IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 18, 2021, accepted April 12, 2021, date of publication April 20, 2021, date of current version April 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3074413

Vision-Based Target Detection and Tracking System for a Quadcopter

SHICHENG WU, RUI LI, (Senior Member, IEEE), YINGJING SHI, (Member, IEEE), AND QISHENG LIU

School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Corresponding author: Rui Li (hitlirui@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61973055, and in part by the Fundamental Research Funds for the Central Universities under Grant ZYGX2019J062.

ABSTRACT Tracking a maneuvering target with a quadcopter is a challenging problem, that involves a variety of fields such as visual tracking, state estimation, and control algorithms. Most existing unmanned aerial vehicle (UAV) systems fail to track targets accurately in the long term and cannot relocate targets after target loss. This paper aims to design and implement a vision-based target tracking system for quadcopters that can steadily and accurately track the ground target as well as the air target without any prior information. We employ a vision detection algorithm to select the target quickly and precisely. To fit complex practical conditions, a target tracking algorithm is developed based on correlation filters, which is capable of tracking targets with large-scale variation and fast motion. In addition, an efficient redetection algorithm based on the support vector machine (SVM) is designed to handle target occlusions and loss. The target states are estimated from the visual information by an improved Lucas-Kanade (LK) optical flow method and an extended Kalman filter (EKF). Moreover, a double closed-loop Proportion Integral Differential (PID) controller using the estimated states is designed to follow the target. By implementing the main algorithms on an onboard NUC computer, an extensive outdoor flight is evaluated for a quadcopter platform equipped with a stereo camera. The experimental results validate the feasibility and practicability of the developed system.

INDEX TERMS Target tracking, unmanned aerial vehicles, state estimation, stereo vision.

I. INTRODUCTION

In the past two decades, multirotor unmanned aerial vehicles, especially quadcopters, have attracted increasing attention due to their vertical take-off and landing abilities, hoving capabilities and exceptional agility. With the development of UAVs, a diverse range of tasks have been appeared, such as search and rescue [1]–[3], surveying [4], surveillance [5], traffic monitoring [6], exploration and mapping [7], [8], delivering goods [9] and autonomous navigation [10]. In particular, visual tracking has become a key task for quadcopters in vision-based applications.

Visual tracking is an important branch of computer vision that aims to automatically determine an object's bounding box in the subsequent frame [11]. Although some works have been made significant progress in tracking, there are still many challenges for practical applications, especially for UAVs. On the one hand, a tracking system frequently tracks targets in the long term, where targets may undergo fast motions, scale variations, heavy occlusion and even moving out-of-view. On the other hand, a real-time tracking algorithm should be considered for airborne computer operations.

Currently, deep learning methods have achieved good tracking performance, but considering their real-time capabilities, they are rarely used in actual engineering applications. Due to their impressive robustness and real-time capabilities, discriminative correlation filter (DCF-)based trackers [13], [14] have gained sustained attention. To address target large-scale variations, the discriminative scale space tracker (DSST) [12] learns separate translation and scale models. A fast version of the DSST (FDSST) has been proposed, which reduces the computational complexity to simultaneously increase a larger target search space and maintain real-time performance [35]. The main methods of tolerating

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

fast motion utilize larger regions of interest (ROIs) and suppressing spatial attention maps [15], [16]. However, these methods may increase the drifting risk of the correlation filter, which may lead to tracking failure after the accumulation of several frames. For occlusion and out-of-view problems in long-term tracking, LCT [17] and PTAV [18] utilize the random forest and the Siamese network [19] as their redetection schemes respectively, where all intermediate tracking results are considered to identify the tracking state of the target. Recently, Wang *et al.* [20] made full use of the baseline tracker for redetection and did not need to incorporate extra features into trackers for redetection purposes.

In addition to researches of tracking algorithms, many works have focused on the application of visual tracking algorithms. In earlier researches, [21] performed road traffic surveillance and autonomous tracking of a moving target by a fixed-wing UAV, where the tracking algorithm was formulated as a normalized cuts problem. Reference [22] developed a system of estimating target GPS coordinates, and the impact of target loss events on the control and estimation algorithms was analyzed. Relying on the color target detection and tracking algorithm, [23] designed a tracking system that enabled the quadcopter to track and hover above the target. To address continuous unrestricted pose variation and low-frequency vibrations of vehicles, [24] developed an approach to create an artificial optical flow by estimating the camera motion. In [25], a vision system for quadcopters was designed to track a moving target on ground, where switching controllers were proposed for stabilizing quadcopters during flight.

In recent years, new functions and improvements have been made for target tracking systems. [26] focused on the trajectory planning and tracking problem, where the relative position of the target in 3D space is obtained by using artificial markers. In [27], Justin *et al.* addressed an autonomous flight system of a small quadcopter that enabled to track a moving object, where the dynamics of the underactuated robot, the actuator limitations, and the field of view constraints were considered. Reference [28], [29], [30] achieved object tracking for quadcopters by combining an object detection algorithm with an improved PID algorithm. A novel technique based on artificial neural networks was proposed in [31] for the identification and tracking of targets. In [32], a detection and multi-object tracking algorithm was proposed based on color and depth data for rescuing victims in disaster environments. Simultaneous target tracking and path planning by a quadcopter were accomplished with a novel system framework proposed in [38], where the target size was set as prior information for target state estimation.

However, limitations still exist in the research on quadcopter visual tracking systems. First, the tracking accuracy is insufficient. On the one hand, limited by the computational capacities of airborne computers, most tracking algorithms with high performance are not suitable for real-time UAV tracking due to complicated computations. On the other hand, the translation drift will accumulate during long-term tracking [15], [16], and the aforementioned DCF-based tracking methods [17], [18], [20] search targets only in a local region or perform redetection at regular intervals, so these tracking methods do not perform well in fast motion or out-of-view scenarios. Second, the application scenarios of these tracking systems are limited. For example, some tracking systems are only able to track the ground target [28], [29], [33] or the target with other prior information [26], [27]. This occurs because of the lack of accurate state estimation on the target. Therefore, some preconditions, such as the height information of the quadcopter and artificial markers, are used for auxiliary estimation on target states, which will affect the versatility and practicability of target tracking systems.

We have concluded that a practicable and stable target tracking system for UAVs should have the following characteristics in common:

1) The overall algorithm of the UAV tracking system should be computationally efficient and demonstrate good real-time performance because the algorithm should run on an airborne computer and deal with many circumstances in real tracking tasks.
2) The target tracking algorithm should provide tracking failure detection, drift correction and target relocation to ensure safe and long-term tracking.
3) It is necessary to obtain as many target states as possible, including target global position and velocity, without any prior information of targets so that sensitive and robust tracking applications can be achieved in more scenes.
4) The UAV system should have the ability to perceive the environment for autonomous obstacle avoidance during the tracking process, and the relationship between tracking and obstacle avoidance can be coordinated.

In our work, by focusing on requirements 1) 2) 3), we provide a complete solution for UAV target tracking systems from hardware construction to software design. Considering the real-time performance of a UAV system, we design and develop our tracking algorithm based on a correlation filter. By improving the classic FDSST [35] and LCT2 [36] algorithms, drift correction and a novel target relocation method are designed for stable tracking performance in practical applications. Moreover, a state estimation method is proposed that can estimate the target position and velocity in global coordinates with no need for target prior information and UAV height information, so that the developed UAV system has the ability to track arbitrary targets. The contributions of this paper are summarized as follows:

1) A systematic solution including software, hardware and algorithms is proposed to realize a UAV tracking system that has low power consumption and high real-time performance.
2) To suppress tracking drift even for fast moving targets, we combine approximate estimation with accurate correction based on an appearance filter. Meanwhile,

a novel full-frame search method is proposed based on a multiscale sliding window, which can relocate the target when it reappears at any area in the whole frame.

