SCIENCE CHINA
Information Sciences

January 2019, Vol. 62 010203:1–010203:16
https://doi.org/10.1007/s11432-018-9578-9

• RESEARCH PAPER •
Special Focus on Guidance, Navigation and Control

Docking navigation method for UAV autonomous aerial refueling

Delin LUO[1*], Jiang SHAO[1], Yang XU[1] & Jinpeng ZHANG[2,3]

[1]*School of Aerospace Engineering, Xiamen University, Xiamen* 361005*, China;*
[2]*China Airborne Missile Academy, Luoyang* 471009*, China;*
[3]*Aviation Key Laboratory of Science Technology on Airborne Guided Weapon, Luoyang* 471009*, China*

Received 2 July 2018/Accepted 19 July 2018/Published online 21 December 2018

Abstract In this paper, a docking navigation method for autonomous aerial refueling (AAR) of unmanned aerial vehicles (UAVs) based on a binocular vision system (BVS) is proposed. A BVS simulation platform is built for simulation research purposes. First, unnecessary scene information in the image is filtered through green light-emitting diodes (LEDs) and filters. Then the image is processed via graying, binarization, and median filtering to highlight the connected area of the LED in the image. Subsequently, the center of mass of the connected area is selected as the feature point (FP), and the FPs are described using an improved Haar wavelet transform. The multidimensional description vector of FP is obtained and matched. Finally, the position and pose of the refueling cone sleeve are estimated. Simulation results show the effectiveness of the presented AAR navigation method.

Keywords autonomous aerial refueling, AAR, position and pose estimation, improved Haar wavelet transform, IHWT, binocular vision system, BVS, BVS simulation platform

Citation Luo D L, Shao J, Xu Y, et al. Docking navigation method for UAV autonomous aerial refueling. Sci China Inf Sci, 2019, 62(1): 010203, https://doi.org/10.1007/s11432-018-9578-9

1 Introduction

Unmanned aerial vehicles (UAVs) have natural advantages over manned aircraft in terms of hardware cost, maneuverability and control performance, environmental adaptability, etc. With the improvement of UAV technology, UAVs will replace manned aircraft in carrying out air combat missions, intercepting enemy aircraft, or fighting for the control of air spaces [1, 2]. However, the limited amount of portable fuel seriously restricts the range, payload, and combat capability of UAVs [3]. It is necessary to draw lessons from the man-machine aerial refueling technology. Autonomous aerial refueling (AAR) of UAVs is a critical area of technology in the field of UAV application [4, 5]. In an airborne UAV, the refueling cone sleeve (RCS) constantly swings in the air owing to the effects of the swirling field and atmospheric turbulence at the end of the tanker aircraft [6]. Therefore, the key aspect of the technology is that UAVs must acquire accurate position and pose information of the RCS in real time and thereafter continue to approach it to achieve accurate docking. UAV refueling can be divided into four stages: rendezvous, close-range docking, refueling, and separation. Their close connection constitutes a set of safe and effective UAV AAR processes. Traditional aeronautical positioning systems, such as the inertial navigation system, global positioning system, and differential global positioning system, are generally used for long-distance measurement with low precision and poor anti-jamming ability.

* Corresponding author (email: luodelin1204@xmu.edu.cn)

Currently, a navigation method based on visual sensors is widely used in the close-range docking stage of aerial refueling. The active vision-based navigation systems developed by researchers at Texas A&M University can provide six higher-precision degrees of freedom navigation information [7, 8]. However, the accuracy of the system is dependent on the sensor and is prone to interference. Pollini et al. [9, 10] developed a relative navigation system based on passive visual sensing. In this system, a charge-coupled device camera with a filter was used for imaging and an orthogonal iterative algorithm was used for calculating the relative position information between the UAV and the RCS. However, the system has some limitations such as poor stability and "position and pose blur". The robust pose estimation algorithm proposed in [11] effectively avoids the problem of "position and pose blur" and improves the accuracy of position and pose estimation. A novel method with an inverse projection ray approach and a new tracking method with a 3D flash light detection and ranging sensor were established to determine the relative position and pose between the UAV and RCS in [12, 13]. A new visual pose estimation method based on a circle feature was proposed in [14]. This method considers a conical refueling port as the visual feature and does not require optical marking points. The relative error in the experiment is less than 6%. In [15], Duan et al. designed a hardware-in-loop simulation platform based on eagle-eye vision mechanism. The relative pose of the RCS and UAV can be measured accurately by using the perspective 3 points algorithm. Subsequently, the vision system designed for boom approach AAR was tested; moreover, an integrated platform was built and the controller was given in [16]. In [6], through wind tunnel experiments, Bloy et al. studied in detail the aerodynamic interference of tanker aircraft and UAV during refueling in the air. In [17], a multiaxis industrial robot was used to test the numerical model of UAV flight environment, which provided a useful tool for studying a multibody contact impact problem such as AAR.

Inspired by the above discussion, this study provides the following contributions: (i) A binocular vision system (BVS) simulation platform is designed, the parameter matrices are calibrated, and the camera distortion coefficient is calculated. (ii) An integral image can simplify the process of obtaining the sum of the pixels in the template, and the description of feature points (FPs) becomes more detailed and accurate with the use of an improved Haar wavelet transform. (iii) The preprocessing of the image, application of a one-step ergodic algorithm, calculation of integral image, improvement of Haar wavelet transform, and minimization of Euclidean distance can accurately estimate the relative position and pose between the RCS and UAV.

The paper is organized in the following manner. In Section 2, the principle of binocular vision is analyzed using coordinate transformation. In Section 3, a BVS simulation platform is built and the parameters of the BVS are calibrated. In Section 4, the image is preprocessed and the FPs are extracted and matched. In Section 5, the position and pose information of the RCS is obtained by fitting the matching coordinates of the FP. In Section 6, a docking and tracking control system for UAV AAR is described and simulated. The conclusion is given in the final section.

2 BVS model

The AAR process and the related coordinate system (CS) of UAV based on BVS are shown in Figure 1. P_1 is the probe of UAV and P_2 is the RCS. The tanker aircraft body CS is given by $O_b - x_b y_b z_b$, the UAV body CS by $O_{bo} - x_{bo} y_{bo} z_{bo}$, the camera CS by $O_c - x_c y_c z_c$, the RCS CS by $O_d - x_d y_d z_d$, and the world CS by $O_w - X_w Y_w Z_w$. The left camera CS is chosen as the world CS in this study. Only by transforming the coordinates of each CS into the same CS (the world CS), the relative position of the UAV and the RCS in the process of UAV AAR can be described accurately and uniformly. The UAV AAR system calculates the position and pose information of P_2 in real time, and controls the flight pose of the UAV to complete the precise docking between P_1 and P_2.

The principle of a convergent BVS is shown in Figure 2. The target object is required to be in the field-of-view overlap area of the left and right cameras. The left and right cameras shoot the spatial point P simultaneously and image P_l and P_r, respectively, in the left and right images. The three-dimensional

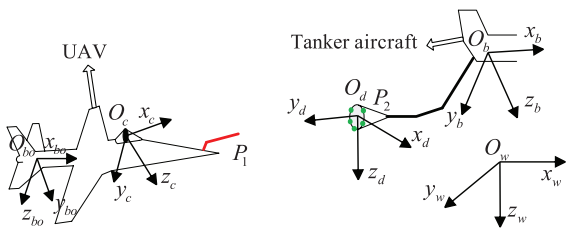


Figure 1 (Color online) AAR diagram of UAV.

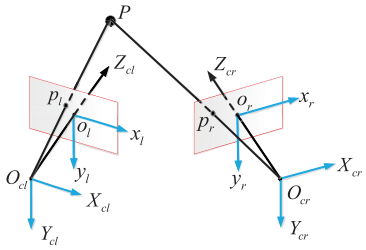


Figure 2 (Color online) Principle of convergent binocular vision.

