IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 9, 2019, accepted September 21, 2019, date of publication September 24, 2019,
date of current version October 7, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2943514*

# A Novel Method for Measuring Drogue-UAV Relative Pose in Autonomous Aerial Refueling Based on Monocular Vision

**YUEBO MA[1,2], RUJIN ZHAO[1], ENHAI LIU[1], ZHUANG ZHANG[1,2], AND KUN YAN[1]**
[1]Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, China
[2]University of Chinese Academy of Sciences, Beijing 100149, China

Corresponding author: Rujin Zhao (zrj0515@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61501429, and in part by the Youth Innovation Promotion Association CAS under Grant 2016335.

**ABSTRACT** The key to the docking phase of autonomous aerial refueling missions is the relative pose measurement between the drogue and unmanned aerial vehicles. A novel measurement method for drogue-UAV (unmanned aerial vehicle) relative poses based on monocular vision is proposed in this paper. An adaptive arc-level structural feature extraction algorithm is applied to obtain the features that can be used for the measurements. In this algorithm, the projection of the end plane of the drogue in the image plane is accurately extracted based on an extraction quality metric. The projection contour and centroid of the internal black part of the drogue form the structural feature for the measurements. A robust pose estimation algorithm is presented to estimate the relative pose between the drogue and the UAV. The pose can be solved using a cone that is composed of the optical center of the camera and the structural feature. A Kalman filter is then applied to robustly estimate the pose of the drogue. A simulation and ground experiment are used to verify the robustness and effectiveness of the proposed method, which achieves good performance compared with other methods.

**INDEX TERMS** Autonomous aerial refueling, pose measurement, monocular vision, structural feature extraction, UAV.

## I. INTRODUCTION

Unmanned aerial vehicles (UAVs) are widely used in military and civilian fields, but the biggest current limitation is their insufficient flight times and effective loads. Aerial refueling has been the most effective method to solve this issue. The boom-receptacle refueling system [1] and drogue-probe refueling system [2] are two currently hardware configurations that are used for aerial refueling. Nevertheless, the aerial refueling of UAVs is limited by the development of autonomous aerial refueling (AAR) technology. For AAR a key issue is the need for a highly accurate measurement of the relative ''drogue-UAV'' pose in the docking phase. The measurement method is used to guide the UAV within a defined range below the tanker. After that, the docking of the UAV probe and the tanker drogue is completed. Many recent studies have been conducted on the t AAR of UAVs. However, there are some limitations that are associated with their use. For example, GPS coverage might not always be available, and GPS signals may be distorted by the tanker airframe. [3] Therefore, the research on high-precision pose measurement based on machine vision (MV) for AAR is gaining increasing attention. The implementation of the vision-based measurement method requires a camera to be installed on the UAV to provide the images of the target (that is, the drogue of the tanker), which are then processed to derive the pose information. The research of AAR measurement technology based on machine vision is mainly divided into three technical directions: monocular vision measurement technology, stereo vision measurement technology and 3D imaging laser radar technology [4].

Many recent efforts have been made to achieve autonomous aerial refueling of UAVs. Sun *et al.* [5] developed a bionic visual close-range navigation control system for AAR. And Paulson *et al.* [6] presented a novel algorithm for mitigating the negative effects of boom occlusion in stereo-based aerial environments. Zhu *et al.* [7] and Johnson *et al.* [8] attempted to combine different sensors

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

system for AAR, such as the inertial navigation system (INS), GPS, stereo vision system and infrared search and track system (IRST). According to the mathematical modeling, Xufeng *et al.* [9] and Tsukerman *et al.* [12] discussed the ability and success probability of AAR, which have great significance for the study of AAR control methods. Jing *et al.* [10] and Chen et al. [11] presented the measurement methods of the drogue's pose, which are vision methods based on dual-quaternion and infrared vision sensor, respectively. Xu *et al.* [13] designed a binocular vision-based UAV autonomous aerial refueling platform for AAR and obtained the poses using the known geometrical relationship of 7 points. Johnson *et al.* [14] designed an extended Kalman filter to combine the stereo camera system and inertial navigation for automated aerial refueling. Parsons *et al.* [15] presents a 3D graphical simulation that replicates a complete aerial refueling scenario. Luo *et al.* [16] stated a binocular vision system that is based on the light-emitting diodes (LEDs) features. The feature points are described using an improved Harr wavelet transform. Afterward, they estimated the pose of the drogue by matching the description vector of the feature points. Since binocular vision measurement is limited by the effect of the baseline length, it is mostly applied to boom AAR, which is a close-range measurement. In addition, these methods always need artificial marks by modifying the drogue. Another option, 3D imaging laser radar technology has been explored as a measurement method for AAR. Chen *et al.* [17], [18], used a 3D Flash LIDAR to estimate the center position of the drogue, and the 3D point cloud depth information data were input to the random sample consensus (RANSAC) algorithm, whereas, a pose information estimation of the drogue was not performed. The drogue's position estimation algorithm based on the 3D point cloud requires more time than the vision measurement method. Generally, the weight and energy consumption of 3D imaging laser radars are larger than those of cameras.

In summary, since the model parameters of the drogue are known, the monocular vision measurement technique can be used for drogue-probe AAR. The technique can measure a larger range than binocular vision for the reason that it is not limited by a baseline [19]. Moreover, the real-time implementation of monocular vision is easier than laser radar, and the equipment weight and energy consumption are much lower. Xin *et al.* [20] addressed a feature matching algorithm based on the perspective transform to estimate the pose of a drogue. Huang *et al.* [21] presented a monocular vision method to estimate the pose of a drogue by using the internal circular refueling port of the drogue. Ma *et al.* [22] designed a drogue detection and measurement method that uses the projection features of a drogue. Recently, Sun *et al.* [23] elaborated a robust landmark detection and position measurement based on monocular vision for the autonomous aerial refueling of UAVs. These researchers used a multitask parallel deep convolution neural network to detect the landmarks of the target drogue. The ellipse fittings of landmarks are used to measure the pose of a drogue. A direct geometric interpretation is used to measure the positions of a drogue's two salient parts and a fusion strategy fuses the measurement results for a drogue's different parts to achieve the position measurement. The landmark detection error with this method is 3.9%, the running speed is 220.2 frames/s with GPU (GTX TITAN X) acceleration, and the average relative error is 1.5%. However, there are quite a few challenges in this state-of-the-art technique. First, the reliability of measuring the pose of a drogue by using the manually marked landmarks is worth considering. Secondarily, the landmark detection error is related to the manually marked landmarks, and the manual marking error should be considered. Whether the measurement method can run in real time on an embedded system or computing resource constrained system requires verification. In this paper, our research focuses on the accurate extraction of a drogue's features and the robust estimation of the drogue's 3D pose. Here, the precise extraction method of a drogue's features effectively solves the shortcomings of a drogue's manually marked features. The high-precision elliptical fitting lays the foundation for the measurement of a drogue's pose. The simple drogue's pose estimation method based on a Kalman filter effectively alleviates the computing resources demands and can operate in real-time on an embedded system.

The motivation of this paper is to provide a real time high-precision pose measurement method based on monocular vision for drogue-probe autonomous aerial refueling. The main contributions of this paper are as follows.

1) A robust high-precision drogue pose estimation method is raised for autonomous aerial refueling. The method includes two parts: i) the accurate extraction of a drogue's features based on arc groupings, and ii) the robust high-precision visual measurement based on 3D analytic geometry.
2) A fast and accurate arc and centroid extraction algorithm is introduced to obtain the drogue's structural features for measurement.
3) Considering the structural constraints of a drogue, a robust pose estimation algorithm based on monocular vision is proposed. A cone is created using the optical center of a camera and the projection ellipse of an image plane. The spatial pose of a drogue is robustly and precisely estimated using 3D analytical geometry and the Kalman filter.
4) The influence of the feature extraction accuracy for the drogue on pose estimation is analyzed in detail. The robustness and effectiveness of the method are verified by the experiment results. In addition, the method is applied to an actual project.

The rest of this paper is structured as follows. The related works are reviewed in Section II. The proposed pose method based on monocular vision for autonomous aerial refueling is introduced in Section III. The experiments and results are shown in Section IV. Finally, this paper is concluded in Section V.

## II. RELATED WORK

The relative pose measurement method for a drogue based on vision for autonomous aerial refueling can full fill imaging and measurement tasks, including the potential to be installed without modification to the drogue, and can obtain more precise five degrees of freedom (DOF) pose, which is why this measurement method has received gradual study. Many scholars have carried out research on this measurement scheme, and the results are described below.

At the beginning of the development of AAR technology, the visual measurement methods for autonomous aerial refueling missions were mostly based on artificial features such as light-emitting diodes (LEDs) or spray marks, which require the drogue to be modified. Valsek et al. [24]–[27] developed a visual navigation system called VisNav. To accomplish the measurement of a drogue's relative pose with the measuring circuit, the LEDs that were mounted on the drogue in this system emit lights with different frequencies, and the position sensing diode (PSD) that was mounted on the receiver acquires the modulated lights and generates a current. Then, the pose of the drogue is calculated by using a Gaussian least-squares differential correction algorithm (GLSDC), which is used in conjunction with a pinhole imaging model based on four or more beacon data points. Pollini *et al.* [28] put forward a fusion switch strategy for the GPS and vision system. LEDs are mounted on the drogue, and a charge-coupled device (CCD) camera and a near infrared filter are used to identify the drogue. GPS is used to measure the relative position between the tanker and the receiver. Machine vision is used to measure the distance between the probe and the drogue with the help of some marker points that were installed. According to the feature extraction and feature matching of the marker points that were projected on the image, the best matching relationship between the 2-dimensional (2D) feature and the 3-dimensional (3D) marker point is constructed, and the rough pose estimation is obtained using the Lu, Hager, and Mjolsness (LHM) algorithm [29] with a fixed number of steps. The best match is selected from the features matching set (FMS) as the element with the lowest collinearity error to complete the pose estimation. Wang et al. [30], [31] brought up using the red ring feature of special materials that were coated with high reflection characteristics on the surface of a drogue to detect the drogue according to color analysis and contour analysis, and also applying a relative position measurement algorithm based on the camera calibration model. The 3D center position of the drogue ring can be calculated using the pinhole imaging model and camera calibration analysis.