3) Target states that do not target position in global coordinates but target velocity are estimated by the proposed method, which is based on an improved LK optical flow and an extended Kalman filter, so that the tracking system is able to track the ground target as well as air target rapidly and robustly.

The rest of this paper is organized as follows. In Section II, we introduce the whole framework and the work process of the vision-based target tracking system. In Section III, the target tracking algorithm is presented. Section IV addresses the target state estimation algorithm. The UAV tracking controller is introduced in Section V. In Section VI, experiments are carried out to validate the developed tracking system. Concluding remarks and future work are discussed in Section VII.

II. SYSTEM OVERVIEW

The vision-based target tracking system consists of a UAV platform and ground station platform. The visual tracking algorithm, target state estimation algorithm and UAV flight controller run on a quadcopter platform in real time. The ground station platform is the interface for monitoring system states and sending quadcopter instruction. The whole framework of this system is shown in Fig. 1.

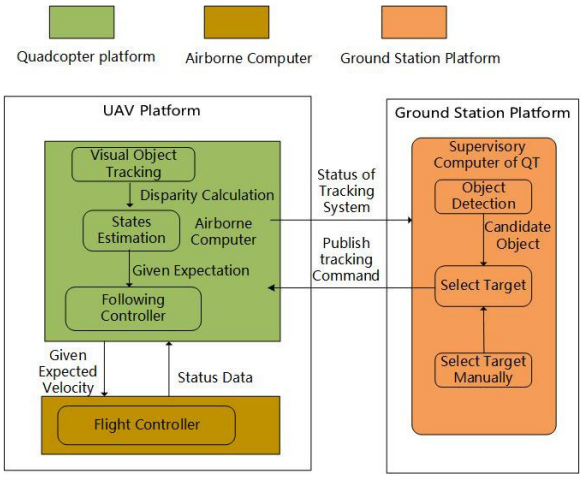


FIGURE 1. The whole framework of target tracking system.

The work process of the target tracking system is shown in Fig. 2. The camera provides visual information for the onboard computer and ground station platform. First, we select the target obtained by the object detection algorithm on the ground station. Then the target position in the subsequent video stream is obtained by the target tracking algorithm. In addition, the local or global position and velocity of the target are estimated by the state estimation algorithm. Furthermore, the controller is designed to correspondingly control the velocity and attitude of the UAV.

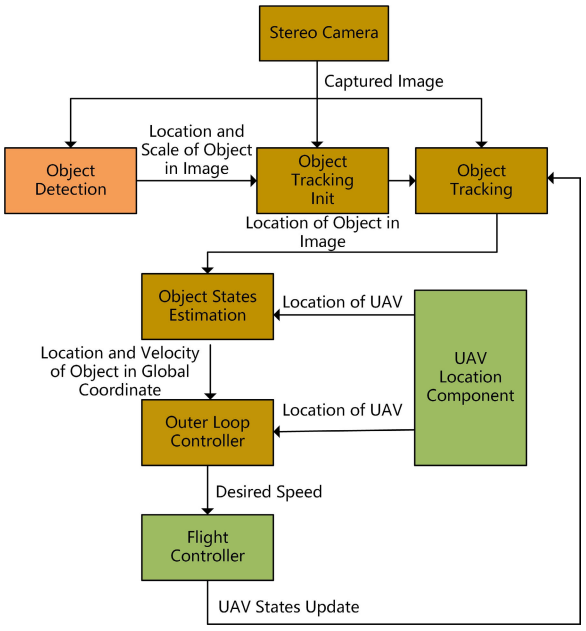


FIGURE 2. The work process of target tracking system.

III. VISUAL TRACKING ALGORITHM

To track the target by UAV, we first need to determine the target position with a 2D bounding box in the image plane by a visual tracking algorithm, which provides an object region for state estimation. Our visual tracking algorithm consists of translation and scale tracking, tracking quality evaluation and drift correction, tracking loss detection and target relocation. Fig. 3 shows the whole framework of our tracking algorithm.

As shown in Fig. 3, the target position p_t is inferred from the correlation response map of the translation filter. Based on p_t, the target scale s_t is predicted by a scale filter. Then we correct the drift of the target position with an appearance filter. Furthermore, the tracking quality is evaluated by the confidence score which is composed of average peak-to-correlation energy (APCE) [37] and the maximum response of the appearance filter in this paper. If the confidence score is higher than the redetection threshold T_r, which means that the target is tracking successfully, we update the translation filter and the scale filter. Otherwise, the SVM classifier is activated for target redetection. Note that the SVM classifier and appearance filter are updated when the confidence score is in the range of the interval threshold T_u.

A. TARGET POSITION AND SCALE TRACKING

The translation filter and the scale filter are applied to estimate the target position and scale in the image plane, respectively, both of which are borrowed from the FDSST algorithm.

The translation filter and scale filter have similar forms as eq. (1).

$$H^l = \frac{\overline{G}F^l}{\sum_{k=1}^{d} \overline{F^k}F^k + \lambda} = \frac{A_t^l}{B_t^l}. \quad l = 1, \ldots, d \tag{1}$$

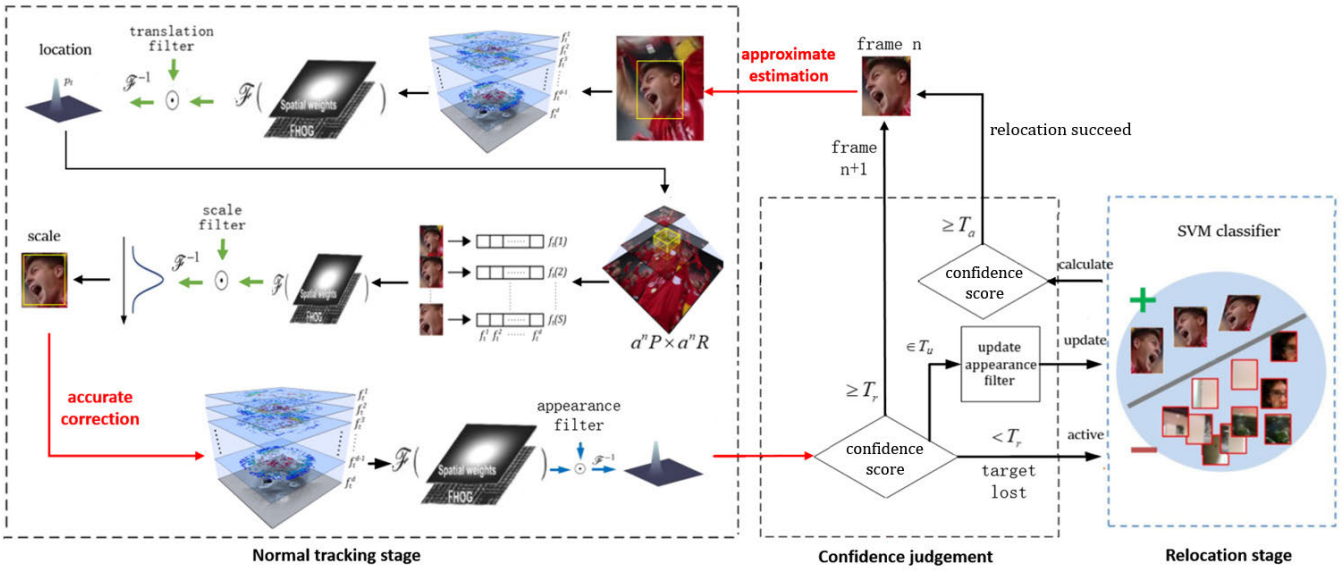


FIGURE 3. The whole tracking process.

