be expressed as

$$\boldsymbol{u} = \boldsymbol{\Phi}\left(\boldsymbol{p}_r, \dot{\boldsymbol{p}}_r, \boldsymbol{a}\right) \tag{20}$$

where $\boldsymbol{u}$ is the control input of the UAV. In our case, a cascade proportional–integral–differential (PID) control law is adopted for position and velocity control. The attitude control is accomplished by PX4 autopilot.

We summarize our method as a control block diagram in Fig. 7. The position $\boldsymbol{p}_c$ in the camera coordinate is transformed (TF) to position $\boldsymbol{p}_r$ in world coordinate with the angle measurement by onboard IMU. The differential position (velocity) $\dot{\boldsymbol{p}}_c$ in the camera coordinate is fused with the acceleration of IMU by a CF to obtain velocity $\dot{\boldsymbol{p}}_r$ in the world coordinate. Finally, the fused state $[\boldsymbol{p}_r, \dot{\boldsymbol{p}}_r]$ is feedback to the PID controller $\boldsymbol{\Phi}_{\text{PID}}$ to output the control u for FC executing stable flight.

IV. Experiment and Discussion

To validate the effectiveness of our ground–air system in bridge inspection, experiments are performed under a highway bridge, which is shown in Fig. 8. Intel RealSense D455 is used as the downward positioning camera. The color and infrared images of the GV are synchronously captured at 30 Hz to run the dual-source positioning algorithm. Real-time state fusion and PID control support a stable flight. When the GV leads the UAV flying under the bridge, a USB camera on the top of the UAV continuously captures the defect point on the bottom of the bridge. In our ground–air system, the effective distance is 1–8 m, and the horizontal and the vertical angle support $[-\pi/4, \pi/4]$ and $[\pi/6, \pi/3]$, respectively.

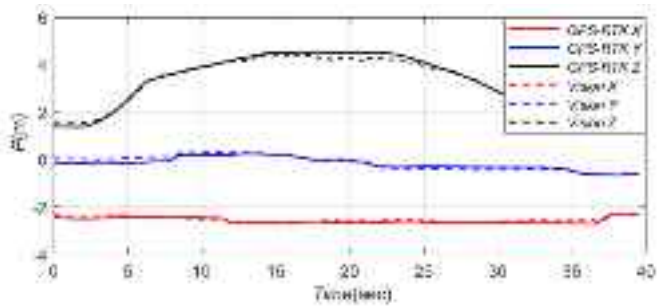

Fig. 9. Position estimation by our dual-source positioning method compared with GPS-RTK.

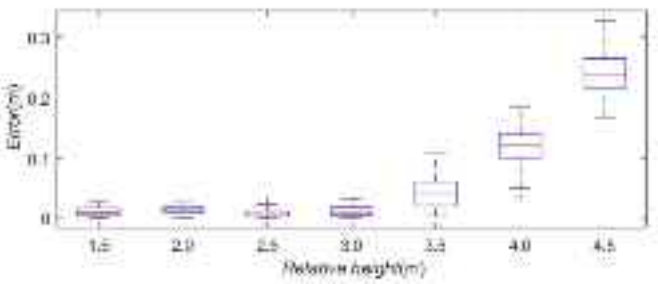

Fig. 10. Boxplot of the position error concerning the relative height between the GV and the UAV.

TABLE I
ERROR ANALYSIS

Axis	x(m)	y(m)	z(m)
RMSE	0.062	0.104	0.111
MAE	0.591	0.086	0.084

A. Precision Analysis

In order to evaluate the accuracy of the visual positioning, the position estimated by our dual-source positioning is compared with GPS-RTK. We perform our experiment under an open sky, where the accuracy of GPS-RTK is about 1 cm. Both the car and the UAV are equipped with GPS-RTK to acquire their ground truth. The UAV is tested by hovering at different altitudes, aiming to adapt to different heights of the bridges. The relative position between the marker and the UAV estimated by our method and the GPS-RTK is plotted in Fig. 9. The error analysis is shown in Table I.

Since the observation accuracy is influenced by the observation distance, we show the position accuracy analysis with different heights in Fig. 10. The accuracy of position estimation declines with the distance. With the relative height growing, the position error (comparing with GPS-RTK) keeps below 0.04 m within 3-m height (1.3% relative error). However, the error rises to 0.12 m when the relative height exceeds 4 m (3% relative error). Finally, the position error climbs to 0.24 m when the height is 4.5 m (5.3% relative error).

