This ICCV Workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Vision-based Online Localization and Trajectory Smoothing for Fixed-wing UAV Tracking a Moving Target

Yong Zhou[1], Dengqing Tang[1], Han Zhou[1], Xiaojia Xiang[1] ,Tianjiang Hu[2]*
[1] National University of Defense Technology, [2] Sun Yat-sen University
{zhouyong13, tangdengqing09, zhouhan, xiangxiaojia, t.j.hu}@nudt.edu.cn

Abstract

Using unmanned aerial vehicle (UAV) to estimate and predict the position and motion of the ground target has been widely focused on in many computer vision tasks. This work aims to address the development of a vision-based ground target localization and estimation for a fixed-wing UAV. Limited by the lightweight onboard processor, it is conflicting with the need for online onboard operation and the computing resource limitation of the platform. In this paper, we develop a practical approach to recover dynamic targets based on extended Kalman filter (EKF) for localization and locally weighted regression for trajectory smoothing. Our methods run online in real time with the only data up to the current timestep. The flight experiment results show the effective tracking and localization of the ground moving target.

1. Introduction

Unmanned aerial vehicles (UAVs) are playing increasing prominent roles in civil and military fields. Equipped with color cameras and other sensors, UAVs are widely used in disaster search and rescue [8], aerial surveillance and mapping [9], wildlife surveillance and protection [7][15] and so on. Among these tasks, moving targets tracking and localization is the foundation of their performance of important tasks [13], which still leaves lots of issues to solve.

Tracking methods vary depending on whether the mobile target is a cooperative one. Because of the uncertainty of target motion, coupled with the high mobility of fixed-wing drone, it is extremely challenging for a fixed-wing UAV to track a ground non-cooperative mobile target in outdoor environment. Most of the previous studies either only studied one aspect of the tracking problem [10][18], or only conducted pure theoretical research [34][22]. In this paper, we focus on the entire process of non-cooperative tracking, from target detection to motion estimation, and carried out the flight experiment verification.

Another challenge, for the non-cooperative tracking, is the online operation with limited computing resource [28]. Due to the restrictions of the payload capability and power consumption of the UAV, the embedded onboard processor has limited resources. If we want to implement onboard processing, algorithm optimization is a necessity. In addition, to achieve robust tracking of the target by the drone, it is often necessary to accurately localize the target and estimate the motion. Therefore, in this work, we focus on the fixed-wing UAV's online tracking and processing for ground mobile targets. The main contributions are summarized as:

(1) We proposed a framework of fixed-wing tracking ground mobile targets, including targets detection, localization and trajectory smoothing.

(2) To complete online processing, the method of EKF-based localization and the locally weighted regression-based [6] trajectory smoothing are proposed and implemented on a lightweight processor.

The remainder of the paper is organized as follows. Section 2 introduces relevant work on target localization and estimation. In Section 3, system overview and frames definition are illustrated briefly. Section 4 describes the proposed methods of EKF localization, outlier detection and online smoothing. The flight experiments results are discussed in Section 5. Finally, the paper ends with some conclusion remarks and future works.

2. Related works

Object detection has witnessed tremendous progress with the development of deep learning. Many CNN-based state-of-the-art object detection methods have been proposed, such as two-stage methods for R-CNN [12], Fast R-CNN [11] and Faster R-CNN [27], and one-stage methods for YOLO [24] and SSD [20]. However, object detection in aerial image is still a challenging task [33] due to the cluttered environment, changing scales and limited resources. Redmon et al. proposed an improved YOLO v3 [26] method based on YOLO [24] and YOLO v2 [25] for predicting across scales, but they must run on GPU. Kaaniche et al. [17] employed feature extraction methods with the Canny-Edge Detector and Harris Detector from aerial images, but it cant meet the real-time requirements with lightweight onboard processors.

Numerous studies have focused on target localization with an aerial vehicle. However, these works mainly concentrate on stationary targets. Zhang et al. [30] designed a stereo vision technique with multiple view to localize singe stationary target using a quadrotor drone. Barber et al. [2] presented four techniques, in terms of recursive least squares (RLS) filtering, bias estimation, flight path selection and wind estimation, to reduce ground stationary localization error. There are also some studies on the motion targets localization. Wang et al. [28] proposed a tracking and localization framework for a single fixed-wing UAV to track a ground target. Minaeian et al. [21], achieved motion target detection and localization onboard, but they need unmanned ground vehicles (UGVs) as landmarks to aid in localization. Unlike the studies mentioned above, our work will focus on onboard online localization with only a UAV and a monocular camera.

