

# An Autonomous Vision-Based Target Tracking System for Rotorcraft Unmanned Aerial Vehicles

H. Cheng, L. S. Lin, Z. Q. Zheng, Y. W. Guan and Z. C. Liu

**Abstract**—In this paper, an autonomous vision-based tracking system is presented to track a maneuvering target for a rotorcraft unmanned aerial vehicle (UAV) with an onboard gimbal camera. To handle target occlusions or loss for real-time tracking, a robust and computationally efficient visual tracking scheme is considered using the Kernelized Correlation Filter (KCF) tracker and the redetection algorithm. The states of the target are estimated from the visual information. Moreover, feedback control laws of the gimbal and the UAV using the estimated states are proposed for the UAV to track the moving target autonomously. The algorithms are implemented on an onboard TK1 computer, and extensive outdoor flight experiments have been performed. Experimental results show that the proposed computationally efficient visual tracking scenario can stably track a maneuvering target and is robust to target occlusions and loss.

## I. INTRODUCTION

Recently, unmanned aerial vehicles (UAVs) have attracted increasing attention from both industrial and academic communities [1]. Vertical takeoff and landing (VTOL) unmanned rotorcrafts equipped with visual sensors have broad applications including environmental monitoring, rescue and search, surveillance, traffic control, etc [2][3][4].

UAVs that can autonomously track maneuvering targets are demanding in a wide range of applications. Many research efforts have been devoted to autonomous tracking for UAVs. However, most of current research work focuses either on visual tracking schemes or on control laws of UAVs [5][6]. Although various visual tracking algorithms with high performance have been developed [5], most of the algorithms are computationally complicated and not suitable for real-time tracking of UAVs with limited onboard computation capacities. The KCF tracker [7] is applied in our autonomous vision-based tracking system for its computational efficiency and impressive performance. In [8][9], visual tracking scenarios of autonomous helicopters have been studied via simulations. Several visual algorithms to track the predefined targets have been implemented on UAVs [10], where simple trackers such as the color and shape filtering are used for specific target tracking [11][12][13]. In [14][15], the OpenTLD [14] and the clustering of static adaptive correspondences (CMT) trackers robust to tracking deformable targets are used for AR Drones to track ground targets autonomously, where the output of the trackers is used for the PD controller and the state estimation of the



Fig. 1. An instant of our flight experiments. The upper left corner is an image of the onboard camera, where the green rectangle is the bounding box indicating the current state of the target.

moving target is not considered. To track a maneuvering target precisely and stably, an efficient state estimator is desirable to estimate the real-time states of the moving target. The gimbal camera is beneficial to alleviate video jiggle and hence improves the stability of the trackers [10]. A visual tracking system using a gimbal camera was proposed in [16], which provides a robust implementation on a helicopter to track predefined targets.

There are some challenging problems for maneuvering target tracking of UAVs: 1) A visual tracker robust to target occlusions and loss is necessary. 2) Accurate state estimation of the target and closed-loop control laws of the gimbal and the UAV should be developed to stably track the maneuvering target. 3) Due to the limited computational capacities, highly computational efficiency of visual tracking, state estimation as well as control algorithms are desirable for onboard implementation to perform real-time tracking.

These challenging problems motivate us to systematically investigate vision-based tracking of UAVs considering both visual trackers and the control laws. Fig. 1 illustrates our tracking system, where a quadcopter is tracking a person autonomously using an onboard gimbal camera. The contributions of the paper are summarized as follow: 1) In the case of target occlusions or loss, the status of the target, i.e. loss or not, is firstly detected based on the KCF tracker, and a computationally efficient redetection method is presented. With this scheme, the UAV can track the target again when it re-appears. 2) An Interacting Multi-Model Extended Kalman Filtering (IMM-EKF) based target state estimator is presented to estimate states of the maneuvering target, and a nonlinear feedback control law is presented to stably track moving targets. 3) A computationally efficient framework implemented on onboard TK1 computer is presented for

\*This work is supported by NSFC-Shenzhen Robotics Projects (U1613211), the Fundamental Research Funds for the Central Universities, and Guangdong Natural Science Foundation (1614050001452).

The authors are with the School of Data and Computer Science, Sun Yat-sen University, China (email: chengh9@mail.sysu.edu.cn).

ground target tracking in unstructured environments.

The rest of the paper is organized as follows: In Section II, we introduce the architecture of the vision-based tracking system. In Section III, we present the KCF tracker, target loss detection and redetection schemes. Section IV addresses the target state estimation algorithms. The control of the gimbal and the UAV is presented in Section V. Experimental results are presented in Section VI. Concluding remarks and future work are discussed in Section VII.