The growing error validates the variation sensitivity V_s part in the observation model. V_s declines with the z_w increasing [see (7)]. The observation accuracy is also influenced by the camera and IMU. Specifically, the higher image resolution will increase the location precision of the detected markers on the image. According to the pinhole camera model, the marker position estimation will also be improved. The precision is proportional to the image resolution. With a camera resolution of 640 $\times$ 480 in our experiment, the theoretical precision is 0.0015 m. Since the IMU is used to transform the marker position in the camera coordinate to UAV coordinate, the attitude precision of IMU also affects the accuracy. In our experiment, the Pixhawk 4 controller (containing ICM-20 689 IMU) is used to estimate attitude. According to the technical parameters, the angle error in pitch θ is 0.01 rad. The position error can be calculated with the small-angle hypothesis. $e = D\sin(\delta\theta) \approx D\delta\theta$. With the observation distance $D = 5$, position precision e is 0.05 m. In our real flight experiment, the mechanical vibration of the propeller increases the noise on the camera and IMU. The observation accuracy can be improved by reducing the vibration in future work.

B. Robustness Analysis

In order to guarantee an effective inspection under the bridge, a robust visual positioning is essential in a light challenging environment. Robustness is defined as correctly extracting the predefined features in a challenging light condition during various flight conditions and GV conditions. The challenging light condition includes high or low environmental illumination intensity as well as complex light noise in the background. The flight condition is the observing position of the UAV. The GV condition includes the appearance of the car, driving situation, and road condition. We present these challenging conditions in three parts, including various light conditions, difficult observation positions, and complex ground scenarios. The experiments about various light conditions are shown in Fig. 11.

In the daytime, the strong sunlight brings more difficulties than the weak sunlight because the infrared markers in strong sunlight are not obvious in the background. The binarization with a dynamic threshold in our algorithm tackles this issue. Markers can be successfully detected under high illumination intensity. On the contrary, at night, the low light intensity poses more challenges to the detection of the marker area than the high light intensity. Due to the limited illumination, a general method by the contour or the color segmentation may fail. Nevertheless, the YOLOv3 network could guarantee a more than 60% true positive detection rate of the car roof with sufficient training data. Then, the following marker detection process could be successfully performed in the candidate area without mismatching other light noises. The complex light condition is difficult too. The optical flow of the marker may mismatch with floodlights on the ground. In this case, our region proposal module could help to avoid this problem by checking the candidate region.

The robustness of the visual positioning is also influenced by the observation position. Fig. 12 shows two difficult positions. The target detection at a long distance is an admitted challenge. Detecting small objects is also the weakness of the YOLOv3 algorithm. However, with the help of the continuous tracking of the marker point by optical flow from the close distance to long distance, the detection problem can be resolved