In addition, many scholars began to explore the potential of visual measurement, using the known structural features of a drogue to estimate the relative pose relationship between drogue and a receiver. Martinez *et al.* [32], [33] introduced a 3D position measurement method that is based on the circular circumscribed rectangle mapping relationship of a drogue based on monocular vision, which ignored the influence of the yaw angel and the pitch angle on the position. Duan *et al.* [34] described a binocular vision-based UAV autonomous aerial refueling platform for boom-and-receptacle refueling (BRR). Meng *et al.* [35] mainly studied the relative pose measure for a tanker and UAV in the docking phase and presented an orthogonal iteration algorithm to estimate the optimal solutions of the rotation matrix and translation vector. Yin *et al.* [36] presented a drogue position measurement algorithm based on monocular vision. The method also ignored the influence of the angle between the drogue and optical axis. When the angle is increased, the accuracy of the two given algorithms will be obviously reduced. Sun *et al.* [23] proposed a measurement method that uses the circle pose determination with a direct geometric interpretation [37] to measure the positions of a drogue's two salient parts. A fusion strategy was used to calculate the position of a drogue using a 3D model of the drogue. Shiu and Ahmad [38] raised a precise, closed-form pose measurement method for the viewpoint determination of circular features. The method is based on the formation of a second degree cone having the center of projection (camera focal point) as the vertex and passing through the ellipse on the image plane.

In all of the previously mentioned methods, the precise extraction of drogue's features greatly affects the results of the drogue pose estimation. And yet, the related research cannot be found on the accurate extraction of a drogue's features. Researchers can only investigate the relevant literatures on the precise extraction of ellipses for vision-based drogue pose measurement methods mostly use the elliptical shape of a drogue's projection as the feature for pose estimation. The technologies of ellipse extraction can be divided into two classes: One class of methods is based on the Hough Transform. The standard Hough Transform [39] involves large computational costs. The other class of ellipse extraction approaches is edge segment detection techniques. Fornaciari *et al.* [40], [41] use geometric constraints as a selection strategy for the arcs belonging to the same ellipse and estimate the parameters by decomposing the parameter space. Mai *et al.* [42] advanced line segments to approximate the potential elliptical arcs according to connectivity and curvature conditions. Bai *et al.* [43] applied the property of elliptical concavity to group arcs to identify candidate ellipses. Liu and Qiao [44] put forward a hierarchical approach that was motivated by the fact that any segment of an ellipse can identify itself in ellipse reconstruction using geometric constraints. Dong *et al.* [45] insisted the accurate detection of ellipses using false detection control based on gradient analysis. The method adopted arc selection, smart grouping, and the repeated utilization of gradient information to significantly reduce the computations that are otherwise needed without compromising the detection effectiveness.

## III. METHOD

In this section, we will systematically expound the proposed method, which includes the accurate extraction of the shape

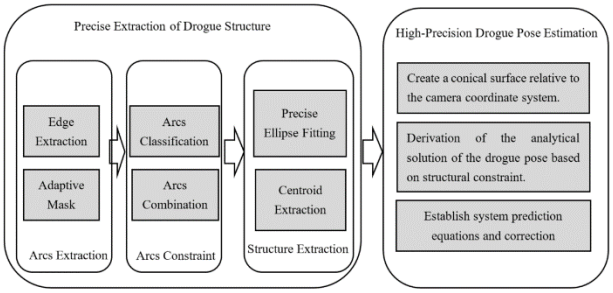


**FIGURE 1. Framework of the proposed pose measurement method based on monocular vision.**

of the drogue and the high-precision pose estimation phases. The implementation of this method is based on the detection of the drogue, and the drogue detection method is introduced in detail in [22]. The main idea is to use an adaptive mask to quickly extract the arcs that may be part of the drogue's shape. Furthermore, according to the prior knowledge of the projection of the drogue on the image plane as an ellipse or circle, the elliptical shape is accurately estimated for the obtained arcs. The estimation of the elliptical shape greatly affects the pose estimation accuracy, and its influence will be analyzed in detail in subsequent experiments. In the pose estimation stage, we establish the analytical equations of the projection ellipse and the end plane of the drogue using the projected relationship of the drogue on the image plane and the structural constraints of the drogue. To obtain the analytical solution of drogue pose. In order to obtain a highly precise and robust pose, a Kalman filter [8] is used to accurately estimate the drogue pose.

### A. DROGUE STRUCTURE EXTRACTION

A fast drogue structure extraction algorithm based on the shape characteristics of the drogue is proposed. First, we create an adapting ring mask, use the Canny operator [46] to obtain the edge map, and use the intersection of the mask and the edge map in the next step. Subsequently, the arcs are constrained to four quadrants according to the gradient direction s, lengths, straightness and positional relationships of the arcs. Next, the extraction quality metric is established based on the fitting errors, the ratio of the circumference and the ratio of the minor axis to the major axis of the ellipse. Finally, the ellipse is selected with the best quality using clustering. The centroid position of the internal black part of the drogue is obtained by generating a circular mask, threshold segmentation and centroid extraction. The entire process is shown in Fig. 1, which describes the precise method for extracting the drogue structure. The key technologies are explained below.

#### 1) ARC EXTRACTION AND CONSTRAINT

In this stage, an adaptive ring mask $M$ is generated based on the size of the image. The outer ring and inner ring radii of the mask $M$ are 1/2 and 1/3 of the longest side of the image, respectively. The canny edge detector [46] is used to extract the edge image $E$ of the image and define the edge points as $e_i = (x_i, y_i, \vartheta_i)$. The edge points consist of the position $(x_i, y_i)$ and the gradient direction $\vartheta_i$. The gradient direction $\vartheta_i$ is calculated using the Sobel operator in the canny edge detection algorithm. The function $J(E, M)$ is used to find the intersection of two images.

$$J(E,M) = \begin{cases} J_{ij} = 1, & E_{ij} \neq 0 \cap M_{ij} \neq 0 \\ J_{ij} = 0, & others \end{cases} \tag{1}$$

The intersection operation can greatly reduce the number of calculations, because the mask only has a ring area that is nonzero, and the edge points that remain after the above operation only exist in the ring region of the mask $M$. Then, according to the gradient direction of the edge points, a piecewise function is defined. Each edge point is classified using the function $D(e_i)$ and divided into four different directions according to $\vartheta_i$.

$$D(e_i) = \begin{cases} I, & 0 < \vartheta_i \leq 90 \\ II, & 90 < \vartheta_i \leq 180 \\ III, & 180 < \vartheta_i \leq 270 \\ IV, & 270 < \vartheta_i \leq 360 \end{cases} \tag{2}$$

The edge points are grouped using directional constraints. The eight neighborhood connectivity is verified in the four groups, and the edge points are connected to form arcs. $N^k$ represents the number of edge points that belong to the arc, and the $Conected(\cdot)$ function is used to verify the eight-neighborhood connectivity of the two edge points.

$$C^k = \left\{ \left( e_1^k \ldots e_{N^k}^k \right) : D\left(e_i^k\right) = D\left(e_j^k\right), Conected(e_i^k, e_j^k) \right\} \tag{3}$$

According to the straightness constraint, $\mathrm{MinB}^k$ is defined as the short side of the smallest bounding rectangle containing all the points of $C^k$. If $\mathrm{MinB}^k < Th_{minb}$, the arcs that are very straight do not belong to the end plane of the drogue projection ellipse. Some arcs, such as $C^k$ $(N^k < Th_{length})$, are not sufficient to characterize the ellipse. Additionally, regarding positional constraints, the image center is defined as $Center(x_c, y_c)$. The center of the smallest bounding rectangle is defined as the center of the arc, which is represented by $Ac^k(x_{ac}, y_{ac})$. According to prior knowledge, the arc that is farthest from the center of the image in each quadrant is most likely to belong to the end plane projection of the drogue. Therefore, Algorithm 1 is proposed to find all the arcs that may belong to the projection of the end plane of a drogue.

$$\delta(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \tag{4}$$

$$C^l = \bigcup_{i=0}^{n-1} C_i^k \delta\left(N^k - Th_{length}\right) \delta\left(MinB^k - Th_{minb}\right) \tag{5}$$

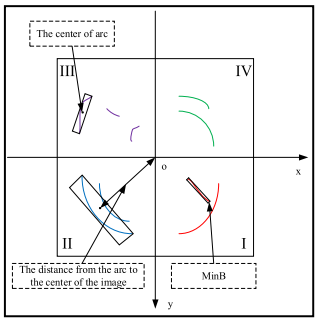


**FIGURE 2.** Arcs constraint diagram.

**Algorithm 1** Arc Extraction Based on the Drogue's Structural Constraints

**Input:** The region of interest (ROI) of the image,
**Output:** The farthest arc from the center of the ROI in each quadrant.

**1.** The mask $M$ is adaptively generated according to the ROI size, and the edge image $E$ of the ROI is acquired by the Canny detector.
**2.** The edge points of $E$ are greatly reduced by $J(E, M)$.
**3.** Edge points are assigned to the four quadrants by the function $D(e_i)$, which based on the gradient direction.
**4.** According to eight neighborhood connectivity, the edge points are connected to form arcs $C^k$.
**5.** According to (5) and (6), some arcs are eliminated, which do not meet the constraints.
**6.** Set the initial distance $d_0$ equal to 0.
**7.** For i=0 to the number of arcs in each quadrant
**8.** Calculate the distance $d_i$ from the arc to the center of the ROI.
**9.** If $d_i > d_0$
**10.** $d_0 = d_i$, $C_f = C_i$.
**11.** End
12. End
**13.** Return the farthest arc in each quadrants.

$$P\left(C^l\right) = \begin{cases} x_{ac} - x_c > 0 \; and \; y_{ac} - y_c < 0, & if \; C^l \in I \\ x_{ac} - x_c < 0 \; and \; y_{ac} - y_c < 0, & if \; C^l \in II \\ x_{ac} - x_c < 0 \; and \; y_{ac} - y_c > 0, & if \; C^l \in III \\ x_{ac} - x_c > 0 \; and \; y_{ac} - y_c > 0, & if \; C^l \in IV \end{cases} \tag{6}$$

### 2) STRUCTURE EXTRACTION

To achieve robust and exact pose of drogue, the structural feature that is used for measurement must be accurately and robustly extracted. The drogue's salient parts include two parts: the internal black part and the end plane of the drogue. The contour shape of the end plane of the drogue and the center position of the internal black part can be used as the image features for the measurement. The shape of the end plane of the drogue can be obtained using the farthest arcs that were obtained above. The position of the center of the internal black part can be obtained using centroid extraction.