Here G is the discrete Fourier transform (DFT) of the desired correlation output, and the bar denotes complex conjugation. λ is a regularization parameter. F^l denotes the DFT of the target training sample of feature channel l, where samples of translation filter f_{trans} consist of 31-dimensional FHOG features and 1-dimensional gray features, and samples of scale filter f_{scale} are composed of 33 feature vectors mapped from image patches of different scales.

After the target is selected, f_{trans} are extracted from the image patch of the target, and the translation filter is obtained by eq. (1). Then f_{scale} is extracted by using variable patch sizes centered around the target, and the scale filter is also calculated by eq. (1).

In the new frame t, the position and scale of the target are orderly determined by the highest response score of the translation filter and the scale filter. The response of the translation filter and that of the scale filter are both calculated by eq. (2).

$$Y_t = \frac{\sum_{l=1}^{d} \overline{A_{t-1}^l} Z_t^l}{B_{t-1} + \lambda}. \tag{2}$$

Here Z_t^l denotes the test sample of feature channel l. For the translation filter, the test samples are extracted at the previous location of the target with padding which has the same form as f_{trans}. The test samples of the scale filter are extracted at the target new center location in the same way as f_{scale}.

The translation filter and scale filter are updated as eq. (3)

$$\begin{aligned} A_t^l &= (1-\eta)A_{t-1}^l + \eta \overline{G} F_t^l, \\ B_t^l &= (1-\eta)B_{t-1}^l + \eta \sum_{k=1}^{d} \overline{F_t^k} F_t^k. \end{aligned} \tag{3}$$

Here η denotes the learning rate. In this paper, η changes with the tracking quality to minimize the effect of false tracking.

In general, the tracking algorithm based on FDSST achieves high accuracy and computing speed. However, this tracking algorithm is unable to evaluate the tracking quality and fails to detect target loss, thus it cannot be applied to UAV systems directly.

B. TRACKING QUALITY EVALUATION AND DRIFT CORRECTION

In LCT2, the appearance filter is presented to evaluate the tracking quality, which is employed to detect if the target is lost and determine if the filter needs to update. By referring to this filter, we design an appearance filter for tracking quality evaluation that has the same structure and updating form as translation filter in Section III-A. However, to avoid excessive update and maintain target appearance memory, the training samples of the appearance filter are extracted with no padding and the appearance filter is updated when the confidence score is within the interval threshold T_u.

Moreover, the predicted position of the target always has a small drift, which will accumulate in the long term and affect the tracking quality. The appearance filter is suitable for correcting this translation drift. Therefore, we design an additional step based on appearance filter for translation drift correction. The specific strategy is as follows: the translation filter still maintains a large padding and high learning rate for tracking fast-moving targets. Meanwhile, we employ the correlation response of the appearance filter calculated by eq. (2) for target translation correction, where the test samples are extracted around the region determined by the translation filter and scale filter with no padding. Note that the position with the highest response score of the appearance filter is the

accurate location. With this strategy, the translation filter is able to employ large padding for tracking fast moving targets, since the drift is then corrected by an appearance filter.

C. TARGET LOSS DETECTION AND REDETECTION

Oftentimes the target is in the presence of full occlusions in UAV tracking missions, which can lead to tracking an incorrect target and even result in crashes. In this paper, a novel target redetection strategy is proposed. When the confidence score is lower than T_r, meaning the target is lost, then the target redetection algorithm based on a SVM is activated. The redetection process is shown in Fig. 4.

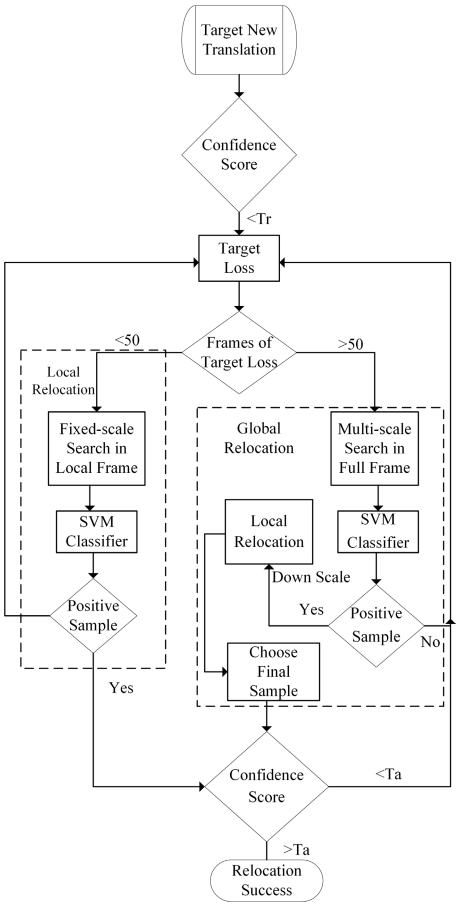


FIGURE 4. Target redetection.

Unlike the LCT algorithm which only relocates targets locally or TLD which sets a small scale factor for relocating targets globally, the relocation process is divided into two different strategies called local relocation and global relocation, which are summarized as follows.

1) Within 50 frames after target loss, the local relocation strategy is applied, where relocation samples are extracted by a sliding window around the target loss position with a fixed scale, and the relocation result is determined by the positive sample of the highest response score with the SVM.
2) After the target is lost exceeding 50 frames, the global relocation strategy is activated. Relocation samples are extracted by sliding windows with 9 different scales, which are much less than TLD. If positive samples exist, we select the positive sample with the highest response score with the SVM. Then, the fixed-scale search is restarted with the 0.8 scale of the positive sample. The final sample is selected with one of the highest response scores with the SVM.

When the positive sample is obtained, we compare the confidence score with T_a to determine whether the relocation is successful. Moreover, a scale filter is applied again to determine the target scale after successful relocation. These strategies reduce the computational complexity and guarantee the accuracy of target redetection, which is suitable for application on UAV systems.

IV. TARGET STATES ESTIMATION

The state estimation of the target and UAV is an important part of the tracking system, which provides inputs for the UAV controller. UAV states are obtained directly by PIXHACK, and the states of the target are estimated based on LK optical flow and the extended Kalman filter algorithm, the whole framework of which is shown in Fig. 5.

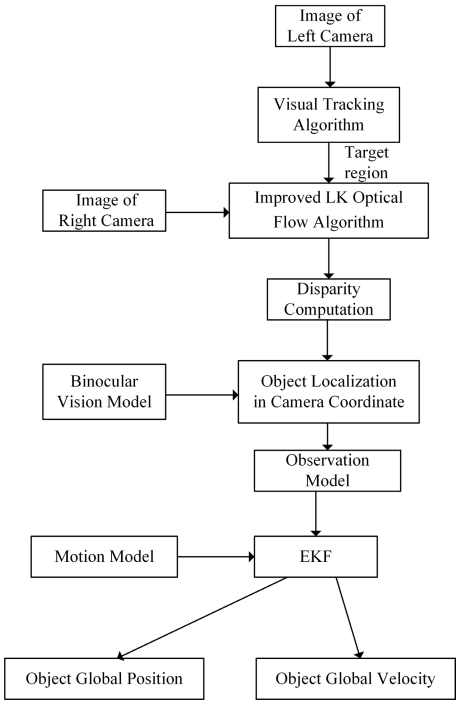


FIGURE 5. Target states estimation.