The position of the target is discrete for the vision-based localization. However, the movement of the target is smooth due to the constraints of kinematic. Canonical smoothers, such as square-root information smoother [4], Rauch-Tung-Striebel (RTS) smoother [23], provide a good solution for smoothing. However, they make use of all the data to estimate every state, which is not causal and cant be used online [3]. Therefore, those methods are not applicable in our case. In the case of online smoothing, Anderson et al. [1] proposed a real-time trajectory generation algorithm for UAV flying. They use a sequence of successive waypoint path segments with kinematic constraints to generation trajectories. Wang et al. [29] presented a trajectory smoothing algorithm based on Dubins-Helix method, which is used to generate an effective trajectory and achieve real-time computation.

3. System Overview

Our focus in this paper is online localization and simultaneously predicting the target motion trajectory where the motion of the target is completely unknown beforehand. As is shown in Figure 1, a fixed-wing UAV equipped with IMU, GPS and a pan-tilt gimbal where the color camera is fixedly mounted, tracks a ground mobile target. The onboard processor keeps the camera pointing to the ground target by controlling the gimbal, captures the images and then detects the bounding box of ground target. With a simple pinhole camera model, the target in camera coordinate frame $\{C\}$ is obtained with the measurement:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \frac{f}{z_c} \begin{bmatrix} x_c \\ y_c \end{bmatrix} \tag{1}$$

where $Z_k = [u\ v]^{\mathrm{T}}$are the center of the target bounding box and f is the focal length of the camera. $P_C = [x_c\ y_c\ z_c\ 1]^{\mathrm{T}}$ denote the relative position of $\{C\}$ with respect to the target. The transform relation between UAV body frame $\{B\}$, gimbal coordinate frame $\{G\}$ and camera coordinate frame $\{C\}$, which is generally called extrinsic, is calibrated offline with iteration and optimization method inspired by the hand-eye calibration idea [32]. Finally, the position of target in world coordinate frame $\{W\}$, namely North-East-Down (NED) coordinate frame in this paper, can be obtained with the GPS and IMU. The transformation from$\{C\}$ to $\{G\}$, $\{G\}$ to $\{B\}$ and $\{B\}$ to $\{W\}$ are denoted as T_C^G, T_G^B and T_B^W respectively. Then, the position of target in $\{W\}$ is:

$$p = T_B^W T_G^B T_C^G P_C \tag{2}$$

where T, including 3x3 rotation matrices $R_{3\times3}$ and 3x1 translation vector $t_{3\times1}$, is 4x4 matrices of the form:

$$T = \begin{bmatrix} R_{3x3} & t_{3x1} \\ 0_{1x3} & 1 \end{bmatrix} \tag{3}$$

Eq. (2) gives the targets world position. However, the uncertainty of target motion brings more challenges to detection and localization; sensor noise and target false detection can cause great localization deviation. To solve this problem, an EKF-based method is used to localize the mobile target. Moreover, we employ locally weighted regression to smooth target trajectory and predict the mobile targets motion in real time.

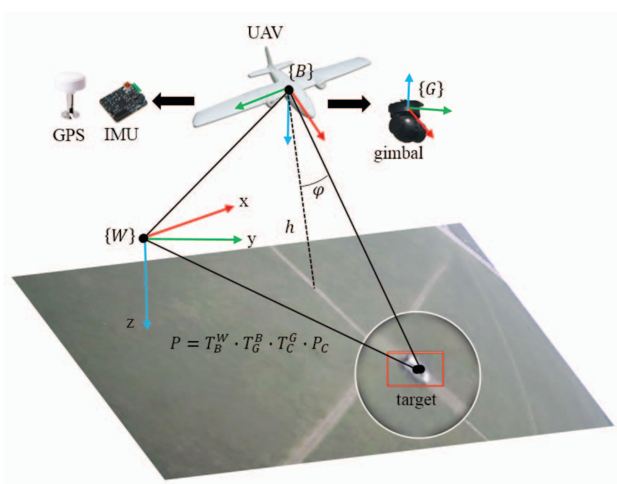


Figure 1. Equipped with GPS, IMU, gimbaled camera and embedded processor, UAV tracks and localizes ground mobile target.