First, the ellipse is fitted using different combinations of the farthest arcs in the four quadrants. $C_f^q$ $(q \in \{I, II, III, IV\})$ is used to represent the farthest arc in each quadrant. A function $ell\left(C_f^q\right)$ is defined to fit the ellipse. The fitting function of the ellipse is based on the least squares ellipse fitting algorithm [47]. Equation (7) of the ellipse in the image coordinate system is obtained using the fitting function.

$$au^2 + bv^2 + cuv + du + ev + f = 0 \tag{7}$$

Additionally, the ellipse fitting quality metric function is established based on the fitting errors, the ratio of the circumference and the ratio of the semi-minor axis to the semi-major axis of the ellipse to evaluate the reliability of the fitted ellipse. We simplify (7) to obtain (8).

$$Ax^2 + By^2 + Cxy + Dx + Ey + 1 = 0 \tag{8}$$

where $A = a/f, B = b/f, C = c/f, D = d/f, E = e/f$

$$\begin{cases} x_c = \dfrac{BE - 2CD}{4AC - B^2} \\ y_c = \dfrac{BD - 2AE}{4AC - B^2} \\ a^2 = \dfrac{2(Ax_c^2 + Cy_c^2 + Bx_cy_c - 1)}{A + C - \sqrt{(A - C)^2 + B^2}} \\ b^2 = \dfrac{2(Ax_c^2 + Cy_c^2 + Bx_cy_c - 1)}{A + C + \sqrt{(A - C)^2 + B^2}} \\ \rho = \dfrac{1}{2} arctan \dfrac{B}{A - C} \end{cases} \tag{9}$$

The center of the fitted ellipse can be represented by $(x_c, y_c)$. $a, b, and \; \rho$ represent the semi-major axis, the semi-minor axis and the angle with respect to the x-axis of the ellipse, respectively. According to the parametric equation of the ellipse, the distance $d$ from the point to the ellipse can be obtained as follows:

$$\begin{cases} X = \dfrac{[(x_i - x_c) \, cos \; \rho + (y_i - y_c) sin \; \rho]^2}{a^2} \\ Y = \dfrac{[(y_i - y_c) \, cos \; \rho - (x_i - x_c) sin \; \rho]^2}{b^2} \\ d = |X + Y - 1| \end{cases} \tag{10}$$

If $d$ is less than a threshold $(Th_{error})$, it implies that the point $(x_i, y_i)$ on the arc group is close enough to the edge of the ellipse. $N_\beta$ is the number of points that satisfy the above conditions. $N_c$ is total number of points that participate in fitting the ellipse. A function $\varphi\left(C_f\right)$ is defined to represent the ratio of the number of points that are less than the fitting error to the total number of points that are used to fit the ellipse.

$$\varphi\left(C_f\right) = \frac{N_\beta}{N_c} \tag{11}$$

A candidate ellipse $\varepsilon_i$ with $\varphi\left(C_f\right) > Th_{score}$ is considered to be valid; otherwise, it is a false detection and is discarded. $C_f$ is the combination of the farthest arcs.

The $l_r$ is defined to represent the ratio of the number of edge pixels of the group to the circumference of the corresponding ellipse.

$$l_r = \frac{Number\ of\ edge\ pixels\ of\ the\ group}{Circumference\ of\ corresponding\ ellipse} \tag{12}$$

The $s_r$ is defined to represent the ratio of the semi-minor axis to the semi-major axis. This is done because the ellipse that we will get based on prior knowledge is not too flat.

$$s_r = \frac{the\ length\ of\ Semi-minor\ axis}{the\ length\ of\ Semi-major\ axis} \tag{13}$$

Each indicator has its shortcomings and false extractions cannot be completely excluded when the indicators are used alone [48]. Therefore, we set a score of ratios. We assign a score ranging from 0 to 0.5 to $l_r$ and $s_r$, and denote them as s ($l_r$) and s($s_r$), respectively. Thus, S is computed as follows:

$$\begin{cases} S = s\,(l_r) + s(s_r) \\ s\,(l_r) = 0.5 \times l_r \\ s\,(s_r) = 0.5 \times \left( \left( \dfrac{(s_r - k_2)}{1 - k_2} - 1 \right) exp\left( \dfrac{k_1\,(s_r - k_2)}{k_2 - 1} \right) + 1 \right), \end{cases} \tag{14}$$

The reliability score is defined as follows:

$$S_{reliable} = 0.5S + 0.5\varphi\left(C_f\right) \tag{15}$$

The clustering method is used to select the most reliable ellipse as the contour shape of the end plane of the drogue. The following formula is used to judge whether the ellipse belongs to the same ellipse.

$$\begin{cases} \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} < T_d \\ \left|a_j - a_i\right| < T_a \\ \left|b_j - b_i\right| < T_b \\ \left|\rho_j - \rho_i\right| < T_\rho \end{cases} \tag{16}$$

Here, $T_d$, $T_a$, $T_b$ and $T_\rho$ are the thresholds of the Euclidean distances of the two different elliptical centers, the semi-major axis, the semi-minor axis and the angles, respectively. After clustering, we chose the ellipse $\varepsilon$ with the highest reliability as the image feature that is used for the measurement.

The centroid of the inner black part of the drogue is used to represent another structural feature of the drogue. The geometric moments of the image area are used to find the centroid [49]. Before getting the centroid, the inner black part of the drogue should be found in the ROI. First, an adaptive mask $M_b$ is generated to reduce the number of computations, but this mask is a circular mask with a radius that is one-third of the longest side of the ROI. The mask $M_b$ acts on the gray image of the ROI to get image H. Then, threshold segmentation and morphological operations are used to obtain the inner black part of the drogue. Finally, the geometric moment of the image area is used to extract the centroid $(i_c, j_c)$. The centroid extraction process is shown in Fig. 3.

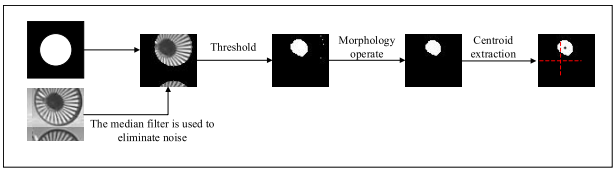


**FIGURE 3.** Centroid extraction process.

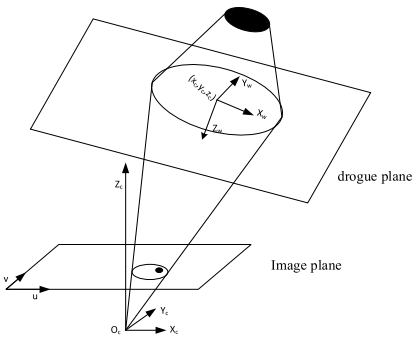


**FIGURE 4.** Drogue target imaging model.

$$\begin{cases} M_{pq} = \int_{a_1}^{a_2} \int_{b_1}^{b_2} x^p y^q f\,(x, y)\,dxdy \\ i_c = \dfrac{M_{10}}{M_{00}},\ j_c = \dfrac{M_{01}}{M_{00}} \end{cases} \tag{17}$$

At this point, the shape ellipse $\varepsilon$ of end plane of the drogue and the centroid $(i_c, j_c)$ of the inner black part have been extracted, and these features can be used to measure the pose using the structural model of the drogue.

### B. ROBUST DROGUE POSE ESTIMATION

In this section, a robust pose measurement method based on the drogue's structural model is proposed. The structure of the drogue's salient part including the internal black part and the end plane of the drogue are known. The projection of the end plane of the drogue is a circle or ellipse in the image plane. The structural constraint can be efficiently used to eliminate ambiguous solutions to the drogue pose from the projection of the drogue. The Kalman filter is used to obtain accurate and robust drogue pose.

#### 1) PRINCIPLE OF SPATIAL CIRCULAR POSE SOLVING

The pose measurement method presented here is based on the 3D location of circular and spherical features [50]–[53]. Due to the rotational symmetry of the drogue, its 3D pose parameter with respect to the camera coordinate system can be represented by five pose parameters: 3 for the position and 2 for the orientation. The drogue imaging model is shown in Fig. 4.

The viewing geometry of the drogue's two salient features is shown in Fig. 4. In Fig. 4, $O_c - X_c - Y_c - Z_c$ is the camera coordinate system and $O_c$ is the optical center of the camera and the apex of the viewing cone. $O_w - X_w - Y_w - Z_w$ is the

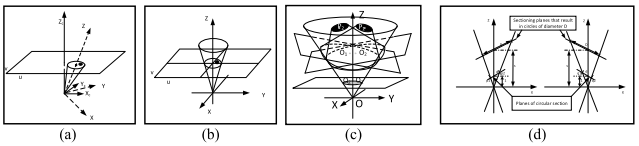


**FIGURE 5.** Process of solving the pose of a drogue.

world coordinate system of the drogue. u − v is the image coordinate system. The process of finding the 3D pose of the drogue using the ellipse that is projected from a circular feature with a known radius and the centroid of the internal black part of drogue is shown in Fig. 5, and the derivation process is as follows.

Fig. 5(a) shows the equation of the cone that passes through the ellipse and has the focal point as the vertex relative to the camera frame. Fig. 5(b) displays a new frame of reference in which the cone has the standard form. Fig. 5(c) presents the ambiguity of the inverse solution from the projection ellipse of the drogue. Fig. 5(d) shows the two planes that intersect the cone in circles having the same diameter as the model feature. This will result in two sets of solutions for the position and normal vector of the circle, and the structural constraint of the drogue is used to obtain the right solution. Then, the solution was transformed back to the camera frame.

First, the elliptical projection equation of the end plane of the drogue and the centroid position $(u_Q, v_Q)$ of the projection of the inner black part of the drogue is obtained in the image coordinate system.

$$Au^2 + Buv + Cv^2 + Du + Ev + F = 0 \tag{18}$$

$$u = f_0 \frac{x_c}{z_c} \quad v = f_0 \frac{y_c}{z_c} \tag{19}$$

By substituting (19) into (18) and rearranging it, the equation of the cone is obtained in terms of the coefficients of the image ellipse and the focal length $f_0$:

$$ax_c^2 + by_c^2 + cx_c y_c + dx_c z_c + ey_c z_c + fz_c^2 = 0 \tag{20}$$

where $a = Af_0^2$, $b = Bf_0^2$, $c = Cf_0^2$, $d = Df_0$, $e = Ef_0$, and $f = F$.