To estimate the target states, we first calculate the disparity of the image patch between the left and right images, which is obtained by the target tracking algorithm. Then the relative position between the UAV and target is obtained by the stereo vision model and coordinate system transformation. Since the calculation is generally inaccurate due to observation

noise, the extended Kalman filter is implemented for the accurate estimation of the target global position and speed by establishing a motion model and measurement model of the target.

A. THE DISPARITY CALCULATION

The disparity of the target is calculated by the method developed from the inverse-additive optical flow method proposed in [39]. Objective function $\|f(m)\|^2$ is constructed as:

$$\min_{\Delta m} \sum_{p_i \in W} \|\boldsymbol{I}_L(p_i + \Delta m) - \boldsymbol{I}_R(p_i + m)\|_2^2. \tag{4}$$

Here $\boldsymbol{I}_L$ and $\boldsymbol{I}_R$ denote the gray values in the left image and right image, respectively. W denotes the neighborhood of point p in the image plane.

The Gauss-Newton iterative method is applied to solve the least square equation, where the Jacobian matrix J is the gradient of the left image gray value $\boldsymbol{I}_L(p)$. Since stereo images have been aligned of rows by [40], we ideally consider the value of J in the y direction to be zero, which reduces errors and improves the computational efficiency. The Jacobian matrix J can be approximately written as:

$$\boldsymbol{J} = \begin{bmatrix} \dfrac{\partial I_L}{\partial x}(p) \\ 0 \end{bmatrix}. \tag{5}$$

The disparity of a single point is calculated by using the following steps.

1) Set the initial estimated value of disparity as m_0;
2) Calculate $f(m)$ and Jacobian matrix $\boldsymbol{J}$;
3) Solve the equation $\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J}\Delta m = -f(m)\boldsymbol{J}$;
4) Update m, that is $m + \Delta m \rightarrow m$, if the increment of disparity Δm is small enough, stop iterating and the current estimated disparity value m is the final value. Otherwise, go to step 2 and continue iteration.

To obtain the disparity of the tracking target between the right image and the left image, 300 points are selected randomly in the region of the target rectangular box in the left image, and the disparity of each point is calculated. Since the disparity of points on the background is generally smaller than that on the target surface, we remove 100 points of smaller disparity, and take the average of 200 remaining points as the final target disparity. The effect of the improved disparity calculation algorithm is shown on the right of Fig. 6. The disparity calculated by the improved algorithm has fewer outliers and is closer to the truth-value.

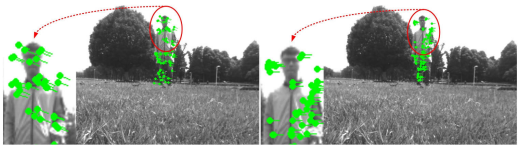


FIGURE 6. The disparity calculation.

B. COORDINATE SYSTEMS TRANSFORMATION

After the disparity of the target is obtained, the target relative position with the UAV is calculated by the stereo vision model which is expressed as P_c in the camera coordinate system. To obtain the target coordinate P_b in the body coordinate system and target coordinate P_w in the global coordinate system, we need to obtain the transforming relationship among these three coordinate systems, which is shown in Fig. 7.

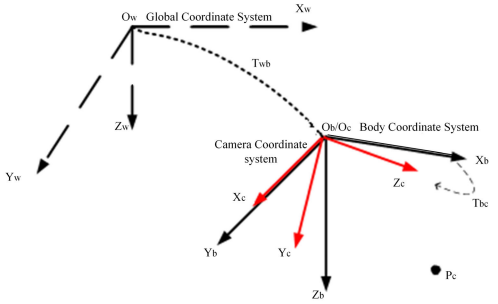


FIGURE 7. Coordinate systems transformation.

The transforming relationship of these three coordinate systems is expressed as eq. (6) and eq. (7)

$$\begin{bmatrix} \boldsymbol{P}_b \\ 1 \end{bmatrix} = \boldsymbol{T}_{bc}\boldsymbol{P}_c = \begin{bmatrix} \boldsymbol{R}_{bc} & \boldsymbol{t}_{bc} \\ \boldsymbol{0}^{\mathrm{T}} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{P}_c \\ 1 \end{bmatrix}. \tag{6}$$

$$\begin{aligned} \begin{bmatrix} \boldsymbol{P}_w \\ 1 \end{bmatrix} &= \boldsymbol{T}_{wb}\boldsymbol{T}_{bc}\boldsymbol{P}_c \\ &= \begin{bmatrix} \boldsymbol{R}_{wb}\boldsymbol{R}_{bc} & \boldsymbol{R}_{wb}\boldsymbol{t}_{bc} + \boldsymbol{t}_{wb} \\ \boldsymbol{0} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{P}_c \\ 1 \end{bmatrix} \\ &= \boldsymbol{T}_{wc}\boldsymbol{P}_c, \end{aligned} \tag{7}$$

where $\boldsymbol{R}_{wb}$ and $\boldsymbol{T}_{wb}$ (obtained from the MAVROS message) denote the rotation matrix and transformation matrix between the body coordinate system and global coordinate system, respectively. $\boldsymbol{R}_{bc}$ and $\boldsymbol{T}_{bc}$ denote the rotation matrix and transformation matrix between the camera coordinate system and body coordinate system, respectively. $\boldsymbol{T}_{wc}$ is the transformation matrix between the camera coordinate system and global coordinate system.

C. EXTENDED KALMAN FILTER

To track the target steadily and quickly, a stable global position and velocity are needed. We take the target position in the camera coordinate system and the rotation matrix between the camera coordinate system and global coordinate system as the inputs. The extended Kalman filter is presented to smooth the target global position and estimate the target global velocity, where the target motion model and measurement model are required.

We adopt the target position and velocity in global coordinates as state variables. The target motion model is

written as

$$\begin{bmatrix} X_t \\ Y_t \\ Z_t \\ \dot{X}_t \\ \dot{Y}_t \\ \dot{Z}_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & T & 0 & 0 \\ 0 & 1 & 0 & 0 & T & 0 \\ 0 & 0 & 1 & 0 & 0 & T \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \\ Z_{t-1} \\ \dot{X}_{t-1} \\ \dot{Y}_{t-1} \\ \dot{Z}_{t-1} \end{bmatrix} + \begin{bmatrix} \frac{T^2}{2} \\ \frac{T^2}{2} \\ \frac{T^2}{2} \\ T \\ T \\ T \end{bmatrix} a + \varepsilon(\boldsymbol{R_t}), \tag{8}$$

where T is the time period of this system and ε_t describes the uncertainty of the state distribution. $\boldsymbol{R}_t$ is the noise covariance matrix.