(3D) coordinates of point P can be uniquely determined. According to the imaging principle of the camera, we can obtain the following relationship:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z_c} \begin{bmatrix} 1/d_x & 0 & u_0 \\ 0 & 1/d_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \frac{1}{z_c} \begin{bmatrix} f/d_x & 0 & u_0 \\ 0 & f/d_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix}, \tag{1}$$

where (u, v) is a point of the image in the pixel coordinates, (u_0, v_0) represents the pixel coordinates of the origin of the image physical CS, (x, y) represents the physical coordinates of a point in the image, d_x and d_y are the physical dimensions of the minimum pixel of the camera, (x_c, y_c, z_c) represents the camera coordinates, and f is the camera focal length. BVS is a kind of non-contact 3D measurement technology, which has the advantages of robust real-time operation, high precision, and strong anti-jamming ability. The main factors that affect the coordinate precision are given as follows:

- the internal and external parameters of the BVS;
- accuracy of FP extraction;
- matching accuracy of FPs.

The internal parameter matrix $\boldsymbol{A}$ of the camera can be defined as

$$\boldsymbol{A} = \begin{bmatrix} f/d_x & \gamma & u_0 \\ 0 & f/d_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} f_x & \gamma & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{2}$$

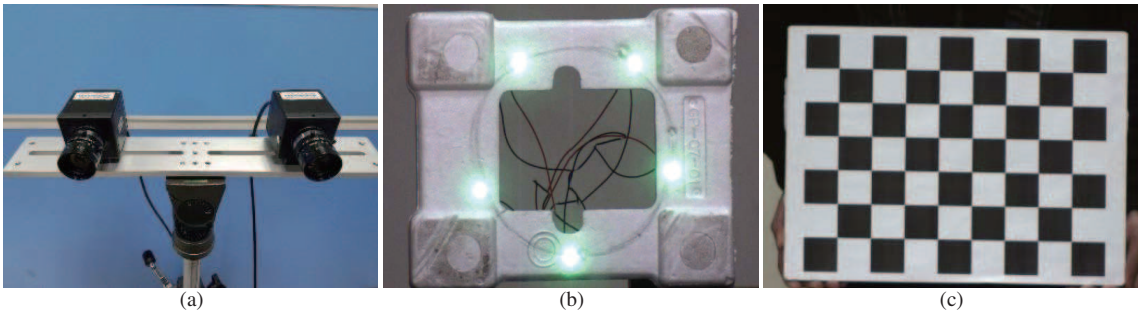
where $f_x = f/d_x, f_y = f/d_y, \gamma$ is the non-vertical factor of two axes of the camera image CS [18]. The conversion from right CS to world CS can be expressed as

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \boldsymbol{M}_{lr} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 & t_x \\ r_4 & r_5 & r_6 & t_y \\ r_7 & r_8 & r_9 & t_z \end{bmatrix}, \tag{3}$$

where $\boldsymbol{M}_{lr} = [\boldsymbol{R}\ \boldsymbol{T}]$ is defined as the external parameter matrix of BVS, $\boldsymbol{R}$ and $\boldsymbol{T}$ are the rotation matrix and translation matrix, respectively. Therefore, the conversion relationship between the left and right camera CS and the world CS is given by

$$z_l \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = \boldsymbol{A}_l \boldsymbol{M}_l \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = \boldsymbol{H}_l \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = \begin{bmatrix} h_{l11} & h_{l12} & h_{l13} & h_{l14} \\ h_{l21} & h_{l22} & h_{l23} & h_{l24} \\ h_{l31} & h_{l32} & h_{l33} & h_{l34} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix}, \tag{4}$$

$$z_r \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = \boldsymbol{A}_r \boldsymbol{M}_r \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = \boldsymbol{H}_r \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = \begin{bmatrix} h_{r11} & h_{r12} & h_{r13} & h_{r14} \\ h_{r21} & h_{r22} & h_{r23} & h_{r24} \\ h_{r31} & h_{r32} & h_{r33} & h_{r34} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix}, \tag{5}$$



(a) (b) (c)

Figure 3 (Color online) UAV autonomous refueling platform. (a) BVS; (b) simulated RCS; (c) CCP.

where (x_w, y_w, z_w) is the world coordinate value of the point P; (u_l, v_l) and (u_r, v_r) are the pixel coordinates of the point P in the pixel CS of the left and right images, respectively; z_l and z_r are proportional coefficients; $\boldsymbol{A}_l, \boldsymbol{A}_r$ and $\boldsymbol{M}_l, \boldsymbol{M}_r$ are the internal and external parameter matrices of the left and right cameras, respectively, which should be obtained via the calibration of BVS; $\boldsymbol{H}_l$ and $\boldsymbol{H}_r$ are mapping relation matrices. Therefore, the process of obtaining 3D coordinates (O3C) can be summarized as Algorithm 1.

Algorithm 1 O3C algorithm

1: Calibration of BVS. $\boldsymbol{A}_l, \boldsymbol{A}_r$ and $\boldsymbol{M}_l, \boldsymbol{M}_r$ are obtained from the left and right cameras of BVS, respectively.
2: Image acquisition. The left and right cameras of BVS are used to capture the same object simultaneously and the captured images are stored in the form of digital images in the computer.
3: Image preprocessing. Graying, binarization, and median filtering are applied to the image.
4: Extraction of image FPs. Pixels containing important features of the image are extracted as FPs.
5: FP matching. The corresponding relationship between the left and right image FPs is determined.
6: 3D coordinate calculation. Based on the binocular vision principle, the 3D coordinates of matching FPs are obtained.

3 Calibration of BVS parameters

3.1 Construction of UAV autonomous refueling system based on binocular vision

The BVS simulation platform is shown in Figure 3. It is mainly composed of BVS, simulated RCS, and checkerboard calibration plate (CCP). The camera is placed in parallel to a frame that can freely rotate in 360° field of view. The cameras are equipped with filters that allow only green light to pass through. Five green light-emitting diodes (LEDs) with the wavelength of 500 nm are installed on the bottom end of the RCS to highlight the imaging area of RCS and to weaken the other scene information. The camera and RCS are connected by a computer, which simultaneously operates two cameras and shoots, saves, and processes the captured images.

Owing to the presence of other background objects in the images captured by the BVS in addition to the RCS, these objects will interfere with the extraction of the feature information of RCS by the BVS. Therefore, installing green LEDs and filters allows only green light to pass through. This approach can highlight the imaging area of the RCS and weaken the other scene information. It is propitious to segment and preprocess the image of LED and other imaging areas, thus simplifying the extraction of image feature information.

3.2 Calculation of calibration parameters of the BVS

(1) The system is calibrated using toolbox_calib based on a calibration method in MATLAB in [19]. A CCP is photographed from a plurality of angles from the BVS. Nine groups of pictures such as those shown in Figures 4(a) and (b) are collected and loaded into the calibration program. The pixel coordinates (u_i, v_i) of the left image group and the right image group, respectively, $i = 1, 2, \ldots, n$, are extracted as shown in Figure 4(c).