Equation (20) can be expressed in terms of the quadratic for $M$:

$$[x_c, \quad y_c, \quad z_c]\, M\, [x_c, \quad y_c, \quad z_c]^T = 0 \tag{21}$$

Then,

$$M = \begin{bmatrix} a & \frac{c}{2} & \frac{d}{2} \\ \frac{c}{2} & b & \frac{e}{2} \\ \frac{d}{2} & \frac{e}{2} & f \end{bmatrix} \tag{22}$$

If $P$ is a diagonal matrix for $M$, $P^{-1}MP = Diag(\lambda_1, \lambda_2, \lambda_3)$, then (21) will be standardized using $P$ as a rotation matrix, which is equivalent to changing the reference frame to the one that is represented by $P$.

$$\lambda_1 X^2 + \lambda_2 Y^2 + \lambda_3 Z^2 = 0 \tag{23}$$

The (23) is aligned with the axis of the cone.

$$\frac{X^2}{k_x^2} + \frac{Y^2}{k_y^2} - \frac{Z^2}{k_z^2} = 0 \tag{24}$$

Here, $\alpha$ is any real constant. The ratio of $k_x$, $k_y$, *and* $k_z$ determines the shape of the cone.

$$k_x = \alpha \frac{1}{\sqrt{|\lambda_1|}}, \quad k_y = \alpha \frac{1}{\sqrt{|\lambda_2|}}, \quad k_z = \alpha \frac{1}{\sqrt{|\lambda_3|}} \tag{25}$$

According to the 3D analytic geometry, the cone is projected onto the X-Z plane to get the following equation.

$$X^2 \left[ \frac{1}{k_x^2} + \frac{1}{k_y^2} \right] - Z^2 \left[ \frac{1}{k_y^2} + \frac{1}{k_z^2} \right] = 0 \tag{26}$$

The two planes of the circular section are confirmed by substituting (25) into (26).

$$Z = \pm \sqrt{\frac{|\lambda_1| - |\lambda_2|}{|\lambda_2| + |\lambda_3|}} X \tag{27}$$

A circle that meets the known diameter can be found by translating the intersecting planes. In Fig. 5(d), the intersection planes translating 1 are used as a benchmark. Then, the diameter d of the circle can be obtained.

$$\begin{aligned} Z &= \pm \sqrt{\frac{|\lambda_1| - |\lambda_2|}{|\lambda_2| + |\lambda_3|}} X + 1, \\ d &= \frac{2}{|\lambda_2|} \sqrt{\frac{|\lambda_1|\,|\lambda_3|\,(|\lambda_2| + |\lambda_3|)}{|\lambda_1| + |\lambda_3|}} \end{aligned} \tag{28}$$

To find the center points of the model circles, we first find the center points of the benchmark circles with diameter d, which are the midpoints of p1 and p2 or q1 and q2. The ratio s = D/d is used to obtain the midpoint of the circle with diameter D.

By formulating the equation, the surface normal vector and the spatial center position of the circle with diameter D are obtained in the X-Y-Z coordinate system.

$$l = \pm \sqrt{\frac{|\lambda_1| - |\lambda_2|}{|\lambda_1| + |\lambda_3|}}, \quad m = 0, \quad n = -\sqrt{\frac{|\lambda_2| + |\lambda_3|}{|\lambda_1| + |\lambda_3|}} \tag{29}$$

$$\begin{aligned} X_o &= \pm \frac{D}{2} \sqrt{\frac{|\lambda_3|\,(|\lambda_1| - |\lambda_2|)}{|\lambda_1|\,(|\lambda_1| + |\lambda_3|)}}, \quad Y_o = 0, \\ Z_o &= \frac{D}{2} \sqrt{\frac{|\lambda_1|\,(|\lambda_2| + |\lambda_3|)}{|\lambda_3|\,(|\lambda_1| + |\lambda_3|)}} \end{aligned} \tag{30}$$

According to the above discussion, we can only obtain two solutions that have not determined the real pose of the drogue. Therefore, the following equations are established by using

the structural constraint of the drogue to obtain the unique solution of the drogue's pose.

$$\sigma\left(u_Q, v_Q\right) = \begin{cases} 1, & u_Q - u_c < 0 \\ -1, & u_Q - u_c \geq 0 \end{cases} \tag{31}$$

$$l = \sigma\left(u_Q, v_Q\right)\sqrt{\frac{|\lambda_1| - |\lambda_2|}{|\lambda_1| + |\lambda_3|}},$$

$$X_o = \sigma\left(u_Q, v_Q\right)\frac{D}{2}\sqrt{\frac{|\lambda_3|\left(|\lambda_1| - |\lambda_2|\right)}{|\lambda_1|\left(|\lambda_1| + |\lambda_3|\right)}} \tag{32}$$

The eigenvectors of $M$ form the transformation matrix $T = [e_1, e_2, e_3]$. According to the transformation matrix, we can obtain the position coordinate $[x_o, y_o, z_0]^T = T\,[X_o, Y_o, Z_o]^T$ and the surface normal vector $\vec{V} = [X_V, Y_V, Z_V]^T = T\,[l, m, n]^T$ of the drogue relative to the camera coordinate system. The yaw angle $\varphi$ and the pitch angle $\omega$ can be obtained using the surface normal vector, as follows:

$$\omega = tan^{-1}Y_V, \quad \varphi = tan^{-1}\frac{X_V}{Z_V} \tag{33}$$

At this point, pose estimation of the drogue target is completed and the pose parameters are obtained as follows:

$$\begin{bmatrix} x_o \\ y_o \\ z_0 \end{bmatrix} = T\begin{bmatrix} X_o \\ Y_o \\ Z_0 \end{bmatrix}, \quad \begin{bmatrix} \omega \\ \varphi \end{bmatrix} = \begin{bmatrix} tan^{-1}Y_V \\ tan^{-1}\dfrac{X_V}{Z_V} \end{bmatrix} \tag{34}$$

#### 2) ROBUST DROGUE POSE ESTIMATION

Since there are some errors in the extraction of the image features for measurement, there is a deviation in the calculated result of the drogue's pose. The Kalman filter is used to accurately estimate the pose of the drogue [28]. By analyzing the motion law between the drogue and the UAV [15], knowing that the UAV approaches the drogue at a uniform speed in the z-axis direction. The drogue swings in a certain frequency range on the X and Y axes direction due to the presence of atmospheric turbulence. Here, $g_k = (x_k, y_k, z_k, \omega_k, \varphi_k)$ is used to represent the measured value.

A time unit dk is defined, which is the time between two consecutive frames. The prediction is performed as follows:

$$\hat{s}_k^- = R\hat{s}_{k-1}^- \tag{35}$$

$$P_k^- = RP_{k-1}R^T + Q \tag{36}$$

Here, $\hat{s}_{k-1}^-$ and $\hat{s}_k^-$ are the system states before and after prediction at time t. $P_k^-$ and $P_{k-1}$ are the prior and post error covariance, respectively, and $Q$ is the process noise covariance. In this method, the system state is defined as a 15-dimensional vector:

$$s_k = (x_k, y_k, z_k, \omega_k, \varphi_k, \dot{x_k}, \dot{y_k}, \dot{z_k}, \dot{\omega_k}, \dot{\varphi_k}, \ddot{x_k}, \ddot{y_k}, \ddot{z_k}, \ddot{\omega_k}, \ddot{\varphi_k}) \tag{37}$$

Here, $x_k$, $y_k$, $z_k$, $\omega_k$, *and* $\varphi_k$ are the spatial position and orientation of the drogue, the velocity $(\dot{x_k}, \dot{y_k}, \dot{z_k})$, the angular velocity $(\dot{\omega_k}, \dot{\varphi_k})$, the acceleration $(\ddot{x_k}, \ddot{y_k}, \ddot{z_k})$ and the angular acceleration $(\ddot{\omega_k}, \ddot{\varphi_k})$, respectively. The model R is define based on the physics equation of the displacement with velocity and acceleration. It is assumed that the acceleration is a constant in $dk$.

$$\begin{cases} x_k = x_{k-1} + \dot{x}_{k-1}\cdot dk + \dfrac{1}{2}\cdot\ddot{x}_{k-1}\cdot dk^2 \\ y_k = y_{k-1} + \dot{y}_{k-1}\cdot dk + \dfrac{1}{2}\cdot\ddot{y}_{k-1}\cdot dk^2 \\ z_k = z_{k-1} + \dot{z}_{k-1}\cdot dk + \dfrac{1}{2}\cdot\ddot{z}_{k-1}\cdot dk^2 \\ \omega_k = \omega_{k-1} + \dot{\omega}_{k-1}\cdot dk + \dfrac{1}{2}\cdot\ddot{\omega}_{k-1}\cdot dk^2 \\ \varphi_k = \varphi_{k-1} + \dot{\varphi}_{k-1}\cdot dk + \dfrac{1}{2}\cdot\ddot{\varphi}_{k-1}\cdot dk^2 \end{cases} \tag{38}$$

$$\begin{cases} \dot{x}_k = \dot{x}_{k-1} + \ddot{x}_{k-1}\cdot dk \\ \dot{y}_k = \dot{y}_{k-1} + \ddot{y}_{k-1}\cdot dk \\ \dot{z}_k = \dot{z}_{k-1} + \ddot{z}_{k-1}\cdot dk \\ \dot{\omega}_k = \dot{\omega}_{k-1} + \ddot{\omega}_{k-1}\cdot dk \\ \dot{\varphi}_k = \dot{\varphi}_{k-1} + \ddot{\varphi}_{k-1}\cdot dk \end{cases} \tag{39}$$

After the prediction, the measurement matrix $H$ can be obtained using $g_k = Hs_k$. $L$ is the measurement noise covariance, represent the noisiness of the measurement. The Kalman gain $K_k$ can be adjusted by manipulating R.

$$\begin{cases} K_k = P_k^- H^T\left(HP_k^- H^T + L\right)^{-1} \\ \hat{s}_k = \hat{s}_k^- + K_k\left(g_k - H\hat{s}_k^-\right) \\ P_k = P_k^- - K_k HP_k^- \end{cases} \tag{40}$$

The position and orientation parameters of $\hat{s}_k$ are used for the pose estimation of the drogue.