## II. SYSTEM CONFIGURATION

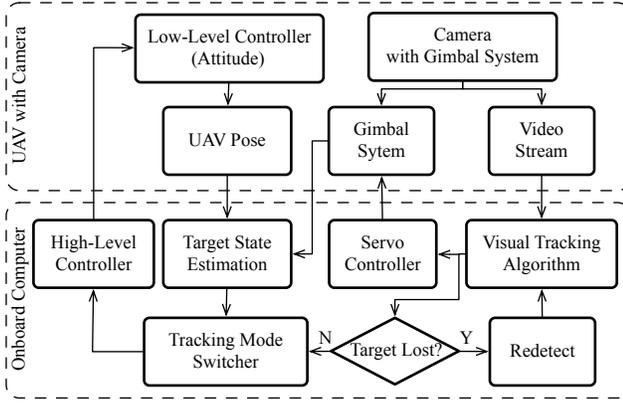


Fig. 2. Architecture of the vision-based tracking system.

A DJI Matrice100 is used as the UAV platform, which is equipped with an onboard TK1 computer and a monocular RGB gimbal camera. An overview of the system configuration is shown in Fig. 2. The gimbal camera mounted on the UAV platform provides the video stream and internal angles for the onboard computer. The visual tracking algorithm obtains position of the target on the image plane, which is feedback to the gimbal controller. In addition, the states of the target are estimated by fusing the inertia measurement unit (IMU) data of the UAV platform and the gimbal. A switching tracking strategy is performed based on the estimated states. The high-level controller computes the desired velocities of the UAV, and the low-level controller controls the attitude correspondingly. The frequencies of the video stream and the control signal are 30Hz and 10Hz, respectively.

## III. A VISUAL TRACKING SCHEME

In this section, a computationally efficient visual tracking scheme robust to target occlusions and loss is presented, which consists of the KCF tracker, the target loss detection and the redetection methods. Fig. 3 shows the configuration of the visual tracking scheme. Firstly, the KCF tracker estimates the state of the target. The status of the target, i.e. loss or not, is then detected based on the regression function of the KCF tracker. Finally, a redetection method is presented to track the target when it come out off ccullisions.

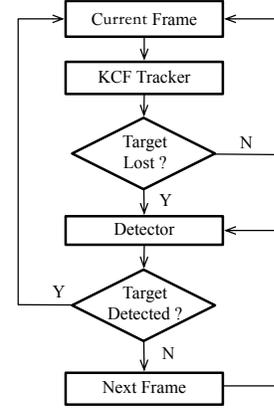


Fig. 3. Architecture of the vision system.

### A. KCF Tracker

KCF develops from Circulant Structure of Tracking-by-detection with Kernels (CSK) [17], applying online learning methods to solve tracking problems. More precisely, it is a machine learning method without any prior knowledge. At the first frame, the object of interested (OOI) region is chosen manually and the KCF tracker transforms the region into a multi-channel HOG feature descriptor. A regression function  $f(\mathbf{z})$  of OOI region  $\mathbf{z}$  is initialized by Ridge Regression with HOG descriptor. For the new frame,  $f(\mathbf{z})$  is evaluated on several regions around the last region of OOI. Finally, the region which has max response of evaluation is considered as the output and applied to update  $f(\mathbf{z})$ .

To accelerate the matrix computation of Ridge Regression, KCF transforms each channel of HOG feature descriptor into a circulant matrix by cyclic shifting. It is known that circulant matrix can be made diagonal by Discrete Fourier Transform(DFT) [18]. Thus matrix computation, especially matrix inversion, can be efficiently processed in fourier domain. Furthermore, a kernel function, which maps the regression function  $f(\mathbf{z})$  into non-linear space, is applied in the KCF tracker to promote the performance of tracking. These solutions are introduced by CSK, and optimized in KCF. In this way, the process speed and mean precision of KCF have reached 172FPS and 73.2% respectively. More details of the KCF tracking algorithm may refer to [7].

### B. Target Loss Detection and Redetection

Various visual trackers have been proposed to tackle illumination variation, scale variation and occlusion problems [19]. However, most of current algorithms can not detect the target in the presence of full occlusions, which are often encountered in tracking mission of UAVs. In this paper, a simple and efficient redetection method is proposed so that the target can be estimated when it appears again.

The status of the target, i.e. loss or not, is estimated by the target loss detection based on the KCF tracker. The max response of regression function  $f_{max}(\mathbf{z})$ , denotes the relevance between the OOI region and the target. When the value of  $f_{max}(\mathbf{z})$  is less than a threshold, it implies that the

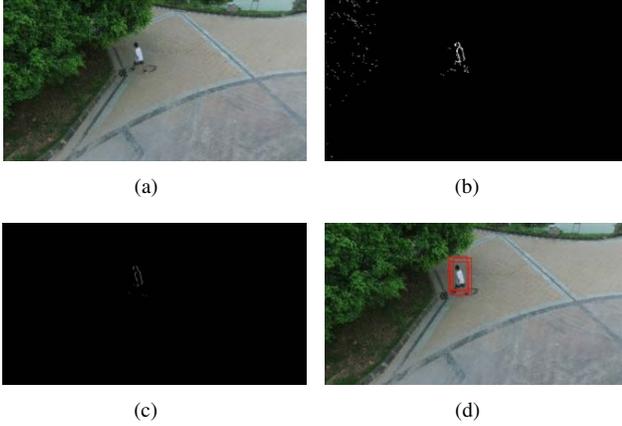


Fig. 4. A moving object detector based on the Frame-Difference method. (a) the original frame; (b) the difference frame with noise; (c) the noise is removed and the foreground is retained using the Gaussian blur; (d) the bounding boxes constructed by the moving object detector.

target may be lost. The redetection works when  $f_{max}(\mathbf{z})$  is less than a threshold. According to flight experiments, the values of  $f_{max}(\mathbf{z})$  vary in the range of (0, 0.5) in outdoor environments. The threshold is experimentally set as 0.17.