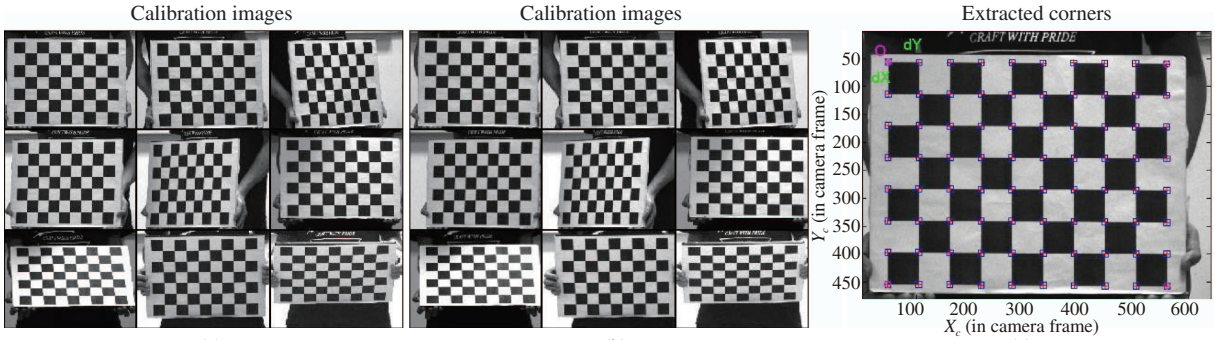


Figure 4 (Color online) CCP images. (a) CCP images captured by the left camera; (b) CCP images captured by the right camera; (c) sample drawing of the fixed point of CCP image.

Table 1 Internal parameters of the BVC

Internal parameter	Left camera	Right camera
f_x	2394.3558	2325.5075
f_y	2393.9677	2325.2000
u_0	692.7227	628.8684
v_0	590.8281	555.3041
k_1	0.0708	0.0661
k_2	−0.1293	0.0147
γ	−4.5025	−0.4948

(2) Calibration of internal and external parameter matrices. The following is the conversion relation on the $Z_w = 0$ plane of CCP.

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \boldsymbol{AM}\begin{bmatrix} X_w \\ Y_w \\ 0 \\ 1 \end{bmatrix} = \boldsymbol{A}\begin{bmatrix} \boldsymbol{R}\ \boldsymbol{T} \end{bmatrix}\begin{bmatrix} X_w \\ Y_w \\ 0 \\ 1 \end{bmatrix} = \boldsymbol{A}\begin{bmatrix} r_1\ r_2\ r_3\ \boldsymbol{T} \end{bmatrix}\begin{bmatrix} X_w \\ Y_w \\ 0 \\ 1 \end{bmatrix} = \boldsymbol{A}\begin{bmatrix} r_1\ r_2\ \boldsymbol{T} \end{bmatrix}\begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix}, \tag{6}$$

where s is a nonzero scale factor and (X_w, Y_w, Z_w) is the world coordinate value.

(3) Calculation of camera distortion coefficients. The points in the actual imaging plane are translated to make them coincident with the imaging points in the ideal imaging plane. The mathematical expressions of the actual physical imaging coordinates and the actual pixel coordinates are as follows:

$$\begin{cases} \tilde{x}_n = x_n[1 + k_1(x_n^2 + y_n^2) + k_2(x_n^2 + y_n^2)^2], \\ \tilde{y}_n = y_n[1 + k_1(x_n^2 + y_n^2) + k_2(x_n^2 + y_n^2)^2], \end{cases} \tag{7}$$

$$\begin{bmatrix} (u - u_0)(x_n^2 + y_n^2) & (u - u_0)(x_n^2 + y_n^2)^2 \\ (v - v_0)(x_n^2 + y_n^2) & (v - v_0)(x_n^2 + y_n^2)^2 \end{bmatrix}\begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \tilde{u} - u \\ \tilde{v} - v \end{bmatrix}, \tag{8}$$

where (u, v) indicates the ideal imaging pixel coordinates, $(\tilde{u}, \tilde{v})$ indicates the actual imaging pixel coordinates, (x_n, y_n) indicates the ideal imaging physical coordinates, and $(\tilde{x}_n, \tilde{y}_n)$ indicates the actual imaging physical coordinates. The maximum likelihood estimation method is used to optimize and calibrate these matrix parameters. The matrix of the internal and external parameters derived from the above steps is shown in Tables 1 and 2.

Table 2 External parameters of the BVC

External parameter	Calibration matrix of BVC
Rotation matrix $\boldsymbol{R}$	$\begin{bmatrix} 0.9986 & 0.0006 & 0.0525 \\ -0.0006 & 1.0000 & -0.0002 \\ -0.0525 & 0.0002 & 0.9986 \end{bmatrix}$
Translation matrix $\boldsymbol{T}$	$\begin{bmatrix} -111.08 & 2.9456 & -25.069 \end{bmatrix}^{\mathrm{T}}$

4 Processing of images and the extraction and matching of the FPs

4.1 Preprocessing of the image

To extract the FP information efficiently, it is necessary to preprocess the original image to enhance the useful information, weaken the unnecessary scene information, and remove the noise points in the image.

(1) Graying the color image. Color images of three channels (red, green, and blue) are converted into single-channel grayscale images. The conversion is represented as

$$\mathrm{Gray} = R \times 0.299 + G \times 0.587 + B \times 0.114, \tag{9}$$

where Gray is the pixel value of the pixel after conversion to a grayscale image, R, G, and B are the red values, green values, and blue values in the color three-channel images, respectively. The image converted to grayscale is shown in Figure 5(b).

(2) Binarization to the grayscale image. The threshold Ts is compared with the gray value $I(i,j)$ of all the pixels in the gray image. The pixel grayscale value greater than the threshold is changed to 255 and that less than the threshold is changed to 0. The mathematical model can be expressed as

$$I(i,j) = \begin{cases} 255, & I(i,j) \geqslant \mathrm{Ts}, \\ 0, & I(i,j) < \mathrm{Ts}. \end{cases} \tag{10}$$

After binarization segmentation, Figure 5(b) is converted to a binarization image as shown in Figure 5(c). Except for the white image area of the LEDs, all the other imaging areas are black.

(3) Median filtering of image. From Figure 5(c), the imaging area of LED is observed to be serrated and some image areas are even surrounded by single granular image noise points. Therefore, the median filter method is used to filter the median value of Figure 5(c), which is shown in Figure 5(d). The noise points of the zigzag and granular images around the imaging area of the LEDs disappear and the image after median filtering is smoother than the original one. Thus, it is easier to extract the centroid of the imaging areas of the LEDs as FPs.

4.2 Extraction of FPs

In this study, an one-step ergodic (OSE) algorithm [20] based on the area growth method is used to search and mark the connected areas with FPs as feature information, which is expressed as Algorithm 2. The algorithm only needs to traverse the image once to mark all the connected areas in the image, which can reduce the computational cost of searching for the connected areas of the image. The results of the lookup and tagging are shown in Figure 6. Each connected area in the graph is surrounded by a red box. The four edges of the red border are the row or column of the connected area at the edge of the pixel in the top, bottom, left, and right directions. Subsequently, the pixel coordinates in each connected area are extracted from the tag matrix obtained by searching the marked connected area, and the centroid of the connected area is obtained as follows:

$$X_c = \frac{\sum_{i=0}^{x_{\max}} P_i x_i}{\sum_{i=0}^{x_{\max}} P_i}, \quad Y_c = \frac{\sum_{i=0}^{y_{\max}} P_i y_i}{\sum_{i=0}^{y_{\max}} P_i}, \tag{11}$$