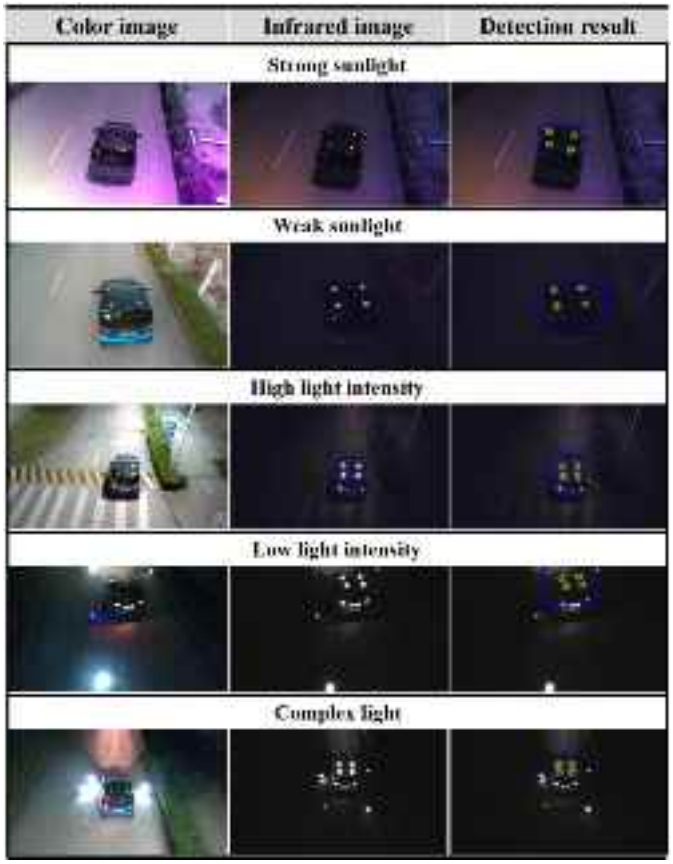

Fig. 11. Visual positioning experiments under various light illumination and light noises.

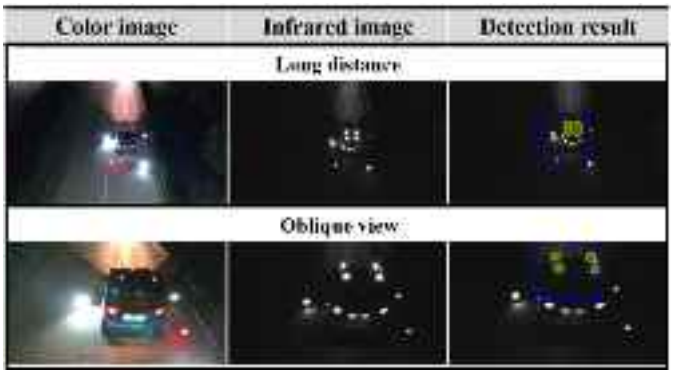

Fig. 12. Visual positioning experiments at different observing positions.

to some extent. The visual positioning with the oblique view is also difficult. Though the dual-source positioning algorithm can track the quick movement of the marker, the fluctuation of the attitude of the UAV may cause the marker to drift out of the view. We try to avoid this situation by adopting an optimized spatial relationship calculated by the observation model.

The GV condition also affects the robustness of visual positioning. Fig. 13 presents the experiment under the various GV conditions. The car turning condition shows the robustness of our algorithm to the rotation of markers. Therefore, if the car's movement is not straight, such as s-driving or turning direction, the UAV can also robustly track the marker and estimate its position. In order to keep the car in the view of

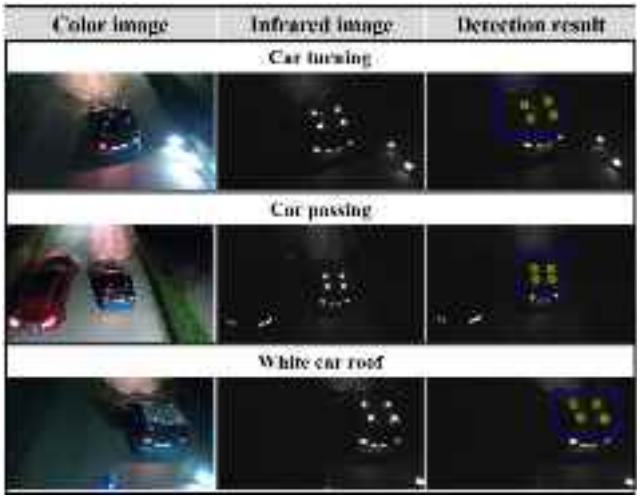

Fig. 13. Visual positioning experiments under various GV conditions.

the UAV's camera, the UAV will execute the same trajectory as the car. In the car passing condition, the tail light or the similar car roof may mix our target car. Though both car roofs may be detected, only the target car roof with infrared markers will continue the marker detection process.

In order to verify the robustness of the algorithm to vehicles with different colors, we decorate the black roof with a white-board. With sufficient training data, our visual positioning algorithm works well. We also successfully test our algorithm on blue and brown cars, which can be viewed in the attached video.

Therefore, varying light conditions, different observing positions, and complex ground conditions all demonstrate the robustness of the dual-source positioning algorithm.

Actually, under sufficient illumination conditions, the UAV can follow any car with nothing installed on the top. The size and the position of the bounding box in the image inferred by YOLOv3 can be utilized to estimate the position of the car coarsely. However, the recall ratio of the detection under varying illuminations is only about 60%. The robustness of observation will be dramatically reduced. In contrast, with the easy attachment of infrared markers, the dual-source positioning algorithm can utilize the information from the car roof and infrared markers. The robustness of observation can be greatly strengthened.