The UAV hovers and starts to search the target when the target loss is detected. Some classical algorithms [14] scan all pixels in the new frame to search the target, and the computational complexity is high. In general, the target is moving when it re-appears in the view of the camera. Hence, the target can be estimated by detecting the moving foreground, instead of searching all pixels in the new frame.

In the paper, a moving object detector based on the Frame-Difference (FD) method [20] is applied, as shown in Fig. 4. In specific, the detector subtracts the last frame from the current frame to obtain the difference image. Although the FD method is computationally efficient, it is sensitive to noise, as shown in Fig. 4(b). The Gaussian blur is hence applied to remove the noise in the difference image. Consequently, the detector constructs the bounding boxes based on the center of the foreground, as shown in Fig. 4(d). It is noted that the size of these boxes is the same as the initial OOI region. Finally, the regions contained by these bounding boxes are evaluated by the regression function. The region, which has the maximum value of  $f_{max}(\mathbf{z})$  and is greater than the threshold, is selected as the position of the target in the new frame.

#### IV. GROUND MANEUVERING TARGET STATE ESTIMATION

The states of the ground maneuvering target are estimated based on the extended Kalman Filtering.

##### A. Distance Estimation Method

As shown in Fig. 5, the relative distance between the UAV and the target is estimated.

The distance between the camera and the UAV center is assumed to be neglectable. Let  $F_B$  denote the body frame of UAV with axes  $X_b$ ,  $Y_b$  and  $Z_b$ , and  $F_C$  denotes the camera's

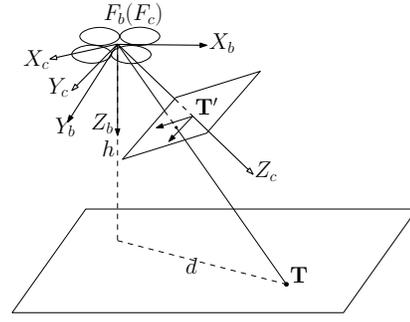


Fig. 5. The relationships between the UAV, the camera and the target.

reference frame with axis  $X_c$ ,  $Y_c$  and  $Z_c$ . The relationships between the ground target  $T$  and the UAV can be shown in Fig. 5. Thus, the transformation of a vector from  $F_C$  to  $F_B$  can be represented by a rotation matrix  $\mathbf{R}_{BC}$ . The target is considered as a point  $T$  on the ground and is represented by a position vector  $\mathbf{p}_B = (x_t, y_t, z_t)^T$  in body frame  $F_B$  of UAV. According to standard pinhole imaging model,  $\mathbf{p}_B$  can be written as:

$$\mathbf{p}_B \sim \mathbf{R}_{BC} \mathbf{K}^{-1}(u, v, 1)^T \quad (1)$$

where the homogenous coordinate  $(u, v, 1)^T$  indicates the position of the target on the image plane, and  $\mathbf{K}$  is the intrinsic matrix of the camera. Hence, the relative distance between the target and UAV can be calculated by

$$d = \frac{h}{z_t} \sqrt{x_t^2 + y_t^2} \quad (2)$$

where  $h$  is the altitude of the UAV.

##### B. Extended Kalman Filtering

The estimated relative position between UAV and target given by Eq. 2 is generally inaccurate due to observation noise. To track the maneuvering target stably and precisely, the velocity and acceleration of the target need to be estimated using appropriate motion models. A single motion model is difficult to accurately describe random movements of the maneuvering target, the IMM-EKF [21] algorithm is presented to estimate its states by fusing the constant velocity model and the current statistical model. In particular, the constant velocity model can be written in a discrete form:

$$\mathbf{X}(k+1) = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \mathbf{X}(k) + \begin{bmatrix} t^2/2 \\ t \end{bmatrix} \mathbf{w}(k), \quad (3)$$

where  $t$  is the sampling interval,  $\mathbf{X}$  is state vector, and  $\mathbf{w}$  is a discrete white process noise.

The current statistical model in discrete form is given as:

$$\mathbf{X}(k+1) = \Phi(k) \mathbf{X}(k) + \mathbf{U}(k) \bar{a}(k) + \mathbf{w}(k), \quad (4)$$

where  $\Phi(k)$  is the state matrix,  $\mathbf{U}(k)$  is the control matrix,  $\bar{a}(k)$  is the mean of current maneuvering acceleration, and  $\mathbf{w}(k)$  is a discrete white process noise. It is noted that the employed model is a Singer model with an adaptive mean [22], which does not require any prior model and can handle targets with rapidly changing speeds.