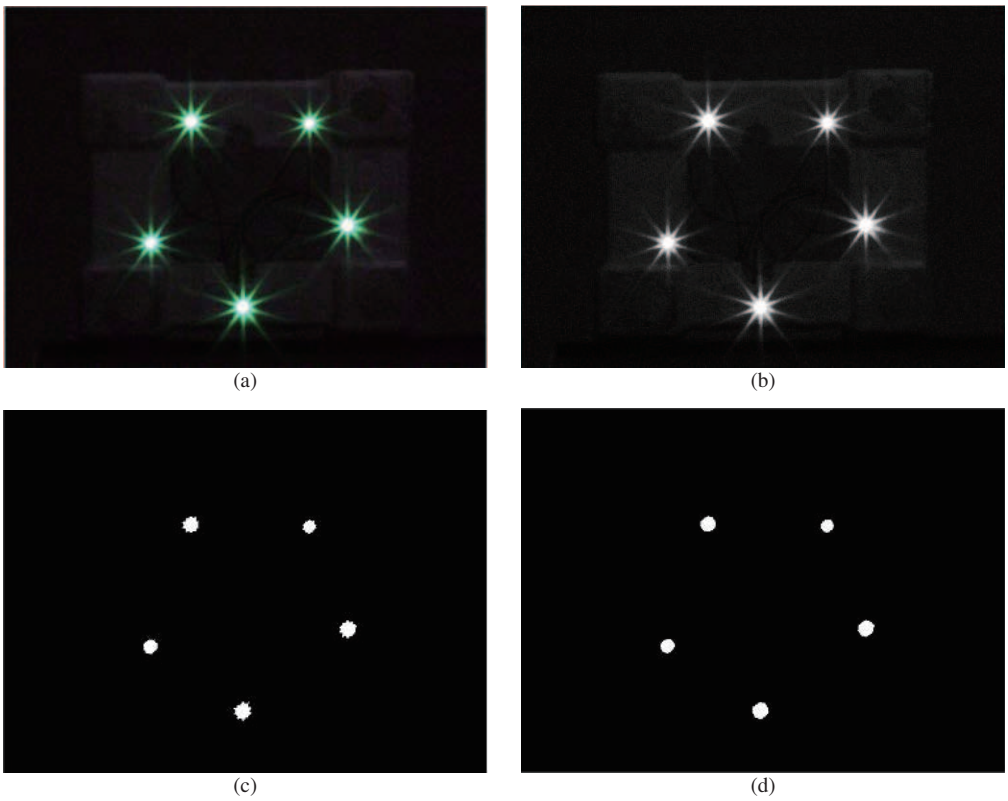


Figure 5 (Color online) Preprocessing of image. (a) Three-channel color image; (b) graying the color image; (c) binarization to the grayscale image; (d) median filtering of image.

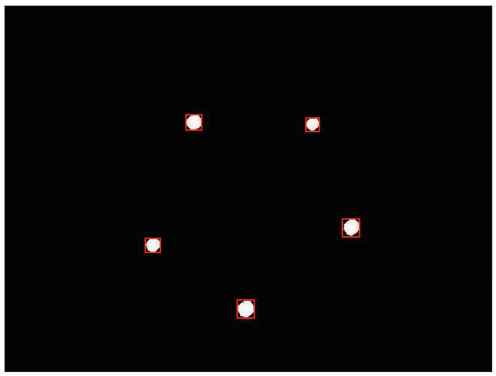


Figure 6 (Color online) Connected area labeling effect diagram.

where $(x_{\max}, y_{\max})$ represents the maximum values of the horizontal and vertical coordinates of the pixel point, P_i is the pixel value of each pixel point, (x_i, y_i) indicates the pixel coordinates of each pixel point, and (X_c, Y_c) indicates the pixel coordinates of the centroid of the connected area.

Algorithm 2 OSE algorithm

Input: Figure 5(d), initialization markup matrix.
Output: The marked image matrix and the number of connected areas.
1: **for** $i = 1, \ldots, y_{\max}$ **do**
2: **for** $j = 1, \ldots, x_{\max}$ **do**
3: Find the pixels that are not marked and mark them in the tag matrix.
4: **end for**
5: **end for**

The obtained FP extraction results are shown in Figures 7(a) and (b). Each FP (red dot) is in the

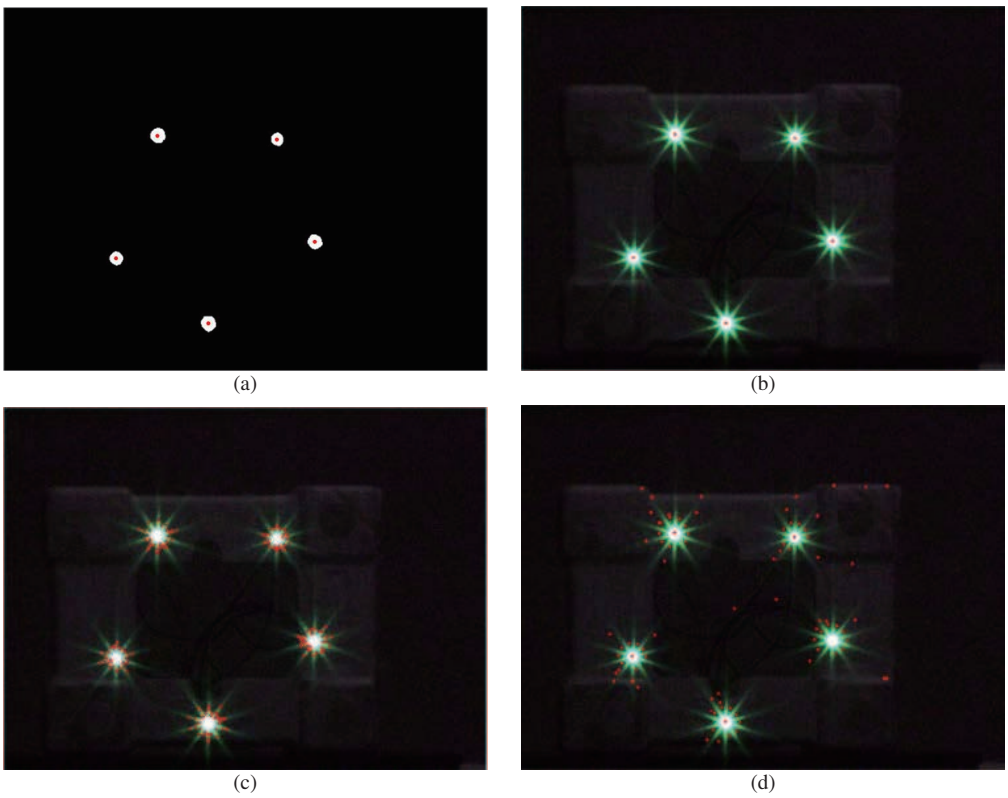


Figure 7 (Color online) Effects of different algorithms. (a) Effect diagram of centroid algorithm; (b) effect of the centroid algorithm in the original image; (c) effect image extracted using the Harris algorithm; (d) effect image extracted using the SURF algorithm.

Table 3 Experimental data of FP extraction

Feature extraction algorithm	Total number of FPs	Number of effective FPs	Effective ratio (%)
OSE	1000	1000	100
Harris	15283	14796	96.81
SURF	18165	16914	93.11

center of mass of the LED. It represents the position of the LED in the image and reflects the actual position of the RCS. Simulation extraction experiments using the Harris algorithm and the sped-up robust features (SURF) algorithm are also carried out in this study. The results are shown in Figures 7(c) and (d), respectively. It can be observed that the FPs extracted using the Harris algorithm are distributed on the light around the imaging area of the LEDs, where the light and shade intersect most evidently. They are not in the center of the imaging area of the LEDs. In comparison with the Harris algorithm, the number of FPs in the SURF algorithm is reduced. However, many of the extracted FPs are not around the LEDs; these are invalid FPs and cannot represent the RCS. Instead, they will increase the amount of calculation. Thus, the accuracy of FP matching is reduced. Therefore, these two algorithms are not suitable for the UAV AAR technology based on binocular vision. One-hundred FP extraction experiments were carried out on the three algorithms and the statistical results are shown in Table 3.

4.3 Obtaining the description vector of FPs

The traditional Haar wavelet transform requires a significant amount of calculation to obtain the pixels of the black and white rectangular area in the Haar wavelet template. Furthermore, the dimension of the description vector is small, and hence, it cannot accurately describe the FP. We make the following improvements.

(1) Calculation of the integral image. The integral image is used to simplify the process of obtaining

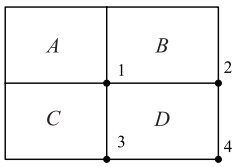


Figure 8 Using the integral image to determine the rectangular area in the image.