4. Mobile Target Localization and Estimation

4.1. EKF Target Localization

The target localization in Eq. (2) provides a one-shot estimate of the target location. Unfortunately, this equation is highly sensitive to measurement errors. For example, rapid movement of the fixed-wing UAV may cause the target to be lost in the image field of view. Considering long-term tracking and localization, as well as target motion and drone observation are statistically independent, it is valid to assume the problem as a Gaussian process. Based on above assumption, we present an EKF-based method to solve the localization problem in this section. Rearranging Eq. (2), assume the target moves at a flat ground and the altitude of the UAV h can be measured with a GPS module. Then, the relationship between the position of target in $\{W\}$ $P_C = [x_c\ y_c\ z_c\ 1]^{\mathrm{T}}$ and pixel coordinate of target $[u\ v]^{\mathrm{T}}$, noted as h_k, can be rewritten as:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{f}{z_c} T_G^C T_B^G T_W^B \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} \tag{4}$$

where $z_c = h/\cos\varphi$ and φ is the angle between the cameras optimal axis and vertical direction as shown in Figure 1, which can be measured from the gimbal drive and IMU. Then, the state equation of target localization can be:

$$x_k = Ax_{k-1} + w_{k-1} \tag{5}$$

where $x_k = (x_W, y_w, v_{wx}, v_{wy})^T$ is the target state variable at the time k; v_{wx},v_{wy} are the targets positions and speed on the X and Y axis; w_{k-1} is a white Gaussian noise sequence state with the covariance of Q_k. The state transition matrix will be:

$$A = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6}$$

where Δt is sampling time. And the observation model is:

$$Z_k = Hx_k + v_k \tag{7}$$

where $Z_k = [u\ v]$ is the system measurement, $v_k$is errors and H is system observation matrix that can be described with the Jacobian matrix of Eq. (4):

$$H = \frac{\partial h_k(u, v)}{\partial x_k} \tag{8}$$

Finally, if the Kalman gain is noted as K_k, the state of the target x_k can be updated:

$$\hat{x}_k = \hat{x}_{\overline{k}} + K_k \left(z_k - H\hat{x}_{\overline{k}}\right) \tag{9}$$

4.2. Outlier Detection

When fixed-wing UAV tracks ground mobile targets at high speed, its attitude and direction change drastically, causing the target out-of-view and re-appearance frequently. As a result, localization results may contain some outliers, which have a large impact for the tracking tasks. In general, outlier detection techniques are offline methods such as clustering-based methods and statistical methods [5]. These supervised learning and unsupervised learning methods require much data and computing resource, which may not be applicable for online tasks. Therefore, an outlier detection algorithm based on target motion state is proposed to get the wrong localization results.

The main idea of this algorithm is to estimate the current position through the kinematics model of the target. We employ the state model as described in Eq. (5), and then estimate current target position with the results of the last time step. The standard deviation of estimated localization results is determined experimentally. Since the localization errors is Gaussian distribution, we can regard the values that are not in the 3-sigma range as outliers. The detailed procedure is summarized as follows:

Algorithm 1 outlier detection for target localization

Input:
 target motion state:
 observation at time k, x_k
 localization results at time $k-1$, $x_{\bar{k}-1}$
 target motion state matrix, A
 standard deviation of localization solution, S

Output: outlier results of target localization

Procedure:
 (1) establish motion state model
  i. establish a target kinematic state space model and get the state matrix, A;
  ii. calculate the standard deviation S, which is obtained from experiment;
 (2) target position prediction
  i. calculate the localization values of prediction, $\hat{x}_{\overline{k}} = A\hat{x}_{k-1}$;
  ii. calculate the target position confidence interval, $x_{\min} = \left(x_k - \hat{x}_{\overline{k}}\right) - 3*S$, $x_{\max} = \left(x_k - \hat{x}_{\overline{k}}\right) + 3*S$;
 (3) outlier detection
  If $x_k \in [x_{\min} \quad y_{\max}]$, x_k is regarded as a normal value; otherwise, it is regarded as an outlier.