## V. CONTROLLER

In this section, feedback control laws are presented for both the gimbal and the UAV with the visual feedback information obtained in Section III and IV. The gimbal and the UAV are controlled simultaneously to perform stable tracking.

### A. Gimbal Controller

A closed-loop controller is developed for controlling the pitch angle and yaw angle of gimbal to keep the target in the view of the camera. As the video stream is captured in a frequency of 30Hz, we assume that the targets acceleration would not change too much within the time interval. The Constant Accelerate(CA) model [22] is applied to forecast the target's state in the 2D image coordinate. Let  $u(k)$  denote the targets position along the x axis on the image coordinates. The position  $u(k+1)$  at the time steps  $k+1$  can be estimated by the CA model as:

$$\begin{bmatrix} u(k+1) \\ \dot{u}(k+1) \\ \ddot{u}(k+1) \end{bmatrix} = \begin{bmatrix} 1 & \Delta t & \frac{1}{2}\Delta t^2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u(k) \\ \dot{u}(k) \\ \ddot{u}(k) \end{bmatrix} \quad (5)$$

where  $\Delta t$  is the time increment between time step  $k$  and  $k+1$ .  $\dot{u}(k)$  and  $\ddot{u}(k)$  are obtained through computing the difference of 4 recent frames:

$$\begin{cases} \dot{u}(k) = \frac{1}{4\Delta t} (u(k-1) - u(k-3) + u(k-2) - u(k-4)) \\ \ddot{u}(k) = \frac{1}{4\Delta t^2} (u(k-1) - u(k-3) - u(k-2) + u(k-4)) \end{cases} \quad (6)$$

Similarly, the target's position  $v(k)$  along y axis on the image coordinates can be estimated. A PD controller is used for the gimbal system to compute its angular velocities in the 2D image plane, i.e.  $e_\alpha(k) = u(k+1) - u_0$  and  $e_\beta(k) = v(k+1) - v_0$ , where  $(u_0, v_0)$  is the center point of the image.

### B. Switchable Tracking Strategy

To perform smooth and stable tracking of a maneuvering target, a switchable tracking strategy consisting of the observing mode and the following mode is presented considering the relative distance  $d$ . Switching between the two modes is determined by the thresholds  $d_{\min}$  and  $d_{\max}$ . The thresholds are calculated by a certain range of the pitch angle  $[\theta_1, \theta_2]$  and the real-time altitude of the UAV:

$$\begin{cases} d_{\max} = h \cdot \tan \theta_2 \\ d_{\min} = h \cdot \tan \theta_1 \end{cases} \quad (7)$$

We use  $\theta_1 = 20^\circ$  and  $\theta_2 = 70^\circ$  for the thresholds, because within this range our relative distance estimator has better results which will be analyzed in Section VI-C.

1) *Observing Mode:* In the observing mode, the UAV only adjusts its yaw angle to the target and there is no horizontal displacement. When the relative distance  $d$  satisfies  $d_{\min} \leq d \leq d_{\max}$ , the observing mode works. In addition, when the targets acceleration  $\mathbf{a}$  estimated from the IMM-EKF, the UAVs maximum acceleration  $\mathbf{a}_{\max}$  and the current velocity  $\mathbf{v}_p$  of the UAV satisfy the following conditions:  $|\mathbf{a}| > |\mathbf{a}_{\max}|$  and  $\mathbf{a} \cdot \mathbf{v}_p < 0$ , which implies that the target's relative distance tends to decrease to the condition of  $d \leq d_{\max}$ . The

UAV rotates to track the ground target while maintaining a certain height and position. A PID controller is utilized to control the UAVs yaw angle.

2) *Following Mode:* In the following mode, the moving of UAV and the rotation of the gimbal are controlled simultaneously to maintain the yaw angle of the gimbal in the UAVs body frame close to zero. When the relative distance  $d \leq d_{\min}$  and  $d \geq d_{\max}$ , the following mode works.

In the following mode, the nonlinear velocity controller of the UAV is proposed based on the Lyapunov theory. Without loss of generality, we assume that the UAV maintains a constant altitude during the tracking process. The target tracking flight control is hence simplified to a planar control problem. The relationship between the UAV and the target in the XY plane will be considered as shown in Fig. 6.

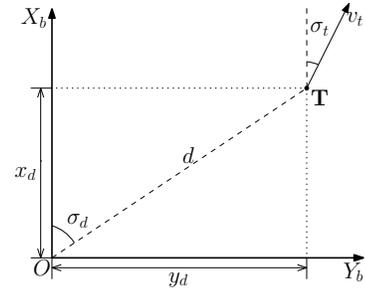


Fig. 6. Relationship between the aircraft and the target in the UAV's body frame  $X_b O Y_b$ .