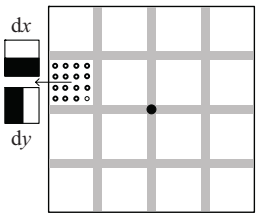


Figure 9 Representation of FP of the IHWT.

an image rectangle [21]. The value of a pixel point is the area of the rectangle formed by the origin of the pixel CS and the pixel point. The equation for integral image calculation can be written as follows:

$$I_{\sum(X)} = \sum_{i=0}^{i\leqslant x_{\max}} \sum_{j=0}^{j\leqslant y_{\max}} I(i,j), \tag{12}$$

where $I_{\sum(X)}$ is the integral image, (i,j) indicates the pixel coordinates of the pixel point, and $I\,(i,j)$ represents the pixel value of the pixel point. The integral image obtained is shown in Figure 8. The area of the rectangle D in the diagram can be expressed as

$$S_D = I_1 + I_4 - I_2 - I_3, \tag{13}$$

where S_D represents the area of D in the rectangular area and $I_k\,(k = 1, 2, 3, 4)$ represents the integral image value of the pixel point k. The value of the top left corner (point 1) of the area D is added to the value of the lower right corner (point 4), and subsequently, the values of the lower left corner (point 3) and the top right corner (point 2) are subtracted. The area of the rectangle D can be obtained. The calculation of this method is simple and the amount of calculation is much less than that of the traditional algorithm.

(2) Improved Haar wavelet transform (IHWT). When the traditional Haar wavelet transform is used to filter the FPs, only four-dimensional description vectors can be obtained. Therefore, the FPs cannot be described in detail and accurately, which will reduce the accuracy of stereo matching of the FPs. The steps of the improved method can be described as follows:

• First, a square area as the center of the FP in the image is considered and it is divided into 16 smaller areas of uniform size.

• The size of the small square area is considered the size of the Haar wavelet template. Some 16 small square areas are transformed using Haar wavelet in the horizontal direction (HD) and the vertical direction (VD). The response values in the HDs and VDs are obtained.

• Finally, the response value in HD, absolute value in HD, response value in VD, and absolute value in VD of the Haar wavelet [22] transform are used to form the 64D description vectors, as shown in Figure 9.

By improving the original method, a description vector of higher dimension is obtained. As the area of the image described becomes larger, the description of FPs is more detailed and accurate, which can effectively improve the accuracy of FP matching.

4.4 Minimum Euclidean distance to match FPs

The Euclidean distance between the description vector of each FP in the left image and the description vector of the FP in the right image is calculated. The two points with the minimum Euclidean distance are considered a pair of matching points. The Euclidean distance equation of the FP description vector is expressed as

$$d_{12} = \sqrt{\sum_{k=1}^{w}(r_{1k} - r_{2k})^2}, \tag{14}$$

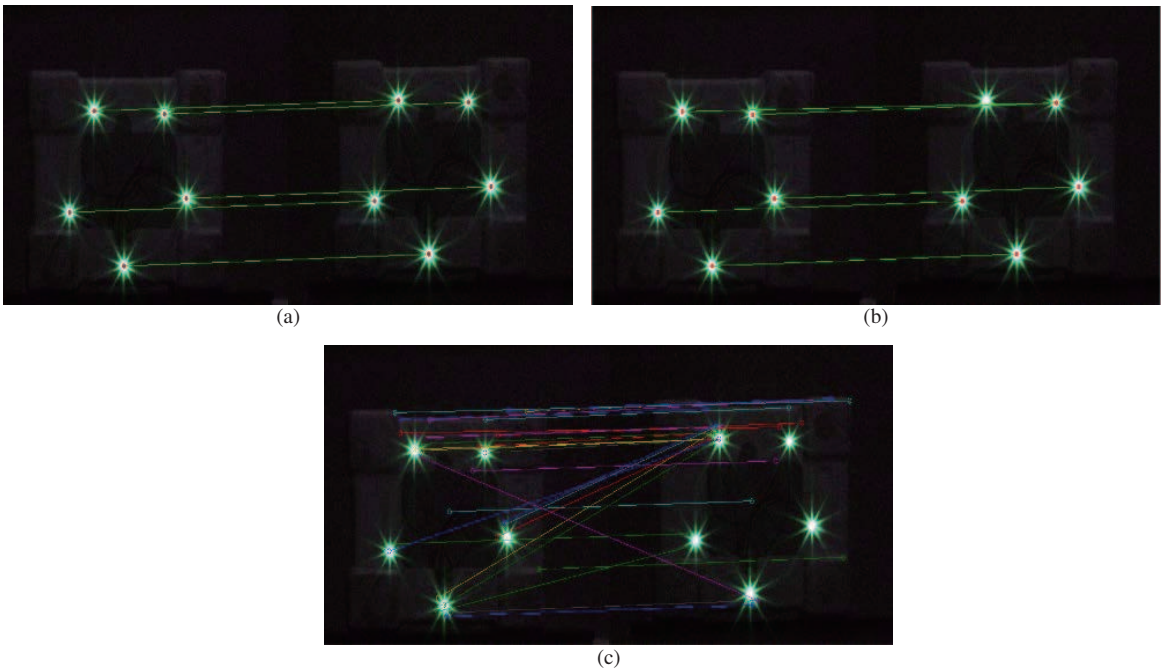


Figure 10 (Color online) Demonstrations of matching effect. (a) Effect of stereo matching of FPs; (b) effect of stereo matching using the ELR algorithm; (c) effect of stereo matching using the SURF algorithm.

Table 4 Experimental data of FP matching

Matching algorithm	Total number of pairs	Correct number of pairs	Matching accuracy (%)
IHWT	50	50	100
ELR	50	43	86
SURF	300	135	45

where d_{12} represents the Euclidean distance between two FPs, w represents the highest dimension of the FP description vector, r_{1k} and r_{2k} represent the values of the left and right image FPs in the k dimension, respectively.

Through descripting and matching the FPs, the results of stereo matching of FPs are shown in Figure 10(a). All the matching results of LEDs are correct. Furthermore, the epipolar line rectification (ELR) algorithm [23] and SURF algorithm are used to perform the simulation experiment and the results are shown in Figures 10(b) and (c), respectively. Although the matching speed of the ELR algorithm is faster, five pairs of FPs are matched and two wrong pairs of matching appear. Thus, this algorithm is not suitable for the high-precision matching requirement of AAR. The matching result of SURF algorithm has too many FPs. Many of them are not relevant to the RCS and are invalid feature points; hence, this algorithm is also not suitable for UAV AAR and close-range docking navigation technology. The statistical results of 100 experiments are shown in Table 4.