C. Comparison With Benchmark

To demonstrate the superiority of robustness of our method in the bridge inspection mission, we compare other state-of-the-art methods for the robust localization of the UAV. First, we compare the V-SLAM method. One of the advanced V-SLAM systems is Intel RealSense Tracking Camera T265, which Intel developed. T265 can run a V-SLAM algorithm and output the current position of the camera from the home point. In the experiment, we install T265 on the UAV for the localization of the UAV. To evaluate the robustness of localization of UAV, the ground truth is provided by GPS-RTK for comparison. The position estimation by T265 compared with GPS-RTK is plotted in Fig. 14. From the figure, the

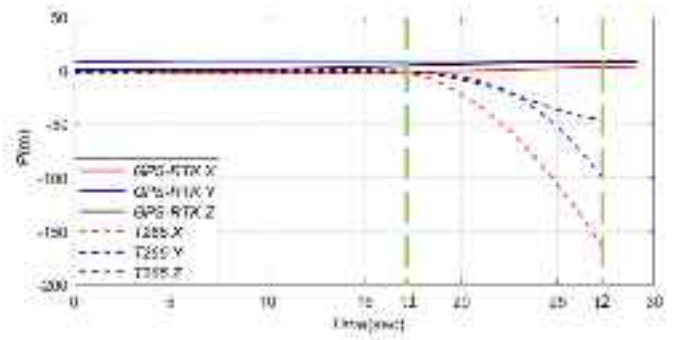

Fig. 14. Position estimation of V-SLAM by T265 and GPS-RTK.

(a) (b)

Fig. 15. (a) Image from T265 camera. (b) Marker positioning experiment with ARTag.

position of T265 is almost consistent with GPS-RTK before $t1$ (17 s). After that, the position of T265 drifts to a negative value and becomes abnormal. After $t2$ (27 s), the position value is totally invalid so that the T265 cannot output its estimation. The main reason for failure may be the adverse environment. The image captured by T265 is shown in Fig. 15(a). The image contains too few environmental features, and then, the T265 cannot robustly run its visual SLAM algorithm to estimate the position. Thus, the V-SLAM localization method by T265 is not robust in the bridge inspection experiment.

As ARTag is a popular approach for relative position estimation, Cantieri *et al.* [24] presented a master–slave positioning system, where an ARTag is placed on the top of the GV to provide positioning for the UAV in powerline inspection. Their experiment is conducted in the daytime with good illuminations to guarantee the effective visual connection between the master and the slave. Since our system needs to execute under bridge inspection with a challenging light environment (varying illumination and complex light noises), the robustness of observation from the slave to the master is the main concern. We test the observation robustness of the ARTag in the challenging light condition. Fig. 15(b) shows an ARTag with 1 m $\times$ 1 m size placed on the car roof. Due to the low illumination and complex light noises, we fail to detect the ARTag board continuously in several trials. Thus, the position estimation from ARTag is not robust in this scenario. In contrast, the dual-source positioning algorithm proposed in our system works well in this condition.

V. Conclusion

In this work, a ground–air system is developed to execute the under bridge inspection. To fulfill this mission, an observation model is first designed to optimize the spatial relationship between the GV and the UAV. A dual-source positioning algorithm is then developed to enhance the aerial visual positioning performance under the environment with different illumination and complex light noises. Finally, the visual position and IMU pose are fused to obtain the relative position for the UAV feedback control. Extensive experiments are conducted to verify the precision and the robustness performance of our visual positioning method of UAV. The comparisons with the advanced V-SLAM system of T265 and the current marker methods prove the superiority of our method in robustness under various illuminations. All the results demonstrate that the proposed ground–air system and related algorithms can provide a robust and precise position for the UAV in the bridge inspection.

In the future, we plan to extend our relative positioning method to other multirobot systems for bridge inspection. For example, the UAV under the bridge can estimate its position by observing another UAV (with GNSS) out of the bridge. A similar case can be applied to the boat and the UAV system for bridges above the river. For the safe flight, an obstacle avoidance system with a laser range finder in our previous bridge inspection work [11] will be deployed in the UAV.