Assuming that the target acceleration conforms to a Gaussian distribution, the standard deviation is σ_a and the mean is 0. When the target velocity is changing, the variation in the estimated velocity is provided by σ_a. Since the coordinate in one direction is only related to the velocity in the same direction and is irrelevant to velocities and coordinates in the other direction, the noise covariance matrix of the motion model is

$$\boldsymbol{R}_t = \begin{bmatrix} \frac{T^4}{4} & 0 & 0 & \frac{T^3}{2} & 0 & 0 \\ 0 & \frac{T^4}{4} & 0 & 0 & \frac{T^3}{2} & 0 \\ 0 & 0 & \frac{T^4}{4} & 0 & 0 & \frac{T^3}{2} \\ \frac{T^3}{2} & 0 & 0 & T^2 & 0 & 0 \\ 0 & \frac{T^3}{2} & 0 & 0 & T^2 & 0 \\ 0 & 0 & \frac{T^3}{2} & 0 & 0 & T^2 \end{bmatrix} \sigma_a^2. \tag{9}$$

The measurement model is designed based on the state variables of the motion model. The target position in camera coordinate system (X_v, Y_v, Z_v) can be calculated by a stereo vision model, which is adopted as the measurement variable. The measurement model of the extended Kalman filter can be written as eq. (10),

$$\begin{bmatrix} X_v \\ Y_v \\ Z_v \end{bmatrix} = R^{-1} \begin{bmatrix} X_t - X_{ch} \\ Y_t - Y_{ch} \\ Z_t - Z_{ch} \end{bmatrix} + \delta(\boldsymbol{Q}_t), \tag{10}$$

where (X_{ch}, Y_{ch}, Z_{ch}) denotes the UAV position in the global coordinate system, R denotes the rotation matrix from the camera coordinate system to the global coordinate system and δ_t describes the uncertainty of the measurement.

The Jacobian matrix of the measurement model is

$$\boldsymbol{H}_t = \begin{bmatrix} \frac{\partial x_v}{\partial X_t} & \frac{\partial x_v}{\partial Y_t} & \frac{\partial x_v}{\partial Z_t} & \frac{\partial x_v}{\partial \dot{X}_t} & \frac{\partial x_v}{\partial \dot{Y}_t} & \frac{\partial x_v}{\partial \dot{Z}_t} \\ \frac{\partial y_v}{\partial X_t} & \frac{\partial y_v}{\partial Y_t} & \frac{\partial y_v}{\partial Z_t} & \frac{\partial y_v}{\partial \dot{X}_t} & \frac{\partial y_v}{\partial \dot{Y}_t} & \frac{\partial y_v}{\partial \dot{Z}_t} \\ \frac{\partial z_v}{\partial X_t} & \frac{\partial z_v}{\partial Y_t} & \frac{\partial z_v}{\partial Z_t} & \frac{\partial z_v}{\partial \dot{X}_t} & \frac{\partial z_v}{\partial \dot{Y}_t} & \frac{\partial z_v}{\partial \dot{Z}_t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}^{-1} & \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \end{bmatrix}. \tag{11}$$

Suppose that the three measurement variables have no relationship with each other, and obey a normal distribution of $N(0, N_t)$. Then, the covariance matrix $\boldsymbol{Q}_t$ is described as eq. (12)

$$\boldsymbol{Q}_t = \begin{bmatrix} \delta_x^2 & 0 & 0 \\ 0 & \delta_y^2 & 0 \\ 0 & 0 & \delta_z^2 \end{bmatrix}, \tag{12}$$

where δ_x, δ_y, δ_z denote the standard deviations of measurement in the x, y, z directions.

V. THE CONTROLLER DESIGN

To follow the target with the quadcopter, two following strategies are designed in this paper: a tail-following controller is based on the relative position between the UAV and target, and a parallel-following controller is based on the global position of the target and UAV.

A. THE TAIL-FOLLOWING CONTROLLER

The tail-following controller includes two closed-loop structures, which are shown in Fig. 8. We adopt the PIXHACK controller as the inner loop, and the outer loop includes the relative position controller and yaw angle controller.

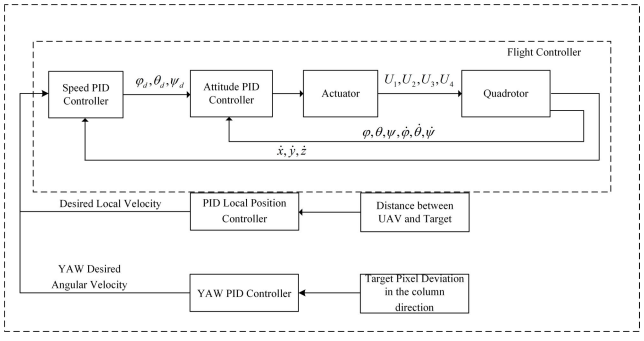


FIGURE 8. The tail following controller.

The UAV follows the target by adjusting the yaw angle and the speed in the direction of the quadcopter nose by the tail-following controller, which provides the shortest flight path of the UAV. However, the controller is only able to track the ground target, since the target height information cannot be obtained only by the relative position between the UAV and target.

B. THE PARALLEL-FOLLOWING CONTROLLER

To achieve accurate path following for the tracking target, the parallel-following strategy is further designed. The structure of the controller is shown in Fig. 9.

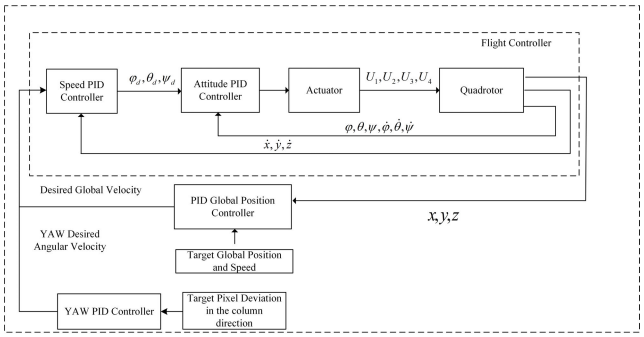


FIGURE 9. The parallel following controller.

Unlike the tail-following controller, the parallel-following controller involves a global position controller and yaw controller in the outer loop. The inputs of the global position controller are the global position and velocity of the target and UAV. The output is the desired global velocity of the UAV; thus, the parallel-following controller can be applied for tracking ground targets as well as air targets.

VI. SIMULATIONS AND EXPERIMENTS

In this section, simulations and experiments of a real system are conducted. We evaluate the error of target state estimation based on the ground truth of Gazebo, and test the effect of the target following controllers in the simulation environment. In experiments of a real system, the hardware is constructed, the software is designed and the communication flow of the whole system is first described. After that, to evaluate the performance of the real system, we carry out the following experiments on ground targets and air targets. More details of this system can be viewed in the video https://www.bilibili.com/video/BV1nK4y1D7sf/.

A. SIMULATIONS

To test the performance of the tracking system, a simulation is first carried out. This simulation system is built with Gazebo and based on the real system. The tracking of the ground target and air target are shown in Fig. 10a and Fig. 10b, respectively.

(a) Ground Target (b) Air Target

FIGURE 10. Tracking simulation experiments.

1) EVALUATION OF THE TARGET STATES ESTIMATION

The estimated position error of the ground target is shown in Fig. 11. The expected distance is set as 8 m and the target is randomly moving with a speed lower than 2 m/s. We only need to estimate the position in the x, y directions since the height of the target is constant. The estimated position error of the air target is shown in Fig. 12. The expected distance is set as 5 m and the target is randomly moving with a speed lower than 2 m/s in the x, y directions and 0.5 m/s in the z direction. It is observed that the measurement noise of the target location in the real world is close to 0.25, so we set δ_x, δ_y, δ_z in eq. (12) as 0.25. and set σ_a^2 in eq. (9) as 0.25.