5 Position and pose estimation of the RCS

5.1 Calculation of the coordinates of FPs

Selecting one of the groups from the experimental image data and matching the FPs clockwise, the pixel coordinates of the feature matching pairs are shown in Table 5. According to the binocular vision

Table 5 Pixel coordinates of FPs

Labels of FP	Pixel coordinates of the left image	
	u-axis coordinates	v-axis coordinates
1	766.3222	294.0458
2	841.5005	296.4669
3	1065.564	452.7329
4	834.3635	596.7836
5	610.3694	437.7926
Labels of FP	Pixel coordinates of the right image	
	u-axis coordinates	v-axis coordinates
1	550.1918	322.0488
2	625.2786	324.3251
3	847.9412	481.3774
4	612.4110	625.5150
5	392.3321	465.6742

Table 6 Coordinates of the FPs

Labels\coordinates of world CS	Coordinates of the FPs obtained using BVS		
	X-axis (mm)	Y-axis (mm)	Z-axis (mm)
1	30.0643	55.2186	957.0100
2	60.1948	55.9709	956.9757
3	147.2115	−9.3191	943.2886
4	55.9019	−64.4386	939.0642
5	−32.2210	−1.2189	947.7082
Labels\coordinates of world CS	Coordinates of the FPs obtained via measurement		
	X-axis (mm)	Y-axis (mm)	Z-axis (mm)
1	29.4972	55.6821	955.7350
2	59.5163	54.7009	955.6184
3	147.5596	−7.6893	946.1418
4	55.4418	−64.1295	938.7077
5	−32.5715	−1.6992	946.9239

principle, the 3D coordinates of FPs in the world CS can be calculated as

$$\begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = \left(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}\right)^{-1}\boldsymbol{M}^{\mathrm{T}}\boldsymbol{U}, \tag{15}$$

where Eq. (15) satisfies the following relationship:

$$\boldsymbol{U} = \begin{bmatrix} u_l h_{l31} - h_{l11} & u_l h_{l32} - h_{l12} & u_l h_{l33} - h_{l13} \\ v_l h_{l31} - h_{l21} & v_l h_{l32} - h_{l22} & v_l h_{l33} - h_{l23} \\ u_r h_{r31} - h_{r11} & u_r h_{r32} - h_{r12} & u_r h_{r33} - h_{r13} \\ v_r h_{r31} - h_{r21} & v_r h_{r32} - h_{r22} & v_r h_{r33} - h_{r23} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = \boldsymbol{M} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix}. \tag{16}$$

Furthermore, the real 3D coordinates of five FPs in the world CS are obtained using other measuring tools. The results are shown in Table 6. Comparing the data, the errors between the actual measured coordinates and those calculated using the method proposed in this paper are in millimeter range. The average absolute error of the data is 0.87 mm. One-hundred experiments were performed and the absolute errors were obtained. The cylindrical distribution is shown in Figure 11. From the distribution law, the absolute errors are mostly within the range 0.85 to 0.95 mm.

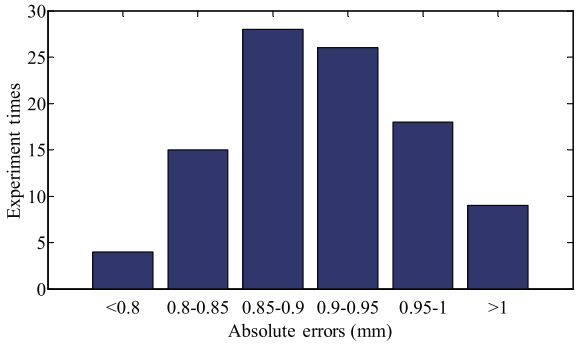


Figure 11 (Color online) Distribution of absolute errors for 100 experiments.

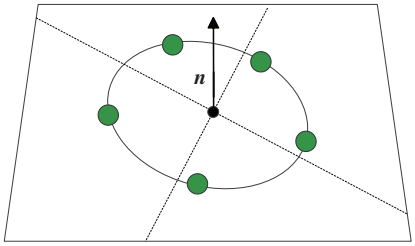


Figure 12 (Color online) Bottom end face of RCS and its normal vector diagram.

5.2 Calculation of the position and pose of the RCS

The coordinates of the center of the bottom of the RCS represent the actual position of the RCS, and they guide the probe of the UAV while inserting the RCS. The normal vector represents the pose of the RCS and indicates the direction in which the UAV inserts the RCS [24]. Therefore, the center of the circle and the normal vector of the bottom face of the RCS indicate the position and pose information of the RCS for UAV AAR, respectively. The principle is shown in Figure 12. According to the knowledge of analytic geometry, the equation of the sphere in 3D space and the plane equation can be established as follows: $(x_i - a)^2 + (y_i - b)^2 - (z_i - c)^2 = R^2$ and $Ax_i - By_i + Cz_i + D = 0$. The coefficients can be obtained by fitting the equations with the least-square method.

$$(x - 41.15)^2 + (y - 133.25)^2 - (z - 26.97)^2 = 933.39^2, \tag{17}$$

$$-0.05x - 0.19y + 1.06z = 1000. \tag{18}$$

According to the above steps, we obtain the relative position and pose information between the UAV and the RCS by combining (19) and (20). The coordinates of the circle of the RCS in the world CS are (52.2303, 7.2426, 948.8093), the radius is approximately 75 mm, and the normal vector is $(-0.082, -0.1468, 0.9891)$.

6 Docking control simulation

6.1 Docking and tracking control system

To verify the presented navigation method, an AAR control simulation is conducted based on the docking and tracking control (DTC) system model given in [25] and the structure diagram is shown in Figure 13. The system is mainly composed of six parts: UAV, RCS, BVC, referencing trajectory generation module (RTGM), observer, and controller. The following Algorithm 3 is the principle of DTC, which can be described in details.

Algorithm 3 DTC algorithm

Input: y, y_d, and y_r.
Output: Errors of relative position and pose of UAV and RCS.
1: **for** $i = 1, \ldots, T$ (time discretization) **do**
2: Find y_r, y^*, $\hat{x}$ and $\hat{u}$.
3: **end for**
4: **Note:** T is the time for close-range docking and refueling.

• First, the relative position and pose information between the UAV and the RCS is measured using BVS. The acquired position and pose information is filtered and used as the input signal of the system. Further, the desirable flight trajectory of the UAV is calculated by the RTGM in real time.

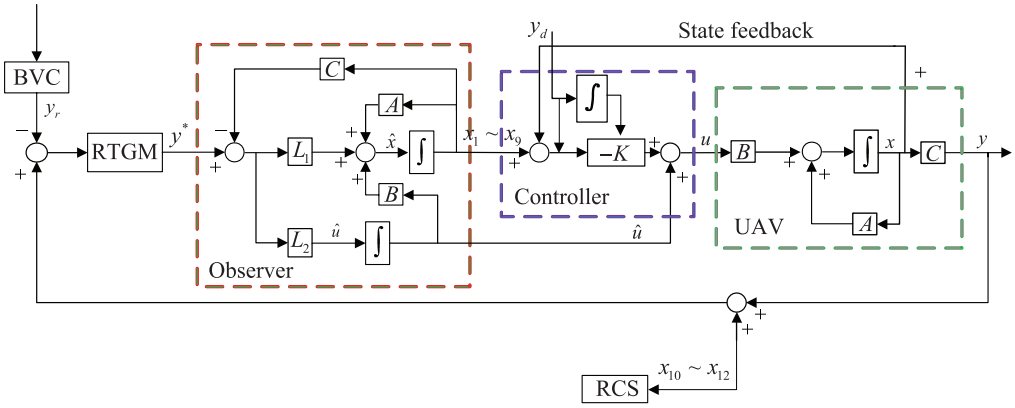


Figure 13 (Color online) Structure diagram of tracking control system for AAR and docking of UAV.

Table 7 Related initial parameters of the UAV and RCS

Parameter variables	Parameter values
Position of UAV & RCS (m)	(0, 0, 0) & (52.2303, 7.2426, 948.8093)
Pose of UAV & RCS (°)	(0, 0, 0) & (−0.083, −0.1468, 0.9891)
Speed of UAV & RCS (m/s)	20
Acceleration of UAV & RCS (m/s^2)	20

• Subsequently, the expected state and control variables of UAV are estimated by the system state observer and the estimated values are input into the controller.

• Finally, a state feedback controller is designed to control the UAV to track the desired trajectory, so that the UAV can dock with the RCS accurately and continue tracking during the refueling process.