References

[1] V. Kumar and N. Michael, "Opportunities and challenges with autonomous micro aerial vehicles," *Int. J. Robot. Res.*, vol. 31, no. 11, pp. 1279–1291, 2012.
[2] L. Duque, J. Seo, and J. Wacker, "Synthesis of unmanned aerial vehicle applications for infrastructures," *J. Perform. Constructed Facilities*, vol. 32, no. 4, Aug. 2018, Art. no. 04018046.
[3] S. Jung *et al.*, "Toward autonomous bridge inspection: A framework and experimental results," in *Proc. 16th Int. Conf. Ubiquitous Robots (UR)*, Jun. 2019, pp. 208–211.
[4] R. R. Murphy *et al.*, "Robot-assisted bridge inspection," *J. Intell. Robotic Syst.*, vol. 64, no. 1, pp. 77–95, Oct. 2011.
[5] B. Chan, H. Guan, J. Jo, and M. Blumenstein, "Towards UAV-based bridge inspection systems: A review and an application perspective," *Struct. Monitor. Maintenance*, vol. 2, no. 3, pp. 283–300, Sep. 2015.
[6] B. Lovelace, "Unmanned aerial vehicle bridge inspection demonstration project," Minnesota Dept. Transp., Oakdale, MN, USA, Tech. Rep. MN/RC 2015-40, 2015, pp. 1–44.
[7] D. Droeschel, M. Nieuwenhuisen, M. Beul, D. Holz, J. Stuckler, and S. Behnke, "Multilayered mapping and navigation for autonomous micro aerial vehicles," *J. Field Robot.*, vol. 33, no. 4, pp. 451–475, Jun. 2016.
[8] H.-J. Jung, J.-H. Lee, and I.-H. Kim, "Challenging issues and solutions of bridge inspection technology using unmanned aerial vehicles," *Proc. SPIE*, vol. 10598, Mar. 2018, Art. no. 1059802.
[9] A. Chen, D. Frangopol, X. Ruan, N. Hallermann, and G. Morgenthal, *Visual Inspection Strategies for Large Bridges Using Unmanned Aerial Vehicles (UAV)*. Boca Raton, FL, USA: CRC Press, 2014, pp. 661–667.
[10] S. Chen, D. F. Laefer, E. Mangina, S. M. I. Zolanvari, and J. Byrne, "UAV bridge inspection through evaluated 3D reconstructions," *J. Bridge Eng.*, vol. 24, no. 4, Apr. 2019, Art. no. 05019001.
[11] J. Chen, J. Wu, G. Chen, W. Dong, and X. Sheng, "Design and development of a multi-rotor unmanned aerial vehicle system for bridge inspection," in *Proc. Int. Conf. Intell. Robot. Appl.*, 2016, pp. 498–510.
[12] B. J. Perry, Y. Guo, R. Atadero, and J. W. van de Lindt, "Streamlined bridge inspection system utilizing unmanned aerial vehicles (UAVs) and machine learning," *Measurement*, vol. 164, Nov. 2020, Art. no. 108048.
[13] J. Seo, L. Duque, and J. Wacker, "Drone-enabled bridge inspection methodology and application," *Autom. Construct.*, vol. 94, pp. 112–126, Oct. 2018.
[14] M. N. Gillins, D. T. Gillins, and C. Parrish, "Cost-effective bridge safety inspections using unmanned aircraft systems (UAS)," in *Proc. Geotech. Struct. Eng. Congr.*, 2016, pp. 1931–1940.
[15] S. Dorafshan and M. Maguire, "Bridge inspection: Human performance, unmanned aerial systems and automation," *J. Civil Struct. Health Monitor.*, vol. 8, no. 3, pp. 443–476, Jul. 2018.