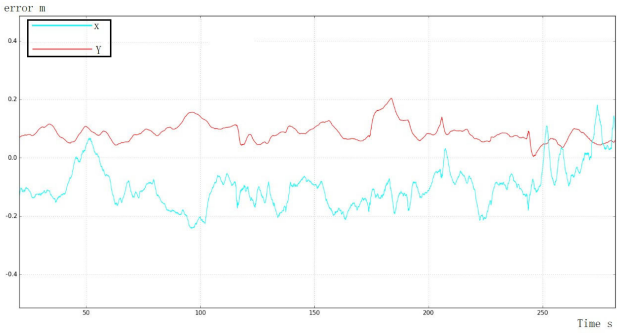


FIGURE 11. The position estimated error of ground target.

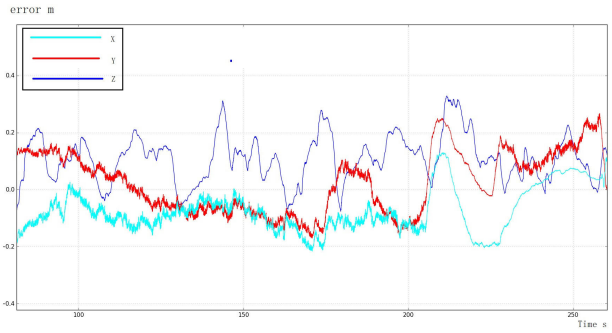


FIGURE 12. The position estimated error of air target.

It can be observed from Fig. 11 and Fig. 12 that the estimation errors in the x, y directions are lower than 0.2 m. Since the default position of the target is the center of the model and the estimated position is the center of the bounding box in the simulation scenario, the estimated overall error can be positive or negative. The estimated error of the air target is lower than 0.3 m in the z direction. The reason for the higher noise is that the appearance of the air target changes considerably when it flies up and down.

The estimated velocity error of the air target is shown in Fig. 13. From 340 s to 360 s, the target is stationary, whereafter, the target is moving with a speed between 0 m/s to 2 m/s.

The velocity error is very small when the target is stationary, and increases as the target moves. Particularly in the

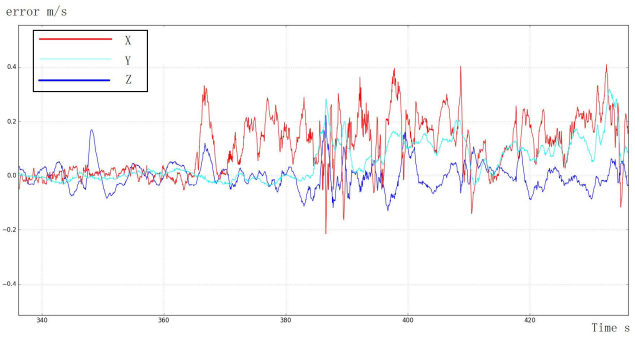


FIGURE 13. The error of estimated velocity.

stages of acceleration and deceleration, the error has a larger fluctuation. However, the error is always lower than 0.4 m/s, which shows the validity of the developed states estimation algorithm.

2) TARGET-FOLLOWING CONTROL SIMULATIONS

The experiment for the tail following controller is conducted first. The distance between the target and the UAV is shown in Fig. 14.

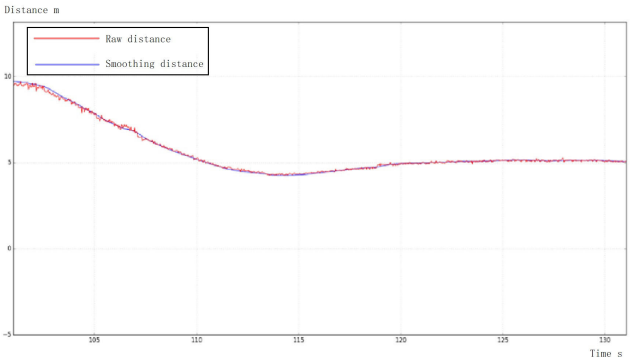


FIGURE 14. The distance in tail-following stage.

The expected distance is 5 m, and the original distance is set as 10 m. In Fig. 14, it can be seen that the relative distance between the UAV and the target is maintained at 5 m after 10 s. Fig. 15 shows the real time trajectory of the UAV and target.

FIGURE 15. The tail-following path.

In the tail-following mode, the UAV maintains the desired distance from the target and tracks the target with the shortest path. However, the following strategy needs to hold the UAV altitude.

The parallel-following controller is designed for tracking the ground target as well as the air target. The following experiments of the ground target and the air target with the parallel-following controller are shown in Fig. 16 and Fig. 17.

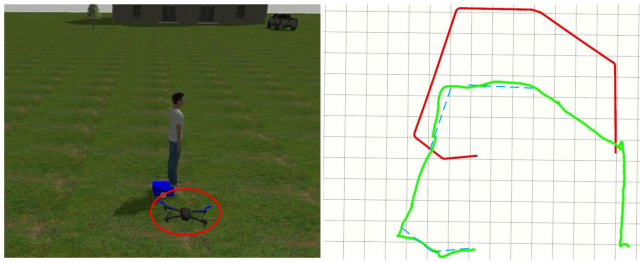


FIGURE 16. Ground target tracking for parallel-following controller.

FIGURE 17. Air target tracking for parallel-following controller.

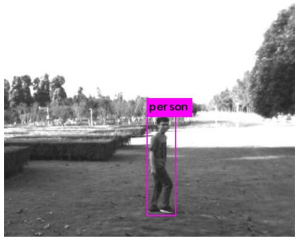
It is observed that the developed system is able to track ground and air targets with good performance. Only when the direction of target motion changes quickly will the tracking trajectory slightly deviate, which results from the estimated error of target global states and the restriction of the UAV kinematics.

B. EXPERIMENTS OF THE REAL SYSTEM

We further carry out experiments on ground targets and air targets with the parallel-following controller for the real platform. The hardware of the UAV platform is mainly composed of several parts: a MYNTEYE camera, an onboard NUC7I7BNH computer, a PIXHACK flight controller, and localized and dynamic components. The UAV platform is shown in Fig. 18. The image resolution is 752*480, and the

FIGURE 18. The quadcopter platform.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			



(a) The Structure of YOLOv3 (b) The Detection Result

FIGURE 19. Target detection algorithm.

frequencies of the camera and IMU are 25 Hz and 100 Hz, respectively.

We employ an Acer P248 computer as our ground station. For the convenience of system operation and system monitoring, we design visualization software in a ground station, which includes a target detection interface and a visualized upper computer. The YOLOv3 [34] algorithm is employed for target detection, the structure of which is shown in Fig. 19a, and the detection result is displayed in Fig. 19b.

The visualized upper computer is designed by C++ based on the QT structure. In addition to the basic functions including UAV flight control, displaying the states of the target and adjusting the parameters of the controller, the upper computer software also has the ability to launch the whole system using one button. It is shown in Fig. 20.

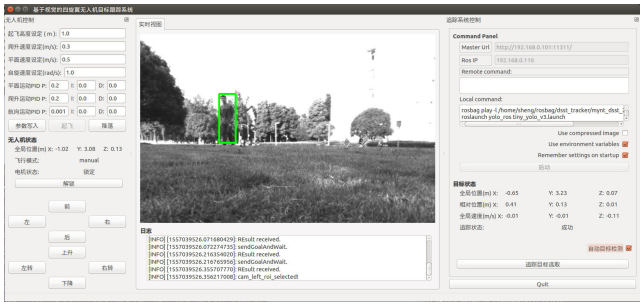


FIGURE 20. The visualization of upper computer.

The ground station communicates with the NUC7I7BNH computer by WI-FI and communicates with the flight controller by radio. The communication flow of the whole system is shown in Fig. 21.