In the structure diagram above, $\boldsymbol{A}$ is the system matrix, $\boldsymbol{B}$ is the input matrix, $\boldsymbol{C}$ is the output matrix, $\boldsymbol{G}$ is the feedback matrix, u represents the actual control of UAV, $\hat{u}$ represents the control estimated by the state observer, x and $x_{1,\ldots,12}$ indicate the state of the UAV, $\hat{x}$ represents the state estimation of the state observer, y is the initial inertial coordinate position of the UAV, y_d is the initial inertial coordinate position of the RCS, y_r is the position of the RCS relative to the UAV, y^* is the reference flight path of the UAV, $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ denote the gain of the state observer, and $\boldsymbol{K}_i$ and $\boldsymbol{K}_p$ denote the gain of the controller. The observer gain $\boldsymbol{L}$ is designed as

$$\boldsymbol{L} = \boldsymbol{P}_0 \boldsymbol{C}'^{\mathrm{T}} \boldsymbol{R}_0^{-1}, \tag{19}$$

where $\boldsymbol{C}' = [\boldsymbol{C}\ \boldsymbol{0}]$ is a solution of Riccati equation and $\boldsymbol{R}_0$ is a positive definite matrix. The control law is given by

$$u = -\boldsymbol{R}^{-1} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{P}^* x, \tag{20}$$

where the controller gain matrix is $\boldsymbol{K} \triangleq -\boldsymbol{R}^{-1} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{P}^*$. It should be emphasized that this study mainly investigates the relative position and pose estimation method of the UAV and RCS. More detailed scheme designs, parameters setting, and system stability proof of the system observer and controller can be found in [25].

6.2 Simulation results

The relative position and pose between the UAV and RCS are used as the reference trajectories and angles. The simulation results of the states and control variables of the UAV at the refueling docking stage under the parameters of Table 7 are shown in Figures 14–16.

In Figure 16, $\Delta\delta_t, \Delta\delta_a, \Delta\delta_e$, and $\Delta\delta_T$ are the position of throttle rod, angle of aileron, angle of elevator, and angle of rudder, respectively, which are relative to the equilibrium state. The simulation results show that, after obtaining the relative position and pose between the UAV and RCS, the UAV will be guided and controlled while moving to the RCS. The process can be divided into two stages.

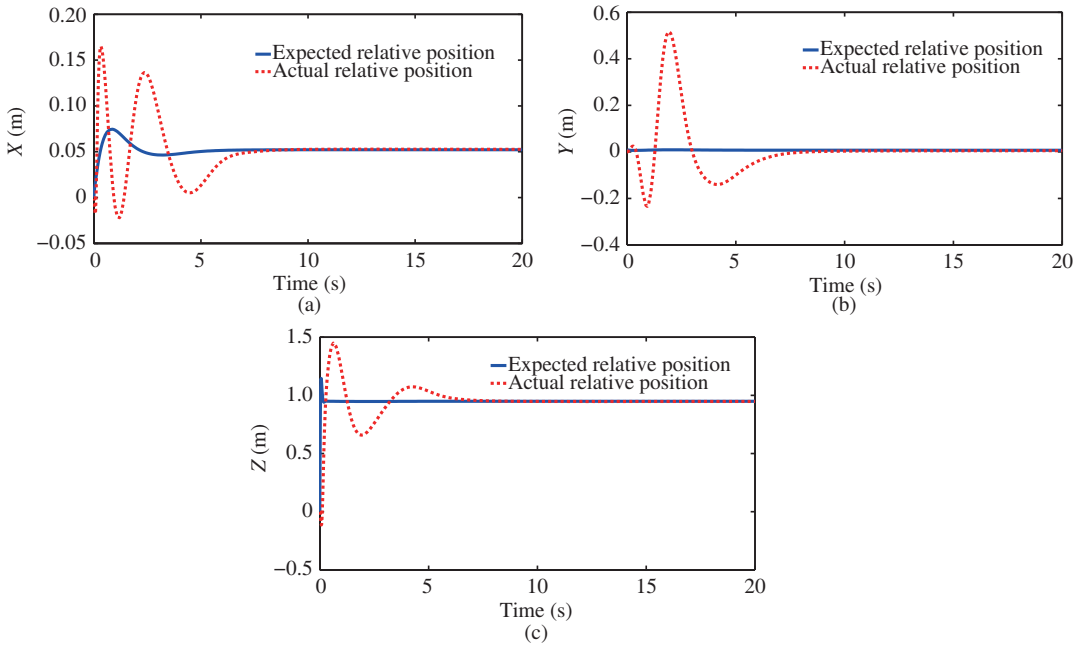


Figure 14 (Color online) Actual relative position and expected relative position of UAV. (a) In the X-axis; (b) in the Y-axis; (c) in the Z-axis.

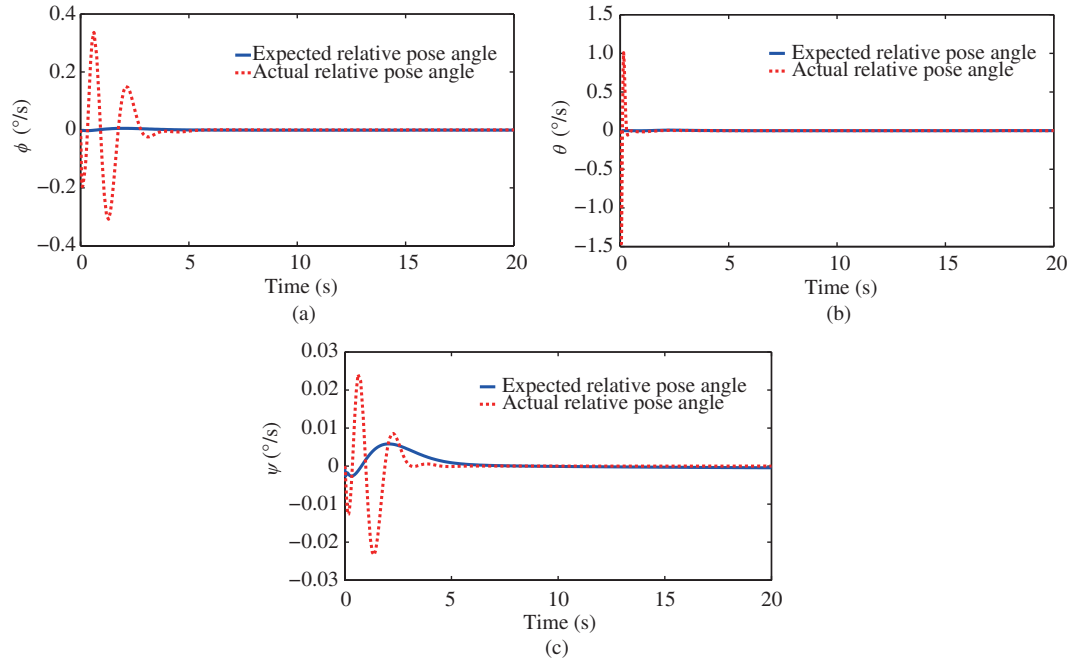


Figure 15 (Color online) Actual relative pose and expected relative pose of UAV. (a) ϕ angle; (b) θ angle; (c) ψ angle.

(1) 0–8 s is the stage of close-range docking. It can be observed that the errors between the measured states and the reference signals gradually attenuate to zero without much oscillation. The overshoot of control also decreases gradually.

(2) 8–20 s is the stage of refueling. During this period, the UAV has achieved precise docking with RCS and maintains this state during the refueling process. All the state variables and control variables reach the steady state.