## IV. EXPERIMENT

### A. ALGORITHM SIMULATION ANALYSIS

The simulation experiment is divided into two parts, which quantitatively analyze the proposed extraction algorithm of the drogue's features and the estimation algorithm of the drogue's pose. The following camera parameters are used in the simulation: a resolution of $2048 \times 2048$ and a focal length of 8.06 mm. According to the motion law of the drogue [34], the effects of illumination variations and motion blur on the extraction algorithm was simulated. The illumination variation was simulated using a hypothetical light source that randomly moves and changes the brightness in the first raw image. The pixel brightness of the image was changed using the distance from the light source. The brightness change formula is as follows:

$$f(x, y) = k \times \left(1 - \frac{\sqrt{(x - x_0)^2 + (y - y_0)^2}}{r}\right) \tag{41}$$

Here, $k$ is a constant that is used to control the range of the changing brightness. $x$ and $y$ represent the position of the pixel. $x_0$ and $y_0$ are used to represent the position of the light source. Finally, $r$ is used to represent the active radius of the light source. The motion blur is simulated using a Gaussian blur function. The simulation results of the extraction algorithm of the drogue's features are shown in Fig. 6.

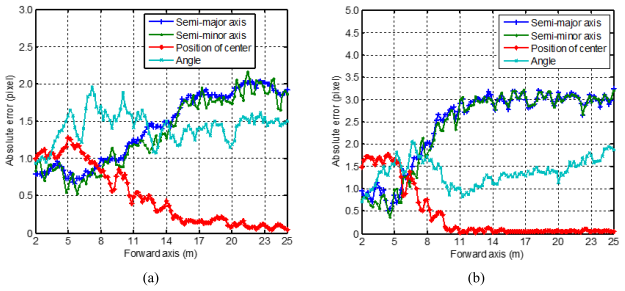


**FIGURE 6.** Simulation results of the extraction algorithm of the drogue's features.

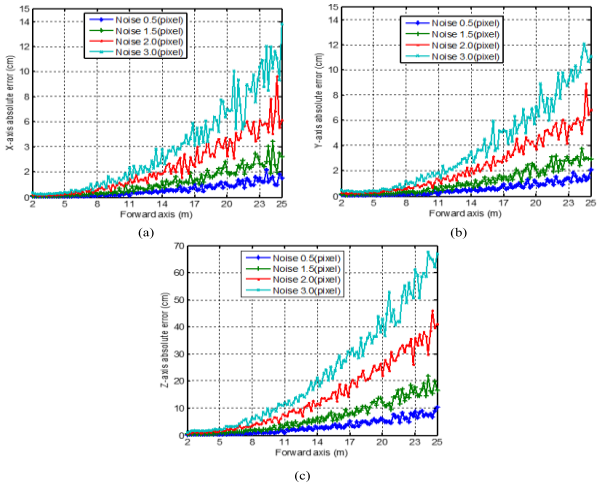


**FIGURE 7.** Position error of pose estimation under different amounts of semi-major axis extraction noise.

**TABLE 1.** Orientation error of the pose estimation algorithm under different semi-major axis extraction noise.

| Performance | Minimum error (°) | Maximum error (°) | Mean error (°) | Standard variance(°) |
|---|---|---|---|---|
| Pitch angle | 0 | 6.32 | 2.98 | 1.84 |
| Yaw angle | 0 | 4.86 | 1.85 | 1.55 |

Fig. 6(a) shows the extraction error of the extraction algorithm under drastic illumination changes. As shown in the results, the extraction error of the end plane ellipse of the drogue was less than three pixels at different distances. Fig. 6(b) shows the extraction error of the extraction algorithm under motion blur. As shown in Fig. 6(b), the extraction error was less than 3.5 pixels at different distances.

After that, the simulation analysis of the pose estimation algorithm based on the results of the feature extraction algorithm was carried out. The different degrees of noise were added to the semi-major axis, semi-minor axis, center and angle of the end plane ellipse of the drogue, respectively. In addition, the pose of drogue was randomly changed at different positions within the field of view. The results of the pose error of the drogue are shown in Fig. 7, Fig. 8 and Fig. 9.

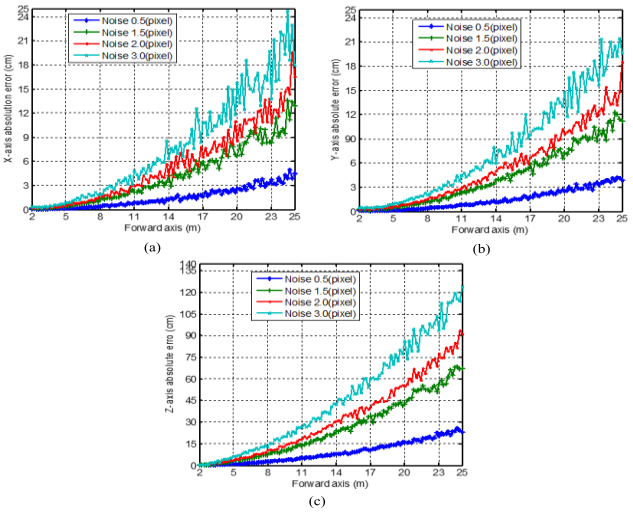


**FIGURE 8.** Position error of the pose estimation algorithm under different semi-minor axis extraction noise.

**TABLE 2.** Orientation error of the pose estimation algorithm under different semi-minor axis extraction noise.

| Performance | Minimum error (°) | Maximum error (°) | Mean error (°) | Standard variance(°) |
|---|---|---|---|---|
| Pitch angle | 0 | 6.78 | 3.05 | 1.83 |
| Yaw angle | 0 | 5.83 | 1.94 | 1.63 |

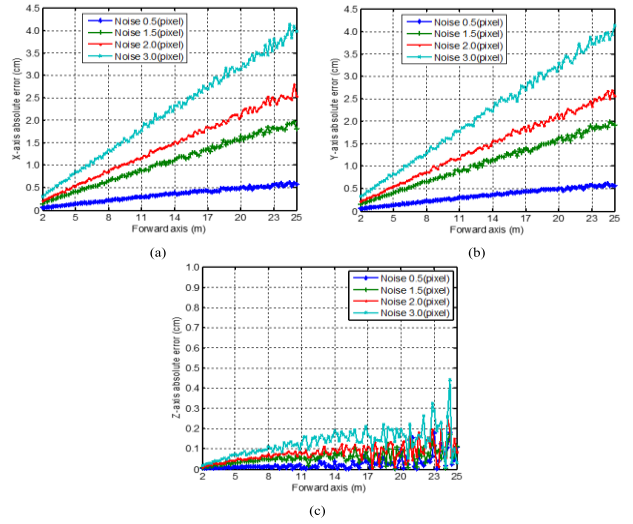


**FIGURE 9.** Position error of the pose estimation algorithm under different center extraction noise.

Fig. 7(a), (b) and (c) show the absolute error of the x-axis, y-axis and z-axis at different distances, respectively. The position error becomes larger as the distance increases, but the x-axis and y-axis error change less with respect to the z-axis.

Compared to the semi-major extraction noise, the semi-minor extraction noise has a greater influence on the absolute error of the z-axis.

**TABLE 3. Orientation error of the pose estimation algorithm under different center extraction noise.**

| Performance | Minimum error (°) | Maximum error (°) | Mean error (°) | Standard variance(°) |
|---|---|---|---|---|
| Pitch angle | 0 | 2.58 | 0.02 | 1.17 |
| Yaw angle | 0 | 2.58 | 0.01 | 1.17 |

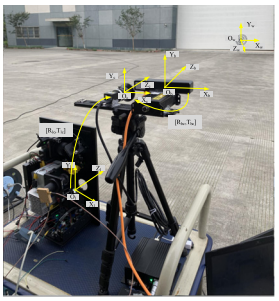

**FIGURE 10.** Experimental platform.

**FIGURE 11.** Some representative images for pose measurement.

The influence of the center extraction noise on the x-axis and y-axis is shown in Fig. 9(a) and Fig. 9(b). The x-axis and y-axis absolute errors have a linear relationship with the distance, and are less affected by the center extraction noise. The z-axis absolute error is slightly affected by the center extraction noise.

In summary, the drogue's feature extraction algorithm is robust under illumination variation and motion blur, and the extraction error is less than 3 pixels within 10 meters. Besides, the pose estimation results are mainly affected by the semi-major axis and semi-minor axis extraction precision based on the simulation of the pose estimation algorithm. The error in the z-axis direction is larger than that of other pose parameters, which is because the monocular vision measurement is not sensitive to the depth. The center's extraction error has a large effect on the positional errors of the x-axis and y-axis direction relative to other parameters, and these positional errors are less than 4.5 cm. In addition, the simulation analysis between the angle extraction error and the pose error is no longer listed, since the length and its own impact on the pose estimation are so small.

### B. GROUND EXPERIMENT ANALYSIS

The experiment platform is composed of LIDAR, a binocular camera, a monocular camera and a drogue target. Since, the effective detection range of the binocular camera is only 8 m, the combination of the laser radar and binocular camera is utilized to obtain the reference pose. The focal length of the monocular measurement camera is 8 mm and the resolution of the camera is $2048 \times 2048$ pixels. The whole experimental platform is shown in Fig. 10.

Some coordinate systems and converted matrixes are shown in Fig. 6. $O_w - X_wY_wZ_w$, $O_l - X_lY_lZ_l$, $O_b - X_bY_bZ_b$ and $O_c - X_cY_cZ_c$ are the coordinate systems of the drogue, laser radar, binocular camera and monocular camera, respectively. $[R_{lc}, T_{lc}]$ is the converted matrix between $O_l - X_lY_lZ_l$ and $O_c - X_cY_cZ_c$, which can be obtained from the joint calibration [54]. $[R_{bc}, T_{bc}]$ is the converted matrix between $O_b - X_bY_bZ_b$ and $O_c - X_cY_cZ_c$, which can be obtained from the multiple camera system calibration [55].

The reference pose of the drogue in the monocular camera coordinate system can be obtained by

$$p_c = [R_{bc}, T_{bc}]\, p_b, \quad p_c = [R_{lc}, T_{lc}]\, p_l \tag{42}$$

Here $p_c$ is the reference pose of the drogue in the camera coordinate system, $p_b$ and $p_l$ are the poses of the drogue in the coordinate system of the binocular camera and laser radar, respectively.

We carried out 103 measurement experiments to verify the validity of the prosed method. These experiments contain the pose of the drogue at different distances and angles. Some representative images from the drogue extraction method are shown in Fig. 11.

The positions were obtained using the method, and were compared to the reference position. The results are shown in Fig. 12. The results show that the proposed algorithm can effectively estimate the position of the drogue within 25 m.

Fig. 13 shows the errors relative to the reference position of the drogue and the relative errors of the drogue, which are defined as the absolute error divided by the reference position on the z-axis. Table 4 shows the pitch angle errors and yaw angle errors. These error results are consistent with the simulation results, and our method can obtain higher precision at the key distance of 10-15 m for aerial refueling. The average error is less than 6 cm on the x-axis and y-axis.