[16] E. Jeong, J. Seo, and J. Wacker, "Literature review and technical survey on bridge inspection using unmanned aerial vehicles," *J. Perform. Constructed Facilities*, vol. 34, no. 6, Dec. 2020, Art. no. 04020113.
[17] L. DeRen, L. Yong, and Y. XiuXiao, "Image-based self-position and orientation method for moving platform," *Sci. China Inf. Sci.*, vol. 56, no. 4, pp. 1–14, 2013.
[18] B. Kakillioglu, J. Wang, S. Velipasalar, A. Janani, and E. Koch, "3D sensor-based UAV localization for bridge inspection," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 1926–1930.
[19] A. P. Tomiczek, T. J. Whitley, J. A. Bridge, and P. G. Ifju, "Bridge inspections with small unmanned aircraft systems: Case studies," *J. Bridge Eng.*, vol. 24, no. 4, Apr. 2019, Art. no. 05019003.
[20] A. Loquercio, E. Kaufmann, R. Ranftl, M. Müller, V. Koltun, and D. Scaramuzza, "Learning high-speed flight in the wild," *Sci. Robot.*, vol. 6, no. 59, Oct. 2021, Art. no. eabg5810.
[21] C. Cadena *et al.*, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1309–1332, Dec. 2016.
[22] G. Wang, Z. Shi, Y. Shang, X. Sun, W. Zhang, and Q. Yu, "Precise monocular vision-based pose measurement system for lunar surface sampling manipulator," *Sci. China Technol. Sci.*, vol. 62, no. 10, pp. 1783–1794, Oct. 2019.
[23] C. Patruno, M. Nitti, E. Stella, and T. D'Orazio, "Helipad detection for accurate UAV pose estimation by means of a visual sensor," *Int. J. Adv. Robotic Syst.*, vol. 14, no. 5, 2017, Art. no. 1729881417731083.
[24] A. Cantieri *et al.*, "Cooperative UAV–UGV autonomous power pylon inspection: An investigation of cooperative outdoor vehicle positioning architecture," *Sensors*, vol. 20, no. 21, p. 6384, Nov. 2020.
[25] M. Kalaitzakis, B. Cain, S. Carroll, A. Ambrosi, C. Whitehead, and N. Vitzilaios, "Fiducial markers for pose estimation: Overview, applications and experimental comparison of the ARTag, AprilTag, ArUco and STag markers," *J. Intell. Robotic Syst.*, vol. 101, no. 4, p. 71, Apr. 2021.
[26] P. Rudol, M. Wzorek, G. Conte, and P. Doherty, "Micro unmanned aerial vehicle visual servoing for cooperative indoor exploration," in *Proc. IEEE Aerosp. Conf.*, Mar. 2008, pp. 1–10.
[27] J. Li, W. Dong, X. Sheng, and S. Xu, "Visual servoing of micro aerial vehicles with the cooperation of ground vehicle," in *Proc. IEEE/ION Position, Location Navigat. Symp. (PLANS)*, Apr. 2020, pp. 150–155.
[28] A. Masselli and A. Zell, "A novel marker based tracking method for position and attitude control of MAVs," in *Proc. Int. Micro Air Vehicle Conf. Flight Competition*, 2012, pp. 1–6.
[29] X. Yan, H. Deng, and Q. Quan, "Active infrared coded target design and pose estimation for multiple objects," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Nov. 2019, pp. 6885–6890.
[30] L. Zhu, S. Zhang, X. Wang, S. Chen, H. Zhao, and D. Wei, "Multilevel recognition of UAV-to-ground targets based on micro-Doppler signatures and transfer learning of deep convolutional neural networks," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.
[31] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[32] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artif. Intell.*, 1981.

Zhaoying Wang received the B.E. degree in mechanical engineering from Wuhan University, Wuhan, China, in 2019. He is currently pursuing the Ph.D. degree with the State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai, China.

His research interests include multiagent cooperative systems and visual positioning of micro aerial vehicles.

Sensen Liu received the B.S. degree in mechanical engineering from Tongji University, Shanghai, China, in 2016. He is currently pursuing the Ph.D. degree with the School of mechanical and engineering, Shanghai Jiao Tong University, Shanghai.

His research focuses on aerial manipulation, and gripper design, planning, and control of unmanned aerial vehicles manipulation systems.

Gang Chen (Graduate Student Member, IEEE) received the B.E. degree in mechanical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2016, where he is currently pursuing the Ph.D. degree with the State Key Laboratory of Mechanical System and Vibration.

His research interests include active perception and high-speed obstacle avoidance of micro aerial vehicles.

Wei Dong (Member, IEEE) received the B.S. and Ph.D. degrees in mechanical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively.

He is currently an Associate Professor with the School of Mechanical Engineering, Shanghai Jiao Tong University. His research mainly focuses on the cooperative intelligence of unmanned systems.