4.3. Online Smoothing

In general, when the UAV tracks the ground target, the object detection takes a certain amount of time, so the localization results are a series of discrete points. However, due to kinematic constraints, the motion of the target is continuous. Therefore, this part proposes an online target trajectory smoother based on locally weighted regression.

Suppose $X = [x_1, x_2, \ldots, x_n]$ is n points ordered in time series and x_n is current time point; w_i is weights; θ is coefficient. Define loss function:

$$J(\theta) = \frac{1}{n}\sum_{i=1}^{n} \omega_i \left(y_i - \theta^* x_i\right)^2 = \frac{1}{n}\omega \left\|Y - X^*\theta\right\|^2 \quad (10)$$

Namely,

$$J(\theta) = \frac{1}{n}\left(Y^T\omega Y - \theta^T X^T \omega Y - Y^T \omega X\theta + \theta^T X^T \omega X\theta\right) \quad (11)$$

Derive Eq. (11) with respect to and let it equals 0, then we can get:

$$\theta = \left(X^T\omega X\right)^{-1} X^T \omega Y \quad (12)$$

Because of $X * \theta = X\left(X^T\omega X\right)^{-1} X^T\omega Y = \hat{Y} = \widehat{H}Y$, thus

$$\hat{H} = X\left(X^T\omega X\right)^{-1} X^T\omega \quad (13)$$

Then, current trajectory prediction is given by weight function $W(x)$ and the regression model $\hat{Y}$:

$$W(x) = \left(1 - \left|\frac{x}{d\max}\right|^3\right)^{3/2} \quad (14)$$

$$\hat{Y} = X\left(X^T\omega X\right)^{-1} X^T\omega Y \quad (15)$$

where $d_{\max} = |x_n - x_1|$

Eq. (13) gives the estimation value of current trajectory points. Unfortunately, considering the real-time requirements of online smoothing, it is necessary to reduce the computational complexity.

$$J(\theta) = \frac{1}{n}\left\|\tilde{Y} - \tilde{X}^*\theta\right\|^2 \quad (16)$$

where $\tilde{X} = \sqrt{\omega}X; \tilde{Y} = \sqrt{\omega}Y$. Then Eq. (13) can be rewritten as:

$$\hat{H} = X\left(X^\tau\omega X\right)^{-1} X^\tau\omega = \sqrt{\omega}^{-1}\tilde{X}\left(\tilde{X}^T\tilde{X}\right)^{-1}\tilde{X}^\tau\sqrt{\omega} \quad (17)$$

Then, QR decomposition for X , namely, X P=QR, where Q is an orthogonal matrix, R is an upper triangular matrix, and P is a column permutation matrix. Then, we have:

$$\begin{aligned}\tilde{X}\left(\tilde{X}^T\tilde{X}\right)^{-1}\tilde{X}^T &= QRP^T\left(PR^TQ^TQRP^T\right)^{-1}PR^TQ^T \\ &= QRP^T\left(P^T\right)^{-1}R^{-1}\left(R^T\right)^{-1}P^{-1}PR^TQ^T \\ &= QRR^{-1}\left(R^T\right)^{-1}R^TQ^T \\ &= QQ^T\end{aligned} \quad (18)$$

Finally, Eq. (13) can be write as:

$$\hat{H} = \sqrt{\omega}^{-1}\tilde{X}\left(\tilde{X}^\tau\tilde{X}\right)^{-1}\tilde{X}^\tau\sqrt{\omega} = \sqrt{\omega}^{-1}QQ^\tau\sqrt{\omega} \quad (19)$$

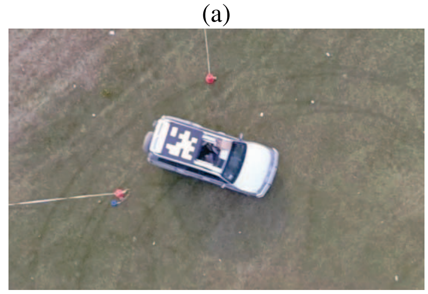
Obviously, Eq. (13) is equivalent to Eq. (19). If we solve Eq. (13), its condition number can be computed with $\kappa\left(X^TX\right) = \kappa\left(V\Sigma^TU^TU\Sigma V^H\right) = \kappa\left(V\Sigma^2V^H\right) = \kappa(X)^2$. That is, the condition number of $(X^TX$ is the square of X. While Eq. (19) obtains the orthogonal matrix Q by QR decomposition. Since the condition number of the orthogonal matrix is 1, it will not amplify the condition number.