Fig. 13(a) shows the positioning errors different axes. Fig. 13(b) shows the relative positioning errors on different axes.

A dynamic experiment was carried out. The results are shown in Fig. 14. The results show that with the dynamic

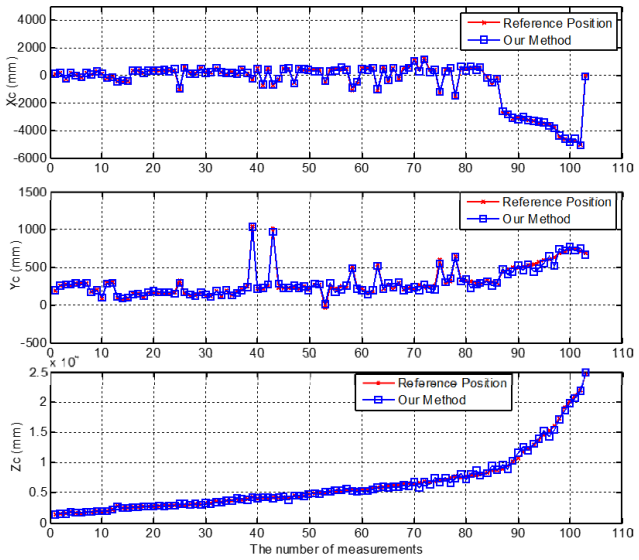


**FIGURE 12.** Position measurement results of the x-axis, y-axis and z-axis.

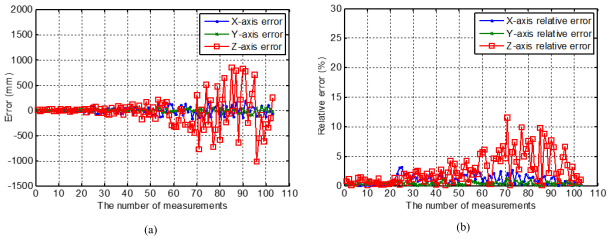


**FIGURE 13.** Errors between the estimated position of the proposed method and the reference position.

**TABLE 4.** Pitch and yaw angle errors.

| Performance | Minimum error (°) | Maximum error (°) | Mean error (°) | Standard variance(°) |
|---|---|---|---|---|
| Pitch angle | 0.43 | 7.21 | 3.34 | 1.92 |
| Yaw angle | 0.21 | 6.32 | 2.97 | 1.57 |

measurement, our algorithm is consistent with the error law that is obtained by the simulation.

Fig. 14(a) shows the dynamic error and displacement of the x-axis. Fig. 14(b) shows the dynamic error and displacement of the y-axis. Fig. 14(c) shows the dynamic error and displacement of the z-axis.

From Fig. 14, we can observe that the error varies with the depth distance. The error is larger with a longer-distance and the smaller with a closer-distance. Moreover, the average error of the x-axis and y-axis is less than 6 cm and the z-axis is less than 15 cm in the measurement range.

To further verify the effectiveness of our measurement method, this method is compared with those introduced by Martinez *et al.* [32], Yin *et al.* [36], Ma and Zhao [22] and Sun *et al.* [23], Where the methods of Martinez *et al.* and Yin *et al.* are simplified pose measurement methods that use

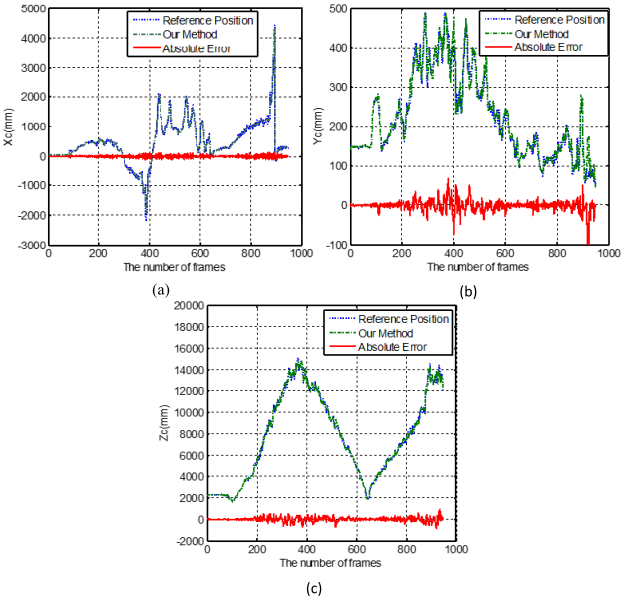


**FIGURE 14.** Results of the dynamic experiment.

**TABLE 5.** Average relative errors of different methods.

| Method | ① | ② | ③ | X-axis | Y-axis | Z-axis | Average |
|---|---|---|---|---|---|---|---|
| Martinez | | | | 1.17% | 1.75% | 4.45% | 2.46% |
| Yin | | | | 2.33% | 2.60% | 10.83% | 5.25% |
| Ma | | | | 1.53% | 2.31% | 2.56% | 2.13% |
| Sun | | | | 1.09% | 1.51% | 1.89% | 1.50% |
| Our | ✓ | | | 1.32% | 2.05% | 2.42% | 1.93% |
| | | ✓ | | 1.13% | 1.93% | 1.78% | 1.61% |
| | | | ✓ | 1.43% | 2.13% | 2.25% | 1.94% |
| | ✓ | ✓ | | 0.89% | 1.03% | 1.65% | 1.19% |
| | ✓ | | ✓ | 1.27% | 1.98% | 2.13% | 1.79% |
| | | ✓ | ✓ | 0.93% | 0.98% | 1.70% | 1.20% |
| | ✓ | ✓ | ✓ | 0.86% | 0.92% | 1.54% | 1.11% |

①:Using Arc-level Extraction, ②:Using Quality Metric, and ③:Using Kalman Filter.

position measurements without considering the angle change of the drogue. They hold the drogue plane perpendicular to the z-axis of the coordinate system of the monocular camera. Besides, the method that was given by Sun *et al.* estimates the position using landmarks that are learned using manual marks, which may cause the error to increase over longer distances. Most of their experiments are measured in a close range distance. Table 5 shows the relative error comparison between our proposed method and other methods over the same distance range. Our method exhibits excellent performance at close range and can also effectively estimate the drogue pose at a long distance. The focal length of the used camera is 8 mm and the resolution of camera is 2048 × 2048 pixels. The algorithm is implemented using a programming language and runs on the NVIDIA TX2 platform. The total time that is required for the algorithm is only 17 ms, and therefore real-time pose estimation can be achieved.

What's more, in Table 5, we also analyze the validity of the key parts of the algorithm. The drogue pose estimation algorithms were compared on the different improvements of key techniques. As shown in Table 5, their average relative errors are 1.93%, 1.61%, 1.94%, 1.19%, 1.79%, 1.20%, and 1.11%. Obviously, the arc-level extraction and the extraction quality metric have greatly improved the average relative errors. The Kalman filter also plays a role in improving the average relative error, and it can smooth the measurement results. Moreover, the relative errors in different directions of our method are lower than those of other methods.

## V. CONCLUSION

This paper proposed a novel method for drogue-UAV relative poses in autonomous aerial refueling based on monocular vision. The method includes two parts. The first is the accurate extraction of the structural features of the drogue, which can be used to estimate the pose of the drogue. The second is the robust pose estimation of the drogue, which is reverse estimated using the two dimensional structural feature and combined with the Kalman filter. To extract the structural features of the drogue, a fast drogue structure extraction algorithm based on the shape characteristics of drogues is designed. In this algorithm, the contour feature of the end plane of the drogue is extracted using the arc-level features which are based on the structural constraints and the extraction quality metric. Additionally, the centroid position of the internal black part of the drogue is extracted using an adaptive mask and threshold segmentation. For the pose estimation, a robust pose measurement algorithm based on monocular vision is proposed that uses the structural features and the Kalman filter. The cone consists of the camera's optical center and the projection of the end plane of the drogue on image plane, which is used to estimate the pose of the drogue in camera coordinate system. In addition, the Kalman filter is used to robustly estimate the pose of the drogue. Finally, the simulation and ground experiment are used to verify the robustness and effectiveness of the method. Future works will be focused on drogue tracking technology, the application of deep learning methods to autonomous aerial refueling and exploring an end-to-end image-to-pose learning method.

## APPENDIX

### A. ERROR ANALYSIS

The real pose of a drogue is defined as $P = [t_x, t_y, t_z, \varphi, \omega]$ on a camera coordinate system. The shape projection of the end plane of the drogue on the image plane can be obtained by the following formulas:

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} n_x & o_x & a_x & t_x \\ n_y & o_y & a_y & t_y \\ n_z & o_z & a_z & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{43}$$

where

$$\begin{aligned} n_x &= cos\,\varphi\, sin\,\varphi - sin\,\varphi\, sin\,\omega\, sin\,\gamma \\ n_y &= cos\,\omega\, sin\,\gamma \\ n_z &= sin\,\varphi\, cos\,\gamma + cos\,\varphi\, sin\,\omega\, sin\,\gamma \\ o_x &= -cos\,\varphi\, sin\,\gamma - sin\,\varphi\, sin\,\omega\, cos\,\gamma \\ o_y &= cos\,\varphi\, cos\,\gamma \\ o_z &= -sin\,\varphi\, cos\,\gamma + cos\,\varphi\, sin\,\omega\, cos\,\gamma \\ a_x &= -sin\,\varphi\, cos\,\omega \\ a_y &= -sin\,\omega \\ a_z &= cos\,\varphi\, cos\,\omega \end{aligned}$$

$$z \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{44}$$

$$x_w^2 + y_w^2 = R^2 \tag{45}$$

From (43), (44), and (45), the general form of the ellipse equation can be obtained.