Overall, it can be observed that the estimated relative positions and pose angles of the RCS to the UAV are accurate and the refueling docking control process is effective. This demonstrates the effectiveness of

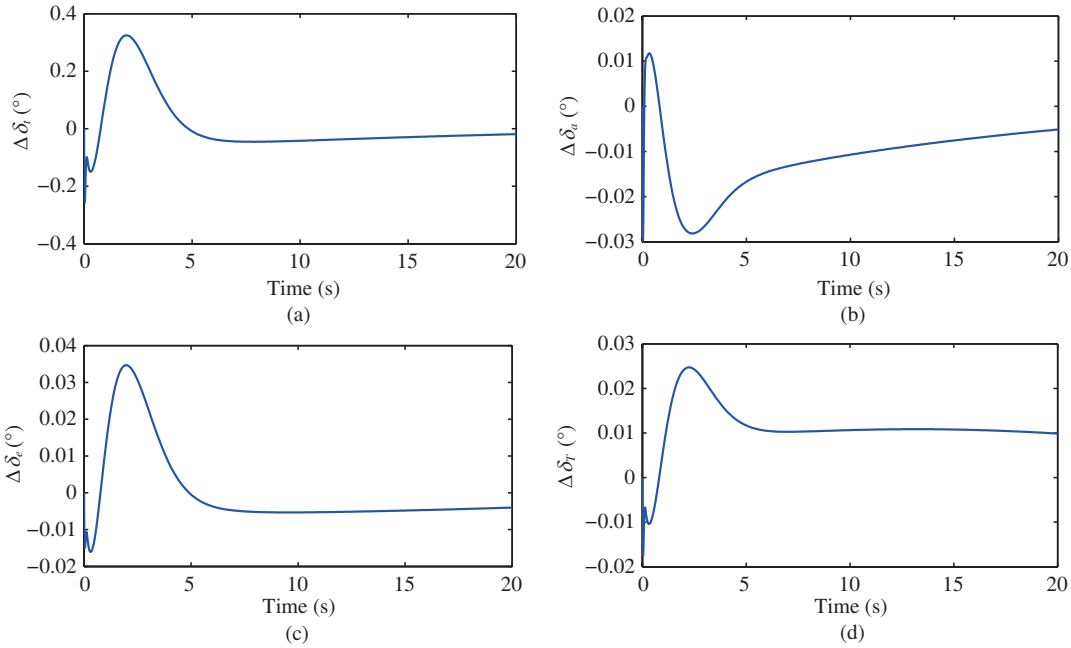


Figure 16 (Color online) Control values for controller output. (a) $\Delta\delta_t$; (b) $\Delta\delta_a$; (c) $\Delta\delta_e$; (d) $\Delta\delta_T$.

the docking navigation method proposed in this paper.

7 Conclusion

As a critical technology in the field of UAV application, AAR can effectively not only increase the range and payload but also improve the combat capability of UAVs. The simulation results show that additional scene information can be weakened using LEDs and filters. The graying, binarization, and median filtering of the image can further highlight the imaging area and reduce the noise. The improved Haar wavelet transform is more accurate than the other algorithms in matching the FPs. Using the least-square method to fit the coordinates of the matching FPs, it is observed that the errors between the obtained results and the actual measured values are in millimeter range. The BVS simulation platform can simulate the unmanned aerial refueling scene. As the subsequent step, we will verify the accuracy of the experimental platform simulation under aerodynamic disturbances and provide the corresponding correction scheme.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 61673327) and Aeronautical Science Foundation of China (Grant No. 20160168001).

References

1 Zhang X, Duan H B, Yu Y. Receding horizon control for multi-UAVs close formation control based on differential evolution. Sci China Inf Sci, 2010, 53: 223–235
2 Xu Y, Lai S P, Li J X, et al. Concurrent optimal trajectory planning for indoor quadrotor formation switching. J Intell Robot Syst, 2018, 4: 1–18
3 Nalepka J, Hinchman J. Automated aerial refueling: extending the effectiveness of UAVs. In: Proceedings of AIAA Modeling and Simulation Technologies Conference and Exhibit, San Francisco, 2013. 15–18
4 Li H, Duan H B. Verification of monocular and binocular pose estimation algorithms in vision-based UAVs autonomous aerial refueling system. Sci China Technol Sci, 2016, 59: 1730–1738
5 Xu Y, Li D Y, Luo D L, et al. Affine formation maneuver tracking control of multiple second-order agents with time-varying delays. Sci China Tech Sci, 2018. doi: 10.1007/s11431-018-9328-2
6 Bloy A W, Trochalidis V, West M G. The aerodynamic interference between a flapped tanker aircraft and a receiver aircraft during air-to-air refueling. Aeronaut J, 2016, 95: 274–282
7 Valasek J, Gunnam K, Kimmett J, et al. Vision-based sensor and navigation system for autonomous air refueling. J Guidance Control Dyn, 2005, 28: 979–989

8 Tandale M D, Bowers R, Valasek J. Robust trajectory tracking controller for vision based probe and drogue autonomous aerial refueling. In: Proceedings of AIAA Guidance, Navigation, and Control Conference and Exhibit, San Francisco, 2005. 15–18

9 Pollini L, Campa G, Giulietti F, et al. Virtual simulation setup for UAVs aerial refueling. In: Proceedings of AIAA Conference on Modeling and Simulation Technologies and Exhibits, Austin, 2003. 11–14

10 Pollini L, Innocenti M, Mati R. Vision algorithms for formation flight and aerial refueling with optimal marker labeling. In: Proceedings of AIAA Modeling and Simulation Technologies Conference and Exhibit, San Francisco, 2005. 15–18

11 Schweighofer G, Pinz A. Robust pose estimation from a planar target. IEEE Trans Pattern Anal Mach Intell, 2006, 28: 2024–2030

12 Zhang S J, Cao X B, Zhang F, et al. Monocular vision-based iterative pose estimation algorithm from corresponding feature points. Sci China Inf Sci, 2010, 53: 1682–1696

13 Chen C I, Stettner R. Drogue tracking using 3D flash lidar for autonomous aerial refueling. Proc SPIE, 2011, 8037: 2362–2375

14 Huang B, Sun Y R, Sun X D, et al. Circular drogue pose estimation for vision-based navigation in autonomous aerial refueling. In: Proceedings of Guidance, Navigation and Control Conference (GNCC), Nanjing, 2016. 960–965

15 Duan H B, Zhang Q F, Deng Y M, et al. Biologically eagle-eye-based autonomous aerial refueling for unmanned aerial vehicles. Chin J Sci Instrum, 2014, 35: 1450–1458

16 Duan H B, Li H, Luo Q N, et al. A binocular vision-based UAVs autonomous aerial refueling platform. Sci China Inf Sci, 2016, 59: 053201

17 Bolien M, Iravani P, Bois J L D. Toward robotic pseudodynamic testing for hybrid simulations of air-to-air refueling. IEEE/ASME Trans Mechatron, 2017, 22: 1004–1013

18 Zhang Z. Flexible camera calibration by viewing a plane from unknown orientations. In: Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV), Kerkyra, 2002. 666–673

19 Zhang Z. A flexible new technique for camera calibration. IEEE Trans Pattern Anal Machine Intell, 2000, 22: 1330–1334

20 Yang J, Liang B, Zhang T, et al. A novel systematic error compensation algorithm based on least squares support vector regression for star sensor image centroid estimation. Sensors, 2011, 11: 7341–7363

21 Molano R, Rodríguez P G, Caro A, et al. Finding the largest area rectangle of arbitrary orientation in a closed contour. Appl Math Comput, 2012, 218: 9866–9874

22 Alemohammad M, Stroud J R, Bosworth B T, et al. High-speed all-optical Haar wavelet transform for real-time image compression. Opt Express, 2017, 25: 9802–9811

23 Papadimitriou D V, Dennis T J. Epipolar line estimation and rectification for stereo image pairs. IEEE Trans Image Process, 1996, 5: 672–676

24 Lu C P, Hager G D, Mjolsness E. Fast and globally convergent pose estimation from video images. IEEE Trans Pattern Anal Mach Intell, 2000, 22: 610–622

25 Murillo O, Lu P. Comparison of autonomous aerial refueling controllers using reduced order models. In: Proceedings of the AIAA Guidance, Navigation and Control Conference and Exhibit, Honolulu, 2008. 18–36