Let  $d$  and  $\sigma_d$  be the target's relative position and relative yaw angle estimated from the IMM-EKF. In the following mode, we expect that the UAV would maintains a certain distance to the target. Therefore, we define the distance error and angular error as follows:

$$\begin{cases} \varepsilon_d = d - E(d) \\ \varepsilon_\sigma = \sigma_d, \end{cases} \quad (8)$$

where

$$E(d) = \begin{cases} d_{\min} & d \leq d_{\min} \\ d_{\max} & d > d_{\max} \end{cases} \quad (9)$$

It is noted that the relative position and  $d$  and relative yaw angle  $\sigma_d$  can also be formulated as  $d = \sqrt{x_d^2 + y_d^2}$  and  $\sigma_d = \arctan(y_d/x_d)$ , then the dynamic error of distance  $\dot{\varepsilon}_d$  and dynamic error of yaw angle  $\dot{\varepsilon}_\sigma$  can be written as:

$$\begin{cases} \dot{\varepsilon}_d = \dot{d} = -v_x \cos \varepsilon_\sigma - v_y \sin \varepsilon_\sigma + v_d \cos(\sigma_t - \sigma_d) \\ \dot{\varepsilon}_\sigma = \dot{\sigma}_d = \frac{\sin \varepsilon_\sigma}{d} v_x - \frac{\cos \varepsilon_\sigma}{d} v_y + \frac{\sin(\sigma_t - \sigma_d)}{d} v_d \end{cases} \quad (10)$$

where  $v_x$  and  $v_y$  are velocities of the UAV in the body frame,  $v_t$  and  $\sigma_t$  are velocity and its yaw angle of the target estimated from the IMM-EKF with respect to the body frame of the UAV.

We seek to control the velocity  $v_x$ ,  $v_y$  and angular velocity  $\omega$  of the UAV to ensure the distance error  $\varepsilon_d$  and angular error  $\varepsilon_\sigma$  converge to zero. The control law of the UAV is

designed as:

$$\begin{cases} v_x = k_1 \varepsilon_d \cos \varepsilon_\sigma + v_d \cos(\sigma_t - \sigma_d) \cos \varepsilon_\sigma \\ v_y = k_1 \varepsilon_d \sin \varepsilon_\sigma + v_d \cos(\sigma_t - \sigma_d) \sin \varepsilon_\sigma \\ \omega = k_2 \varepsilon_\sigma + \frac{v_d}{d} \sin(\sigma_t - \sigma_d) \end{cases} \quad (11)$$

The stability of the feedback control system can be proved using Lyapunov's second theory. The Lyapunov function candidate can be formulated as:

$$V_1 = \frac{1}{2} (\varepsilon_d^2 + \varepsilon_\sigma^2) \quad (12)$$

Note that,  $V_1 \geq 0$  and  $V_1 = 0$  if and only if  $[\varepsilon_d \ \varepsilon_\sigma]^T = [0 \ 0]^T$ . The time derivative of  $V_1$  can be written as:

$$\begin{aligned} \dot{V}_1 &= \varepsilon_d \dot{\varepsilon}_d + \varepsilon_\sigma \dot{\varepsilon}_\sigma \\ &= v_{px} \left( -\varepsilon_d \cos \varepsilon_\sigma + \varepsilon_\sigma \frac{\sin \varepsilon_\sigma}{d} \right) \\ &\quad + v_{py} \left( -\varepsilon_d \sin \varepsilon_\sigma - \varepsilon_\sigma \frac{\cos \varepsilon_\sigma}{d} \right) \\ &\quad + v_d \varepsilon_d \cos(\sigma_v - \sigma_d) + v_d \varepsilon_\sigma \frac{\sin(\sigma_v - \sigma_d)}{d} \end{aligned} \quad (13)$$

By substituting  $v_x, v_y$  and  $\omega$  into (13), we simplify the time derivative of  $V_1$  as:

$$\dot{V}_1 = -k_1 \varepsilon_d^2 - k_2 \varepsilon_\sigma^2. \quad (14)$$

Equation (14) ensures that  $\dot{V}_1 \leq 0$ , while  $k_1, k_2 \geq 0$ . And  $\dot{V}_1 = 0$ , if and only if  $[\varepsilon_d \ \varepsilon_\sigma]^T = [0 \ 0]^T$ . Thus, the control system is asymptotically stable with the designed control law.

## VI. EXPERIMENTAL VALIDATION

To verify the vision-based ground target tracking system, extensive flight experiments were performed in outdoor environments. The performance of the visual tracking scheme with the redetection method is firstly evaluated on several videos collected using the UAV platform. The IMM-EKF based state estimation of the maneuvering person is then experimentally verified and analyzed. Finally, experimental results of the visual tracking system in real-world are presented. The computation is fully implemented on the onboard TK1 computer in the experiments.

### A. Experimental Setup



Fig. 7. The experimental test-bed: DJI Matrice 100 UAV with onboard computer and monocular RGB gimbal camera system.

A DJI Matrice 100 UAV is used as an experimental test-bed, as shown in Fig. 7. Onboard equipments include a DJI Manifold embedded Linux computer (NVIDIA Tegra

TK1 processor consisting of a Kepler GPU with 192 CUDA cores, a NVIDIA 4-Plus-1 quad-core A15 CPU of 1.5 GHz, 2 GB RAM and a WiFi interface), a monocular Zenmuse X3 gimbal camera, a GPS receiver, an Inertial Measurement Unit (IMU), a barometer, and one downward-pointing sensor module of a DJI Guidance visual sensing system.