5. Flight experiment

5.1. Implementation Details

We have conducted a series of field experiments to evaluate our proposed methods. The fixed-wing UAV and ground target, as shown in Figure 2, are used as our experiment platforms. The main characteristics of the experimental UAV are listed in Table 1. The aerial platform is equipped with a gimbaled camera that make it always point to the ground target and capture the images with 30 HZ and 1280x720 frame size. The autopilot of the drone is PIXHAWK which fuses the IMU and GPS data. An ODROID-XU4 with eight CPU cores was selected as the onboard processor. It is run in the ROS system and all our algorithms are running on the system. The ground target platform is a Pajero, which carries a DGPS to record its position as ground truth.

(a)



(b)

Figure 2. Aerial and ground target platform used in experiments: (a) Fixed-wing UAV; (b) ground mobile target.

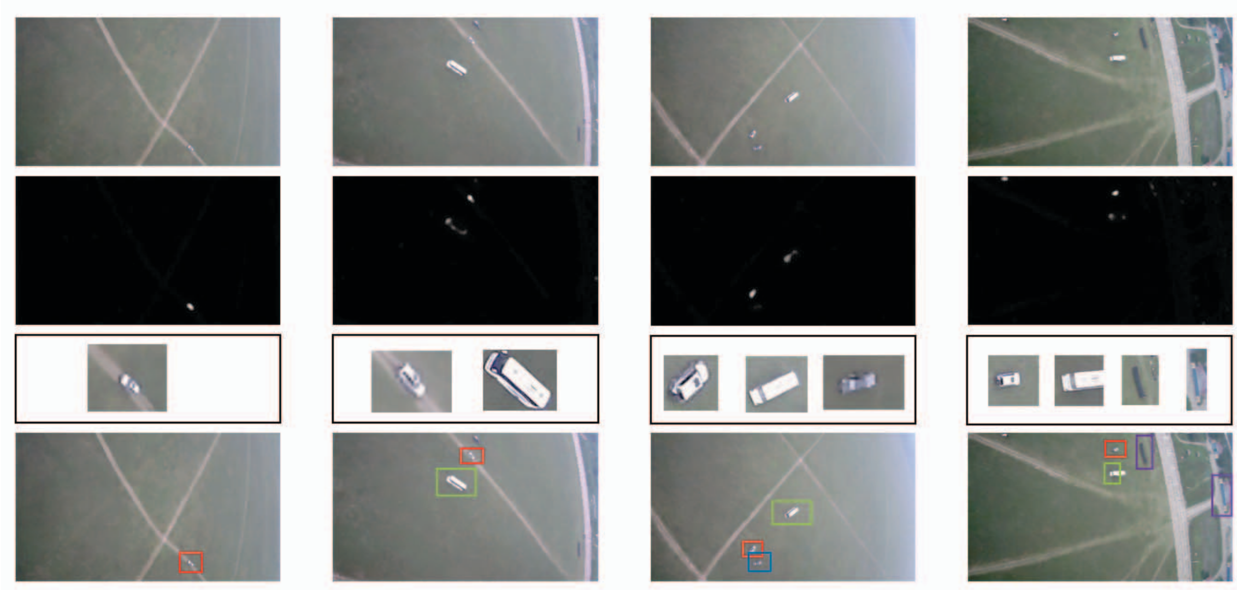


Figure 3. Onboard ground target detection results. From top to bottom: origin images, saliency map, proposal regions, and classification results. The classification results of Pajero, bus, pickup trucks and others are indicated by red, green, blue and purple.

The fixed-wing UAV tracks the ground target with the manner of loitering. The flying height is about 100m. Datasets captured by the onboard gimbaled camera includes different challenging scenarios, such as multiple objects, background interference and illumination. The detection results and estimated trajectory are shown in subsection 5.2 and 5.3 respectively.

Parameter	Value	Unit
Wingspan	1.8	m
Maximum weight	4.2	kg
Cruising speed	16	m/s
Flight speed	14-23	m/s

Table 1. Specifications of the experimental UAV prototype.