$$au^2 + bv^2 + cuv + du + ev + f = 0 \tag{46}$$

where

$$\begin{aligned} a &= a_z^2\left(-R^2 + t_x^2 + t_y^2\right) + t_z^2\left(n_x^2 + o_x^2\right) \\ &\quad - 2a_x n_x t_x t_z - 2a_x o_x t_y t_z \\ b &= a_y^2\left(-R^2 + t_x^2 + t_y^2\right) + t_z^2\left(n_y^2 + o_y^2\right) \\ &\quad - 2a_y n_y t_x t_z - 2a_y o_y t_y t_z \\ c &= 2(a_x a_y\left(-R^2 + t_x^2 + t_y^2\right) + t_z^2 o_x o_y - t_x t_z\left(a_x n_y + a_y n_x\right) \\ &\quad - t_y t_z\left(a_x o_y + a_y o_x\right)) \\ d &= 2f_0(a_x a_z\left(-R^2 + t_x^2 + t_y^2\right) + t_z^2\left(n_x n_z + o_x o_z\right) \\ &\quad - t_x t_z\left(a_x n_z + a_z n_x\right) - t_y t_z\left(a_z n_x + a_z o_x\right)) \\ e &= 2f_0(a_y a_z\left(-R^2 + t_x^2 + t_y^2\right) + t_z^2\left(n_y n_z + o_y o_z\right) \\ &\quad - t_x t_z\left(a_y n_z + a_z n_z\right) - t_y t_z\left(a_y o_z + a_z o_y\right)) \\ f &= f_0^2(a_z^2\left(-R^2 + t_x^2 + t_y^2\right) + t_z^2\left(n_z^2 + o_z^2\right) \\ &\quad - 2t_x t_z a_z n_z - 2t_y t_z a_z o_z) \end{aligned}$$

The parameters ($a'$, $b'$, $u_c$, $v_c$ *and* $\rho$) of projection ellipse can be obtained with (8) and (9). The general form of the equation of a cone derived using the above equation is as follows:

$$Ax_c^2 + By_c^2 + Cx_c y_c + Dx_c z_c + Ey_c z_c + Fz_c^2 = 0 \tag{47}$$

where

$$\begin{aligned} A &= f_0^2\left(a'^2 sin\,\rho^2 + b'^2 cos\,\rho^2\right) \\ B &= 2f_0^2\left(b'^2 - a'^2\right) sin\,\rho\, cos\,\rho \\ C &= f_0^2\left(a'^2 cos\,\rho^2 + b'^2 sin\,\rho^2\right) \\ D &= -f_0\left(2Au_c + Bv_c\right) \\ E &= -f_0\left(Bu_c + 2Cv_c\right) \\ F &= Au_c^2 + Bu_c v_c + Cv_c^2 - a'^2 b'^2 \end{aligned}$$

From (22), $\lambda_1$, $\lambda_2$, $\lambda_3$ and T can be used to estimate the pose of the drogue.

$$
\begin{aligned}
T &= (e_1, e_2, e_3) = \begin{bmatrix} e_{1x} & e_{2x} & e_{3x} \\ e_{1y} & e_{2y} & e_{3y} \\ e_{1z} & e_{2z} & e_{3z} \end{bmatrix} \\
x_0 &= e_{1x}R\sqrt{\frac{|\lambda_3|(|\lambda_1|-|\lambda_2|)}{|\lambda_1|(|\lambda_1|+|\lambda_3|)}} + e_{3x}R\sqrt{\frac{|\lambda_1|(|\lambda_2|+|\lambda_3|)}{|\lambda_3|(|\lambda_1|+|\lambda_3|)}} \\
y_0 &= e_{1y}R\sqrt{\frac{|\lambda_3|(|\lambda_1|-|\lambda_2|)}{|\lambda_1|(|\lambda_1|+|\lambda_3|)}} + e_{3y}R\sqrt{\frac{|\lambda_1|(|\lambda_2|+|\lambda_3|)}{|\lambda_3|(|\lambda_1|+|\lambda_3|)}} \\
z_0 &= e_{1z}R\sqrt{\frac{|\lambda_3|(|\lambda_1|-|\lambda_2|)}{|\lambda_1|(|\lambda_1|+|\lambda_3|)}} + e_{3z}R\sqrt{\frac{|\lambda_1|(|\lambda_2|+|\lambda_3|)}{|\lambda_3|(|\lambda_1|+|\lambda_3|)}} \\
V_x &= e_{1x}\sqrt{\frac{|\lambda_1|-|\lambda_2|}{|\lambda_1|+|\lambda_3|}} - e_{3x}\sqrt{\frac{|\lambda_2|+|\lambda_3|)}{|\lambda_1|+|\lambda_3|}} \\
V_x &= e_{1y}\sqrt{\frac{|\lambda_1|-|\lambda_2|}{|\lambda_1|+|\lambda_3|}} - e_{3y}\sqrt{\frac{|\lambda_2|+|\lambda_3|)}{|\lambda_1|+|\lambda_3|}} \\
V_x &= e_{1z}\sqrt{\frac{|\lambda_1|-|\lambda_2|}{|\lambda_1|+|\lambda_3|}} - e_{3z}\sqrt{\frac{|\lambda_2|+|\lambda_3|)}{|\lambda_1|+|\lambda_3|}} \\
\omega_0 &= tan^{-1}V_y \\
\varphi_0 &= tan^{-1}\frac{V_x}{V_z} \\
F &= \begin{bmatrix} F_1 = x_0 \\ F_2 = y_0 \\ F_3 = z_0 \\ F_4 = \omega_0 \\ F_5 = \varphi_0 \end{bmatrix}
\end{aligned}
\tag{48}\tag{49}
$$

According to the above formulas, the analytical solution of the estimated drogue pose was achieved. The error of the estimated drogue pose is defined as $PN = [e^x, e^y, e^z, e^\omega, e^\varphi]^T$. The elliptical shape of the end plane of the drogue is defined as $I = [a', b', u_c, v_c, \rho]^T$. The extraction error is defined as $DI = [e^{a'}, e^{b'}, e^{u_c}, e^{v_c}, e^\rho]$.

$$PN = \frac{\partial F}{\partial I} \times DI \tag{50}$$

The analytical form of the error can be obtained from (50) to analyze the robustness of the system to image extraction errors at different distances. However, because the analytical formulas are very complicated, they are impossible to use to directly analyze the variation of the pose error. In the text, the error analysis was performed through simulations under general conditions.

## ACKNOWLEDGMENT

The authors would like thank the accompaniers working with them in department of the Institute of Optics and Electronics, Chinese Academy of Sciences. *(Yuebo Ma and Rujin Zhao contributed equally to this work.)*

## REFERENCES

[1] L.-N. C. Hugh, "Apparatus for aircraft-refuelling in flight and aircraft-towing," U.S Patent 2 716 527, Aug. 30, 1955.
[2] C. J. Leisy, "Aircraft interconnecting mechanism," U.S Patent 2 663 523, Dec. 22, 1953.
[3] R. Korbly, "Sensing relative attitudes for automatic docking," *AIAA J. Guid. Control Dyn.*, vol. 6, no. 3, pp. 213–215, 1983.
[4] A. Al-Kaff, D. Martín, F. García, A. de la Escalera, and J. M. Armingol, "Survey of computer vision algorithms and applications for unmanned aerial vehicles," *Expert Syst. Appl.*, vol. 92, pp. 447–463, Feb. 2018. doi: 10.1016/j.eswa.2017.09.033.
[5] Y. Sun, H. Duan, X. Xu, and Y. Deng, "Bionic visual close-range navigation control system for the docking stage of probe-and-drogue autonomous aerial refueling," *Aerosp. Sci. Technol.*, vol. 91, pp. 136–149, Aug. 2019.
[6] Z. Paulson, S. Nykl, J. Pecarina, and B. Woolley, "Mitigating the effects of boom occlusion on automated aerial refueling through shadow volumes," *J. Defense Model. Simul.*, vol. 16, no. 2, pp. 175–189, Apr. 2019.
[7] Y. Zhu, Y. Sun, B. Huang, L. Wu, and W. Zhao, "Relative navigation for autonomous aerial refueling rendezvous phase," *Optik*, vol. 174, pp. 665–675, Dec. 2018.
[8] D. T. Johnson, S. L. Nykl, and J. F. Raquet, "Combining stereo vision and inertial navigation for automated aerial refueling," *J. Guid., Control, Dyn.*, vol. 40, no. 9, pp. 2250–2259, May 2017.
[9] W. Xufeng, L. Jianmin, X. Dong, B. Zhang, and K. Xingwei, "An approach to mathematical modeling and estimation of probe-drogue docking success probability for UAV autonomous aerial refueling," *Int. J. Aerosp. Eng.*, vol. 2017, Aug. 2017, Art. no. 6427209.
[10] J. Li, "Binocular vision measurement method for relative position and attitude based on dual-quaternion," *J. Modern Opt.*, vol. 64, no. 18, pp. 1846–1853, Apr. 2017.
[11] S. Chen, H. Duan, C. Li, G. Zhao, Y. Xu, and Y. Deng, "Drogue pose estimation for unmanned aerial vehicle autonomous aerial refueling system based on infrared vision sensor," *Opt. Eng.*, vol. 56, no. 12, Dec. 2017, Art. no. 124105.
[12] A. Tsukerman, M. Weiss, D. Löbl, F. Holzapfel, and T. Y. Shima, "Trajectory shaping autopilot-guidance design for civil autonomous aerial refueling," in *Proc. AIAA Guid., Navigat., Control Conf.*, Jan. 2017, p. 1025.
[13] Y. Xu, H. Duan, and C. Li, "On-board visual navigation system for unmanned aerial vehicles autonomous aerial refueling," *Proc. Inst. Mech. Eng., G, J. Aerosp. Eng.*, vol. 233, no. 4, pp. 1193–1203, Dec. 2017.
[14] D. T. Johnson, S. L. Nykl, and J. F. Raquet, "Combining stereo vision and inertial navigation for automated aerial refueling," *J. Guid., Control, Dyn.*, vol. 40, no. 9, pp. 1–10, May 2017.
[15] C. Parsons, Z. Paulson, W. Dallman, B. G. Woolley, J. Pecarina, and S. Nykl, "Analysis of simulated imagery for real-time vision-based automated aerial refueling," *J. Aerosp. Inf. Syst.*, vol. 16, no. 3, pp. 77–93, Jan. 2019.
[16] D. Luo, J. Shao, J. Zhang, and Y. Xu, "Docking navigation method for UAV autonomous aerial refueling," *Sci. China Inf. Sci.*, vol. 62, no. 1, Jan. 2019, Art. no. 10203.
[17] C. I. Chen, R. Koseluk, C. Buchanan, A. Duerner, B. Jeppesen, and H. Laux, "Autonomous aerial refueling ground test demonstration—A sensor-in-the-loop, non-tracking method," *Sensors*, vol. 15, no. 5, pp. 10948–10972, May 2015.
[18] C. I. Chen and R. Stettner, "Drogue tracking using 3D flash LiDAR for autonomous aerial refueling," *Proc. SPIE*, vol. 8037, Jun. 2011, Art. no. 80370Q.
[19] H. Li and H. B. Duan, "Verification of monocular and binocular pose estimation algorithms in vision-based UAVs autonomous aerial refueling system," *Sci. China Technol. Sci.*, vol. 59, no. 11, pp. 1730–1738, 2016.
[20] L. Xin, D. Luo, and H. Li, "A monocular visual measurement system for UAV probe-and-drogue autonomous aerial refueling," *Int. J. Intell. Comput. Cybern.*, vol. 11, no. 6, pp. 166–180, Jun. 2018.
[21] B. Huang, Y.-R. Sun, J.-Y. Liu, and X.-D. Sun, "Circular drogue pose estimation for vision-based navigation in autonomous aerial refueling," in *Proc. IEEE Chin. Guid., Navigat. Control Conf.*, Aug. 2016, pp. 960–965.
[22] Y. Ma, E. Liu, Z. Zhang, K. Yan, and R. Zhao, "A novel autonomous aerial refueling drogue detection and pose estimation method based on monocular vision," *Measurement*, vol. 136, pp. 134–142, Mar. 2019.