### B. Visual Tracking Test

The performance of the visual tracking scheme is evaluated on several videos collected using the UAV. Three cases are considered, i.e., no occlusion, partial occlusion and full occlusion. Full occlusion for a long term occurs in all videos, and the target losses for at least 3s (about 90 frames). The traditional KCF fails to track the targets correctly in the case of full occlusion. The visual tracking scheme consisting of the KCF tracker and the redetection method is evaluated. For videos of resolution, the KCF tracker and the redetection method achieve 30FPS and 10FPS, respectively. The accuracy of the visual tracking scheme is measured by  $P_{\text{accuracy}} = \frac{N_{\text{td}}}{N_t}$ . The experimental results of the visual tracking scheme are summarized in Table I. Table I indicates that the proposed visual tracking scheme can effectively track and redetect the target. The accuracy decreases when the background is complex or more variations are introduced.

Test No.	$N_s$	$N_t$	$N_{\text{td}}$	Situation	Accuracy
1	660	332	312	SB	94.87%
2	930	872	838	SB	96.10%
3	1169	1019	964	SB	94.60%
4	1349	1222	1033	CB	84.53%
5	2212	1881	1609	CB	86.07%
6	840	601	459	IV,CB	76.37%
7	1049	863	712	IV,CB	82.50%

TABLE I: Visual tracking results while the target with occlusion in different scenarios.  $N_s$  is the total number of frames in each test,  $N_t$  is the number of frames containing the target,  $N_{\text{td}}$  is the number of frames where the target is correctly tracked and redetected. SB: Simple Background; CB: Complex Background; IV: Illumination Variation.

### C. Evaluation of the Target State Estimation Test

Experiments were conducted to evaluate the accuracy of the target state estimation, where the measured values are used as the ground truth. The relative error is calculated by  $\delta_d = \frac{\Delta d}{d_{\text{true}}} = \frac{d_{\text{est}} - d_{\text{true}}}{d_{\text{true}}}$ , where  $d_{\text{est}}$  is the relative position estimated by the IMM-EKF based on estimator, and  $d_{\text{true}}$  is the measured relative position. To examine the effects of the pitch and the yaw angles of the gimbal on the state estimation precision, the relative position between the target and the UAV is fixed. Fig. 8 shows the relative distance estimation errors  $\delta_d$  with respect to the pitch and the yaw angles of the gimbal, where the greatest/smallest errors occur in the red/blue area.

Fig. 8 indicates that the accuracy of the target state estimation decreases along with the increasing of the pitch angles. The following reasons may lead to the decreasing estimation accuracy: 1) lens distortion; 2) deviation of calculating actual distance in image edge; 3) inherent noise of the gimbal camera. In general, the results show that the

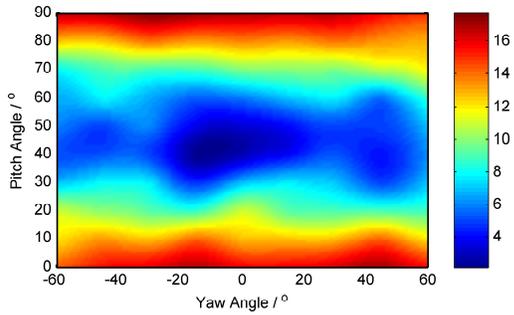


Fig. 8. The pseudocolor plot of relative distance estimation errors (%). The test is under a certain range of the gimbal's angle:  $-60^\circ$  to  $60^\circ$  of yaw angle and  $20^\circ$  to  $70^\circ$  of pitch angle. The distribution of the error can be seen from the legend on the right side of the figure

target state estimation achieves acceptable performance in a certain range between 3% and 8% of the gimbals attitude angles. In addition, the yaw angle between  $-20^\circ$  and  $20^\circ$  and pitch angle between  $30^\circ$  and  $55^\circ$  achieves smallest estimated errors. In summary, the yaw and the pitch angle of gimbal should be controlled within a certain range to provide acceptable localization accuracy.

#### D. Target Following Control Simulations

In this subsection, the simulation is presented to compare the performance between the proposed nonlinear control law and the PID controller.

In the simulations, the UAV is represented by a rigid body-mass point and the parameters of the UAV were set as the following: the angular velocity of the yaw angle is  $90 \text{ deg/s}$ , and the max velocity of the UAV is  $6 \text{ m/s}$ . The trajectory of target is generated by a set of random way-points with varying velocities and accelerations. The switchable tracking strategy is considered both in proposed nonlinear controller and the PID control law. The simulation results of the target tracking control are shown in Fig.9.

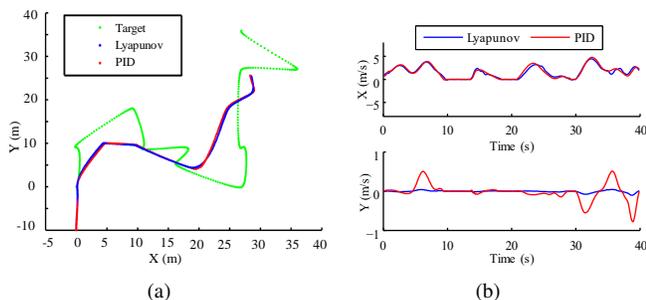


Fig. 9. The UAV tracks a maneuvering target with random movements. (a)trajectories; (b)velocities.