First, we adopt the pedestrian as the ground target. The UAV takes off to a height of 1.5 m with the control of the ground station, and then we choose the target among objects detected by YOLOv3, where the expected distance between the target and UAV is set as 5.5 m. It is observed that the measurement noise of the target location in the real world is

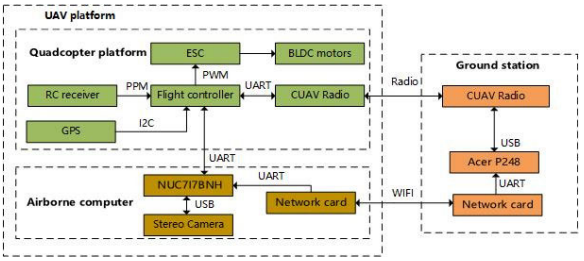


FIGURE 21. The hardware communication of target tracking system.

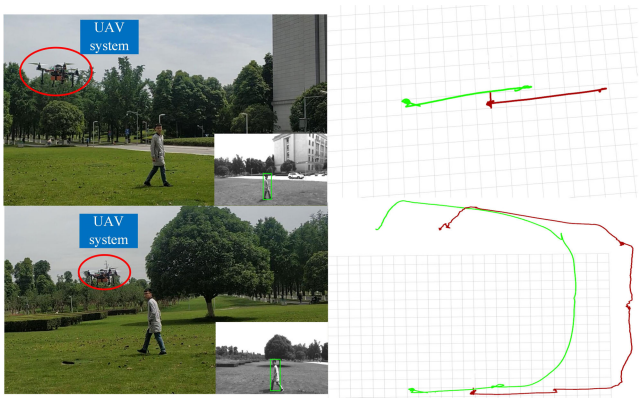


FIGURE 22. Ground target tracking experiment in real world.

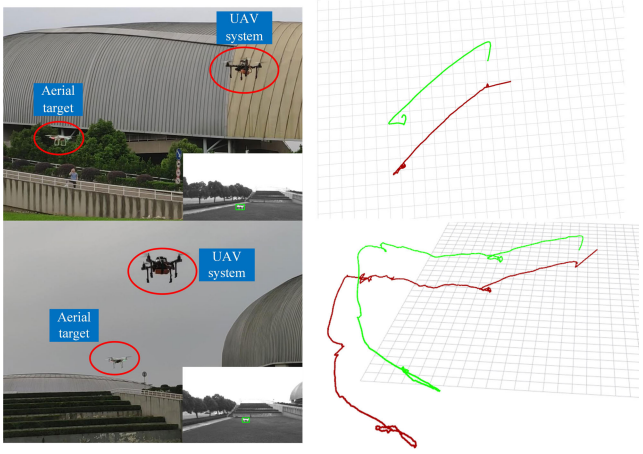


FIGURE 23. Air target tracking experiment in real world.

close to 0.4, so we set δ_x, δ_y, δ_z in eq. (12) respectively as 0.4 and set σ_a^2 in eq. (9) as 0.25.

Fig. 22 shows the actual trajectories of the UAV and target at different times. In this scenario, our target tracking system exhibits good performance with a parallel trajectory.

In air target tracking experiments on a real platform, we employ DJI PHANTOM2 as an air target. The expected distance between the target and tracking system is set as 4 m, and for security, the expected height of the tracking system is 1 m above the target.

Fig. 23 shows the tracking scenario of the air target tracking experiment. Our system has good trajectory-following

performance for tracking air target, which indicates that the tracking algorithm and target state estimation method are effective and stable even if the target is small and moving fast.

VII. CONCLUSION

In this paper, an autonomous vision-based target tracking system is designed and implemented for a quadcopter platform which is developed in our lab. The target is easily selected by the detection result of YOLOv3. Utilizing a new update strategy of a tracking filter and an additional drift correction step, our system is able to track targets accurately in the long term. Meanwhile, a novel target redetection method is developed, which is able to relocate the target accurately in an effective way. An improved LK optical flow and an extended Kalman filter algorithm can estimate both local and global states of the target, which correspond to the inputs of the tail-following controller and parallel-following controller, respectively. Simulation and real-world experimental results show that the developed tracking system achieves stable tracking performance. In the future, a new algorithm will be investigated to track the target indoors by combining this system with indoor localization and obstacle avoidance methods. Meanwhile, 3D object detection and multi-target tracking algorithms will be considered for tracking tasks.

REFERENCES

[1] T. Tomic, K. Schmid, P. Lutz, A. Domel, M. Kassecker, E. Mair, I. Grixa, F. Ruess, M. Suppa, and D. Burschka, "Toward a fully autonomous UAV: Research platform for indoor and outdoor urban search and rescue," *IEEE Robot. Autom. Mag.*, vol. 19, no. 3, pp. 46–56, Sep. 2012, doi: 10.1109/MRA.2012.2206473.

[2] C. Wang, P. Liu, T. Zhang, and J. Sun, "The adaptive vortex search algorithm of optimal path planning for forest fire rescue UAV," in *Proc. IEEE 3rd Adv. Inf. Technol., Electron. Autom. Control Conf. (IAEAC)*, Oct. 2018, pp. 400–403, doi: 10.1109/IAEAC.2018.8577733.

[3] G. Bevacqua, J. Cacace, A. Finzi, and V. Lippiello, "Mixed-initiative planning and execution for multiple drones in search and rescue missions," in *Proc. Int. Conf. Autom. Plan. Sched. (ICAPS)*, Apr. 2015, pp. 315–323.

[4] S. Siebert and J. Teizer, "The adaptive vortex search algorithm of optimal path planning for forest fire rescue UAV," *Autom. Constr.*, vol. 41, pp. 1–14, May 2014, doi: 10.1016/j.autcon.2014.01.004.

[5] R. W. Beard, T. W. McLain, D. B. Nelson, D. Kingston, and D. Johanson, "Decentralized cooperative aerial surveillance using fixed-wing miniature UAVs," *Proc. IEEE*, vol. 94, no. 7, pp. 1306–1323, Aug. 2006, doi: 10.1109/JPROC.2006.876930.

[6] Y. Qu, L. Jiang, and X. Guo, "Moving vehicle detection with convolutional networks in UAV videos," in *Proc. 2nd Int. Conf. Control, Autom. Robot. (ICCAR)*, Apr. 2016, pp. 225–229, doi: 10.1109/ICCAR.2016.7486730.

[7] N. Batsoyol, Y. Jin, and H. Lee, "Constructing full-coverage 3D UAV ad-hoc networks through collaborative exploration in unknown urban environments," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–7, doi: 10.1109/ICC.2018.8422396.

[8] H. Qin, Z. Meng, W. Meng, X. Chen, H. Sun, F. Lin, and M. H. Ang, "Autonomous exploration and mapping system using heterogeneous UAVs and UGVs in GPS-denied environments," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1339–1350, Feb. 2019, doi: 10.1109/TVT.2018.2890416.

[9] B. Y. Li, H. Lin, H. Samani, L. Sadler, T. Gregory, and B. Jalaian, "On 3D autonomous delivery systems: Design and development," in *Proc. Int. Conf. Adv. Robot. Intell. Syst. (ARIS)*, Sep. 2017, pp. 1–6, doi: 10.1109/ARIS.2017.8361592.

[10] R. P. Padhy, F. Xia, S. K. Choudhury, P. K. Sa, and S. Bakshi, "Monocular vision aided autonomous UAV navigation in indoor corridor environments," *IEEE Trans. Sustain. Comput.*, vol. 4, no. 1, pp. 96–108, Jan. 2019, doi: 10.1109/TSUSC.2018.2810952.