[23] S. Sun, Y. Yin, D. Xu, and X. Wang, "Robust landmark detection and position measurement based on monocular vision for autonomous aerial refueling of UAVs," *IEEE Trans. Cybern.*, vol. 49, no. 12, pp. 4167–4179, Dec. 2019.
[24] J. Valasek, K. Gunnam, J. Kimmett, M. D. Tandale, J. L. Junkins, and D. Hughes, "Vision-based sensor and navigation system for autonomous air refueling," in *Proc. 1st AIAA Unmanned Aerosp. Vehicles, Syst., Technol., Oper. Conf. Exhibit*, May 2002, pp. 20–22.
[25] J. Kimmett, J. Valasek, and J. L. Junkins, "Autonomous aerial refueling utilizing a vision based navigation system," in *Proc. AIAA Guid., Navigat., Control Conf. Exhibit*, Monterey, CA, USA, Aug. 2002, pp. 5–8.
[26] J. Kimmett, J. Valasek, and J. L. Junkins, "Vision based controller for autonomous aerial refueling," in *Proc. IEEE Int. Conf. Control Appl.*, Glasgow, U.K., Sep. 2002, pp. 1138–1143.
[27] M. Tandale, R. Bowers, and J. L. Valasek, "Robust trajectory tracking controller for vision based probe and drogue autonomous aerial refueling," in *Proc. AIAA Guid., Navigat., Control Conf. Exhibit*, San Francisco, CA, USA, Aug. 2005, pp.846–857.
[28] L. Pollini, R. Mati, and M. Innocenti, "Experimental evaluation of vision algorithms for formation flight and aerial refueling," in *Proc. AIAA Modeling Simulation Guid. Control*, Providence, RI, USA, Aug. 2004, p. 4918.
[29] C.-P. Lu, G. D. Hager, and E. Mjolsness, "Fast and globally convergent pose estimation from video images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 6, pp. 610–622, Jun. 2000.
[30] W. Xufeng, D. Xinmin, and K. Xingwei, "Feature recognition and tracking of aircraft tanker and refueling drogue for UAV aerial refueling," in *Proc. 25th IEEE Chin. Control Decis. Conf.*, May 2013, pp. 2057–2062.
[31] X. Wang, K. Xingwei, Z. Jianhui, C. Yong, and D. Xinmin, "Real-time drogue recognition and 3D locating for UAV autonomous aerial refueling based on monocular machine vision," *Chin. J. Aeronaut.*, vol. 28, no. 6, pp. 1667–1675, Dec. 2015.
[32] C. Martínez, T. Richardson, P. Thomas, J. L. du Bois, and P. Campoy, "A vision-based strategy for autonomous aerial refueling tasks," *Robot. Auton. Syst.*, vol. 61, no. 8, pp. 876–895, 2013.
[33] C. Martínez, T. Richardson, and P. Campoy, "Towards autonomous air-to-air refuelling for UAVs using visual information," in *Proc. IEEE Int. Conf. Robot. Automat.*, Karlsruhe, Germany, May 2013, pp. 5736–5742.
[34] H. Duan and Q. Zhang, "Visual measurement in simulation environment for vision-based UAV autonomous aerial refueling," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 9, pp. 2468–2480, Sep. 2015.
[35] M. Ding, L. Wei, and B. Wang, "Vision-based estimation of relative pose in autonomous aerial refueling," *Chin. J. Aeronautics*, vol. 24, no. 6, pp. 807–815, Dec. 2011.
[36] Y. Yin, D. Xu, X. G. Wang, and M. R. Bai, "Detection and tracking strategies for autonomous aerial refuelling tasks based on monocular vision," *Int. J. Adv. Robotic Syst.*, vol. 11, no. 1, pp. 399–412, 2014.
[37] Z. Chen and J.-B. Huang, "A vision-based method for the circle pose determination with a direct geometric interpretation," *IEEE Trans. Robot. Autom.*, vol. 15, no. 6, pp. 1135–1140, Dec. 1999.
[38] Y. C. Shiu and S. Ahmad, "3D location of circular and spherical features by monocular model-based vision," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, vol. 2, Nov. 1989, pp. 576–581.
[39] R. O. Duda and R. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Commun. ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972.
[40] M. Fornaciari, R. Cucchiara, and A. Prati, "A mobile vision system for fast and accurate ellipse detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2013, pp. 52–53.
[41] M. Fornaciari and A. Prati, "Very fast ellipse detection for embedded vision applications," in *Proc. 6th Int. Conf. Distrib. Smart Cameras*, Oct./Nov. 2013, pp. 1–6.
[42] F. Mai, Y. S. Hung, H. Zhong, and W. F. Sze, "A hierarchical approach for fast and robust ellipse extraction," *Pattern Recognit.*, vol. 41, no. 8, pp. 2512–2524, Aug. 2008.
[43] X. Bai, C. Sun, and F. Zhou, "Splitting touching cells based on concave points and ellipse fitting," *Pattern Recognit.*, vol. 42, no. 11, pp. 2434–2446, 2009.
[44] Z.-Y. Liu and H. Qiao, "Multiple ellipses detection in noisy environments: A hierarchical approach," *Pattern Recognit.*, vol. 42, no. 11, pp. 2421–2433, Nov. 2009.
[45] H. Dong, D. K. Prasad, and I.-M. Chen, "Accurate detection of ellipses with false detection control at video rates using a gradient analysis," *Pattern Recognit.*, vol. 81, pp. 112–130, Sep. 2018.
[46] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
[47] D. K. Prasad, M. K. H. Leung, and C. Quek, "ElliFit: An unconstrained, non-iterative, least squares based geometric ellipse fitting method," *Pattern Recognit.*, vol. 46, no. 5, pp. 1449–1465, May 2013.
[48] S. Chen, R. Xia, Y. Chen, M. Hu, and J. Zhao, "A hybrid method for ellipse detection in industrial images," *Pattern Recognit.*, vol. 68, pp. 82–98, Aug. 2017.
[49] D. R. Radev, H. Jing, D. Tam, and M. Styś, "Centroid-based summarization of multiple documents," *Inf. Process. Manage.*, vol. 40, no. 6, pp. 919–938, Nov. 2004.
[50] Z. Zhao and Y. Liu, "Applications of projected circle centers in camera calibration," *Mach. Vis. Appl.*, vol. 21, no. 3, pp. 301–307, 2010.
[51] Y. Wu, H. Zhu, Z. Hu, and F. Wu, "Camera calibration from the quasi-affine invariance of two parallel circles," in *Computer Vision—ECCV*. Berlin, Germany: Springer, 2004, pp. 190–202.
[52] C. Colombo, A. D. Bimbo, and F. Pernici, "Metric 3D reconstruction and texture acquisition of surfaces of revolution from a single uncalibrated view," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 1, pp. 99–114, Jan. 2005.
[53] R. Safaee-Rad, I. Tchoukanov, K. C. Smith, and B. Benhabib, "Three-dimensional location estimation of circular features for machine vision," *IEEE Trans. Robot. Autom.*, vol. 8, no. 5, pp. 624–640, Oct. 1992.
[54] Y. Park, S. Yun, C. S. Won, K. Cho, K. Um, and S. Sim, "Calibration between color camera and 3D LIDAR instruments with a polygonal planar board," *Sensors*, vol. 14, no. 3, pp. 5333–5353, 2014.
[55] B. Li, L. Heng, K. Koser, and M. Pollefeys, "A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Tokyo, Japan, Nov. 2013, pp. 1301–1307.

**YUEBO MA** received the B.S. degree in mechanical engineering from the Jincheng College of Sichuan University, Chengdu, China, in 2014, and the M.S. degree in automobile engineering from Xi Hua University, Chengdu, in 2017. He is currently pursuing the Ph.D. degree with the Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu.

His current research interests include object detection, object tracking, measurement, computer vision, and deep learning.

**RUJIN ZHAO** received the B.S. and M.S. degrees from the Southwest University of Science and Technology, Mianyang, China, and the Ph.D. degree from the Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, China.

He is currently a Researcher with the Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu. His current research interests include visual measurement, SLAM, space robotic detection, and so on. At present, he has published relevant articles, and is a member and peer reviewer of several journals. He was supported by the National Natural Science Foundation of China and the Youth Innovation Promotion Association CAS.

**ENHAI LIU** is a Researcher and Doctoral Tutor. He is currently the Deputy Director of the Institute of Optoelectronic Technology, Chinese Academy of Sciences.

He is a member of the Academic and Academic Degree Committee and is involved in research and engineering research on photoelectric precision measurement and automatic control technology. His research interests include space optoelectronic precision measurement, photodetector application and photodetection technology, signal and information processing, optoelectronic measurement system error theory analysis, system integration technology research, and so on. He has undertaken and completed a number of manned spaceflight and lunar exploration in related fields. The research (development) of the engineering, 863, and 973 engineering projects won four awards for provincial and ministerial level scientific and technological progress. He is a member of the Chinese Optical Society, the Optical Engineering Society, and the Space Optical Engineering Society.

**ZHUANG ZHANG** received the B.S. degree from the Hebei University of Science and Technology, Shijiazhuang, China, in 2013, and the Ph.D. degree from the Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, China, in 2019.

His current research interests include visual SLAM, motion detection, 3D reconstruction, sensor fusion, and so on.

**KUN YAN** received the B.S. degree from Sichuan Normal University, Chengdu, China, in 2013, and the Ph.D. degree from the Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, in 2018.

His research interests include binocular vision, computer vision, pose estimation, photogrammetry, and image processing.