Fig.9(a) shows that the switchable tracking strategy can steadily track the target while moves smoothly. It is noted that the proposed nonlinear controller achieves similar performance compared to the PID controller under the same switchable tracking strategy. Fig. 9(b) illustrates that the computed velocity of the nonlinear controller is smoother

than the PID controller, that implies unnecessary movements of the UAV can be reduced.

One of critical concerns of the control system is the adjustment of the controller's parameters. There are nine parameters in the PID algorithm, and it is quite difficult to adjust the parameters to achieve satisfactory control performance. In contrast, there are only two parameters to be adjusted in the proposed nonlinear controller, the two parameters can be set in a relatively wide range. The nonlinear controller is hence more conveniently implemented than the PID controller in real-world applications.

#### E. Target Tracking Flight Test

To systematically investigate the performance of the proposed vision-based tracking scenario, extensive flight experiments were performed for autonomous target tracking using fully onboard computation. An un-predefined person is selected manually as the target. The UAV tracks the maneuvering person in unstructured outdoor environments. The flying altitude of the UAV is set to a height of 5 meters. The trajectory of the target person is not pre-set.

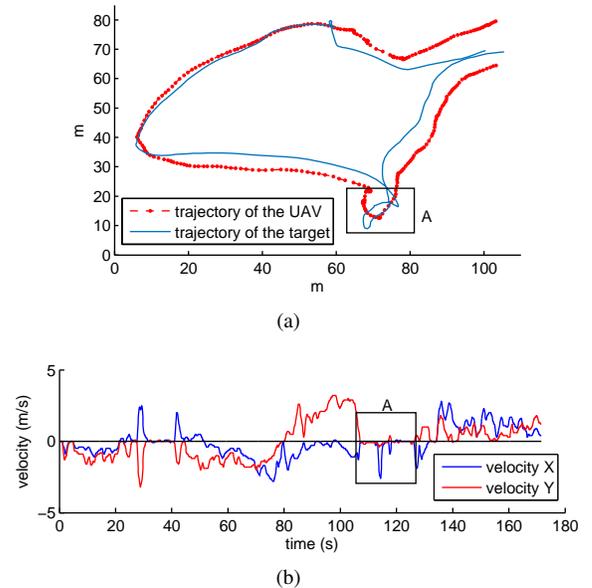


Fig. 10. (a) Trajectories of the UAV(red) and the target(blue). Their position information is derived from the GNSS; (b)UAV's of the x direction and y direction velocities in the body coordinates

The flight results are shown in Fig. 10, where the total distance of the tracking process is 330 meters with about 3 minutes. The trajectories of the target person and the UAV are obtained from the mobile phone's GPS data and the UAV's onboard GPS data, respectively. Fig. 10(a) shows that the UAV can maintain a certain distance to the target around 5 meters, and can successfully track the maneuvering target person.

It is noted from Fig. 11(a) that the trajectory of the UAV is more stable than that of the ground target. It takes the advantages of the proposed switchable target tracking strategy, where the observation mode works when the target person frequently changes its direction. In this case, the

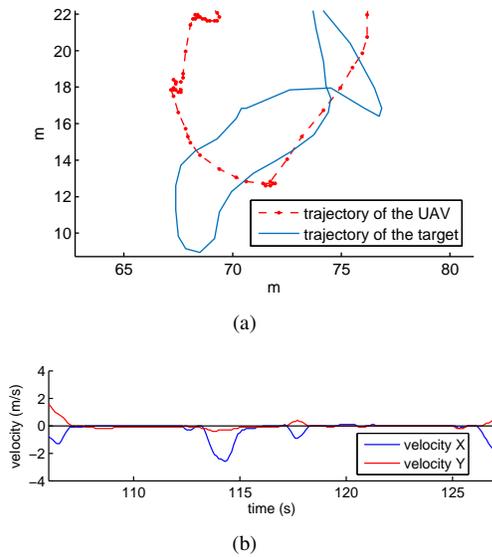


Fig. 11. (a) Partial enlarged views of box A in Fig. 10(a), showing the experimental results of the switchable tracking strategy of UAVs when the target frequently changes its direction. (b) Partial enlarged views of box A in Fig. 10(b)

UAV rotates to track the target person without horizontal displacement, which can be shown from the Fig. 11(b). The experimental results illustrate the UAV can stably track the maneuvering target person using the proposed vision-based tracking scenario.

A demo video of our target tracking UAV system in outdoor environments can be seen in <https://youtu.be/7dc4etU0IHs>, working entirely onboard without prior knowledge of the target.