[11] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1409–1422, Jul. 2012.

[12] M. Danelljan, G. Häger, F. Shahbaz Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proc. Brit. Mach. Vis. Conf.*, 2014, pp. 2–12, doi: 10.5244/C.28.65.

[13] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015, doi: 10.1109/TPAMI.2014.2345390.

[14] D. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2544–2550, doi: 10.1109/CVPR.2010.5539960.

[15] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4310–4318, doi: 10.1109/ICCV.2015.490.

[16] A. Lukezic, T. Vojir, L. C. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4847–4856, doi: 10.1109/CVPR.2017.515.

[17] C. Ma, X. Yang, C. Zhang, and M.-H. Yang, "Long-term correlation tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5388–5396, doi: 10.1109/CVPR.2015.7299177.

[18] H. Fan and H. Ling, "Parallel tracking and verifying," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4130–4144, Aug. 2019, doi: 10.1109/TIP.2019.2904789.

[19] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, "Fully-convolutional siamese networks for object tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 850–865, doi: 10.1007/978-3-319-48881-3_56.

[20] N. Wang, W. Zhou, and H. Li, "Reliable re-detection for long-term tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 3, pp. 730–743, Mar. 2019, doi: 10.1109/TCSVT.2018.2816570.

[21] K. Kaaniche, B. Champion, C. Pegard, and P. Vasseur, "A vision algorithm for dynamic detection of moving vehicles with a UAV," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2005, pp. 1878–1883, doi: 10.1109/ROBOT.2005.1570387.

[22] V. N. Dobrokhodov, I. I. Kaminer, K. D. Jones, and R. Ghabcheloo, "Vision-based tracking and motion estimation for moving targets using small UAVs," in *Proc. Amer. Control Conf.*, 2006, pp. 1428–1433, doi: 10.1109/acc.2006.1656418.

[23] S. Azrad, F. Kendoul, D. Perbrianti, and K. Nonami, "Visual servoing of an autonomous micro air vehicle for ground object tracking," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2009, pp. 5321–5326, doi: 10.1109/IROS.2009.5354728.

[24] G. R. Rodríguez-Canosa, S. Thomas, J. del Cerro, A. Barrientos, and B. MacDonald, "A real-time method to detect and track moving objects (DATMO) from unmanned aerial vehicles (UAVs) using a single camera," *Remote Sens.*, vol. 4, no. 4, pp. 1090–1111, Apr. 2012, doi: 10.3390/rs4041090.

[25] J. E. Gomez-Balderas, G. Flores, L. R. G. Carrillo, and R. Lozano, "Tracking a ground moving target with a quadrotor using switching control," *J. Intell. Robotic Syst.*, vol. 70, nos. 1–4, pp. 65–78, Apr. 2013, doi: 10.1007/s10846-012-9747-9.

[26] J. Chen, T. Liu, and S. Shen, "Tracking a moving target in cluttered environments using a quadrotor," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 446–453, doi: 10.1109/IROS.2016.7759092.

[27] J. Thomas, J. Welde, G. Loianno, K. Daniilidis, and V. Kumar, "Autonomous flight for detection, localization, and tracking of moving targets with a small quadrotor," *IEEE Robot. Autom. Lett.*, vol. 2, no. 3, pp. 1762–1769, Jul. 2017, doi: 10.1109/LRA.2017.2702198.

[28] A. Rohan, M. Rabah, and S.-H. Kim, "Convolutional neural network-based real-time object detection and tracking for parrot AR drone 2," *IEEE Access*, vol. 7, pp. 69575–69584, 2019, doi: 10.1109/ACCESS.2019.2919332.

[29] M. Rabah, A. Rohan, M.-H. Haghbayan, J. Plosila, and S.-H. Kim, "Heterogeneous parallelization for object detection and tracking in UAVs," *IEEE Access*, vol. 8, pp. 42784–42793, 2020, doi: 10.1109/ACCESS.2020.2977120.
[30] M. Rabah, A. Rohan, S. A. S. Mohamed, and S.-H. Kim, "Autonomous moving target-tracking for a UAV quadcopter based on fuzzy-PI," *IEEE Access*, vol. 7, pp. 38407–38419, 2019, doi: 10.1109/ACCESS.2019.2906345.
[31] K. Boudjit, C. Larbes, and N. Ramzan, "ANN design and implementation for real-time object tracking using quadrotor AR.Drone 2.0," *J. Experim. Theor. Artif. Intell.*, vol. 30, no. 6, pp. 1013–1035, Nov. 2018, doi: 10.1080/0952813X.2018.1509896.
[32] A. Al-Kaff, M. J. Gomez-Silva, F. M. Moreno, A. De La Escalera, and J. M. Armingol, "An appearance-based tracking algorithm for aerial search and rescue purposes," *Sensors*, vol. 19, no. 3, pp. 652–682, 2019, doi: 10.3390/s19030652.
[33] H. Cheng, L. Lin, Z. Zheng, Y. Guan, and Z. Liu, "An autonomous vision-based target tracking system for rotorcraft unmanned aerial vehicles," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 1732–1738, doi: 10.1109/IROS.2017.8205986.
[34] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: http://arxiv.org/abs/1804.02767
[35] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Discriminative scale space tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 8, pp. 1561–1575, Aug. 2017, doi: 10.1109/TPAMI.2016.2609928.
[36] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Adaptive correlation filters with long-term and short-term memory for object tracking," *Int. J. Comput. Vis.*, vol. 126, no. 8, pp. 771–796, Aug. 2018, doi: 10.1007/s11263-018-1076-4.
[37] M. Wang, Y. Liu, and Z. Huang, "Large margin object tracking with circulant feature maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4800–4808, doi: 10.1109/CVPR.2017.510.
[38] Y. Liu, Q. Wang, H. Hu, and Y. He, "A novel real-time moving target tracking and path planning system for a quadrotor UAV in unknown unstructured outdoor scenes," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 11, pp. 2362–2372, Nov. 2019, doi: 10.1109/TSMC.2018.2808471.
[39] S. Baker, R. Patil, G. Cheung, and I. Matthews, "Lucas-Kanade 20 years on: An unifying framework: Part 5," *Int. J. Comput. Vis.*, vol. 56, no. 3, pp. 221–255, 2004, doi: 10.1023/B:VISI.0000011205.11775.fd.
[40] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 1280–1286, doi: 10.1109/IROS.2013.6696514.

SHICHENG WU received the B.E. degree in automation from Zhengzhou University, Zhengzhou, China, in 2018. He is currently pursuing the M.S. degree with the University of Electronic Science and Technology of China. His research interests include UAV systems, visual object tracking, navigation, and state estimation.

RUI LI (Senior Member, IEEE) received the Ph.D. degree from the Harbin Institute of Technology, Harbin, China, in 2008. She is currently an Associate Professor with the University of Electronic Science and Technology of China. Her research interests include multi-agent systems, UAV control, and principle of automatic control.

YINGJING SHI (Member, IEEE) received the Ph.D. degree from the Harbin Institute of Technology, Harbin, China, in 2008. He is currently an Associate Professor with the University of Electronic Science and Technology of China. His research interests include principle of automatic control, aircraft control, and multi-agent systems.

QISHENG LIU received the B.E. degree in automation from Nanchang University, Nanchang, China, in 2016, and the M.S. degree from the University of Electronic Science and Technology of China, in 2019. His research interests include computer vision, machine learning, deep learning, and intelligence driver.