## VII. CONCLUSIONS

In this paper, the design and implementation of an autonomous vision-based ground target tracking system for rotorcraft UAV are presented. A visual tracking scheme integrating the KCF tracker and the redetection method was designed, which is robust to target occlusion and loss. The IMM-EKF based estimator is presented to estimate the states of the maneuvering target. The gimbal and the UAV are controlled simultaneously to perform stable tracking. The gimbal camera is controlled to keep the target in the center of the frame. Moreover, a switchable tracking strategy including the following mode and the observing mode is proposed for UAVs, and a Lyapunov theory-based nonlinear controller was designed. Extensive real-time flight experiments have been executed in outdoor environments, where the computation is fully implemented on the onboard computer. Experimental results illustrate that the proposed real-time vision-based tracking system achieves stable and robust tracking performance. Future research will consider autonomous vision-based obstacle avoidance and object grasping of UAVs.

## REFERENCES

- [1] K. P. Valavanis and G. J. Vachtsevanos, *Handbook of unmanned aerial vehicles*. Springer Publishing Company, Incorporated, 2014.
- [2] R. He, A. Bachrach, M. Achtelik, A. Geramifard, D. Gurdan, S. Prentice, J. Stumpf, and N. Roy, "On the design and use of a micro air vehicle to track and avoid adversaries," *The International Journal of Robotics Research*, 2009.
- [3] G. Bevacqua, J. Cacace, A. Finzi, and V. Lippiello, "Mixed-initiative planning and execution for multiple drones in search and rescue missions," in *Twenty-Fifth International Conference on Automated Planning and Scheduling*, 2015.
- [4] K. Kanistras, G. Martins, M. J. Rutherford, and K. P. Valavanis, "Survey of unmanned aerial vehicles (uavs) for traffic monitoring," in *Handbook of Unmanned Aerial Vehicles*. Springer, 2015, pp. 2643–2666.
- [5] A. W. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1442–1468, 2014.
- [6] R. Mahony, V. Kumar, and P. Corke, "Multirotor aerial vehicles," *IEEE Robotics and Automation magazine*, vol. 20, no. 32, 2012.
- [7] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015.
- [8] H. Deng, X. Zhao, and Z. Hou, "Tracking ground targets using an autonomous helicopter with a vision system," in *Image and Signal Processing (CISP), 2010 3rd International Congress on*, 2010, pp. 1704 – 1708.
- [9] A. Razinkova and H.-C. Cho, "Tracking a moving ground object using quadcopter uav in a presence of noise," in *Advanced Intelligent Mechatronics (AIM), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1546–1551.
- [10] F. Lin, X. Dong, B. M. Chen, K.-Y. Lum, and T. H. Lee, "A robust real-time embedded vision system on an unmanned rotorcraft for ground target following," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 2, pp. 1038–1049, 2012.
- [11] C. Teuliere, L. Eck, and E. Marchand, "Chasing a moving target from a flying uav," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*. IEEE, 2011, pp. 4929–4934.
- [12] A. G. Kendall, N. N. Salvapantula, and K. A. Stol, "On-board object tracking control of a quadcopter with monocular vision," in *Unmanned Aircraft Systems (ICUAS), 2014 International Conference on*. IEEE, 2014, pp. 404–411.
- [13] A. Chakrabarty, R. Morris, X. Bouyssounouse, and R. Hunt, "Autonomous indoor object tracking with the parrot ar. drone," in *Unmanned Aircraft Systems (ICUAS), 2016 International Conference on*. IEEE, 2016, pp. 25–30.
- [14] J. Pestana, J. L. Sanchez-Lopez, S. Saripalli, and P. Campoy, "Computer vision based general object following for gps-denied multirotor unmanned vehicles," in *American Control Conference (ACC), 2014*. IEEE, 2014, pp. 1886–1891.
- [15] M. A. Olivares-Méndez, P. Campoy, C. Martínez, and I. Mondragón, "A pan-tilt camera fuzzy vision controller on an unmanned aerial vehicle," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*. IEEE, 2009, pp. 2879–2884.
- [16] K. Krishnamoorthy, D. Casbeer, and M. Pachter, "Minimum time uav pursuit of a moving ground target using partial information," in *Unmanned Aircraft Systems (ICUAS), 2015 International Conference on*. IEEE, 2015, pp. 204–208.
- [17] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, *Exploiting the circulant structure of tracking-by-detection with kernels*, 2012.
- [18] R. M. Gray *et al.*, "Toeplitz and circulant matrices: A review," *Foundations and Trends® in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.
- [19] Y. Wu, J. Lim, and M.-H. Yang, "Object tracking benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1834–1848, 2015.
- [20] C. Zhan, X. Duan, S. Xu, Z. Song, and M. Luo, "An improved moving object detection algorithm based on frame difference and edge detection," in *Image and Graphics, 2007. ICIG 2007. Fourth International Conference on*. IEEE, 2007, pp. 519–523.
- [21] L. A. Johnston and V. Krishnamurthy, "An improvement to the interacting multiple model (imm) algorithm," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 2909–2923, 2001.
- [22] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. part i. dynamic models," *IEEE Transactions on Aerospace Electronic Systems*, vol. 39, no. 4, pp. 1333–1364, 2004.