5.2. Onboard Ground Object Detection

Object detection on the fixed-wing UAV platform is still a challenging task due to large scales and limited sources. Feature-based methods provide a good solution for object detection, but they have poor robustness in cluttered environment. Deep learning based object detection methods show greater advantages, especially in terms of robustness and accuracy. The disadvantage is that they require more computing resources. In consideration of the real-time detection, localization and online smoothing, as well as a limited computing resource of onboard processor detections, we use a two-stage method [31]: the first step is saliency detection for region of interest (ROI) and the second is classification with a shallow convolutional neural network (CNN). Figure 4 shows the run time of classic saliency detection algorithms, such as SR [16], PQFT [14] and HFT [19]. In order to reduce the time of image processing, we use SR to obtain saliency map. The results of SR detection and CNN-based classification are shown in Figure 3. We use the Fourier transform to obtain a saliency map from the frequency domain. The saliency maps will provide the targets proposal regions. Once we get the proposal regions, a CNN-based classification network will be used to classify different categories of targets. For roads, houses, flocks, etc. in the environment, we classify them as other categories.

To select suitable resolution for image processing, origin images of 1280x720 resolution are downsampled 4x, 9x, 16x and 25x. The measured run times on ODROID-XU4 at different resolutions is shown in Table 2. Both of the consumed time and number of saliency region increase with higher resolution. But low resolution may cause missing detection. In order to reduce the run time of detection, achieving online localization, as well as in consideration of the accuracy, we chose the resolution of 320x180.

5.3. Mobile Target Online Localization and Smoothing

Given the estimated target detection bounding box, the target tracking results could be projected to the world frame W with perspective projection and rigid-body transform. And further, we use EKF to improve the robustness of the localization results. In our experiment, a Pajero (see Figure 2 (b)) is adopted as the tracking target, which is controlled

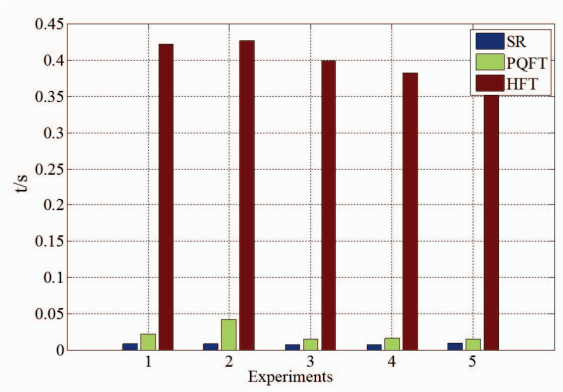


Figure 4. Run time of saliency region detection.

Resolution	SR/ms	SR+CNN/ms	Number of objects
1280x720	778.44	978.62	17.83
640x360	178.92	266.04	7.81
426x240	111.42	169.54	4.71
320x180	44.44	98.28	4.36
256x144	28.30	82.01	4.55

Table 2. Run time of objection detection.

on the ground with random motion. Figure 5 illustrates the EKF localization error of X axis (red), Y axis (green) and X-Y planar (blue) respectively. At starting point, the fixed-wing UAV searches for the ground target in the air, leading to a large localization errors. When the ground target is captured by the aerial platform, a stable localization result is obtained. Sometimes the localization results are close to 2 m distance from the actual position. Unfortunately, long time ground target tracking can result in loss of some information during transmission. At this time, the localization error may increase. But as long as the UAV recaptures the target, the localization result will tend to be stable. In the whole process of initial search, stable tracking, information loss, search and re-tracking, the average absolute error is 11.48 m in X axis, 11.63 m in Y axis and 18.03 m in X-Y planar (see Table 3).

In order to better optimize the localization results and position prediction, we use an online smoother as shown in Eq. (19). We only employ data up to the current timestep to estimate the state of ground target. For the initial value of the smoother, we use its EKF localization results directly. Then the smoother smooths the trajectory using span as the number of points used to compute current state estimation. Table 3 shows the average absolute error, memory size and processing time with different value of span. As the smoothing data used increases, the optimization results are significantly improved. Correspondingly, memory size and processing time increase as well. Although growing number of data can reduce the localization errors, it may lengthen initialization and processing time. The memory usage also increases. Therefore, the selection of span from 91 to 121 is a suitable choice.

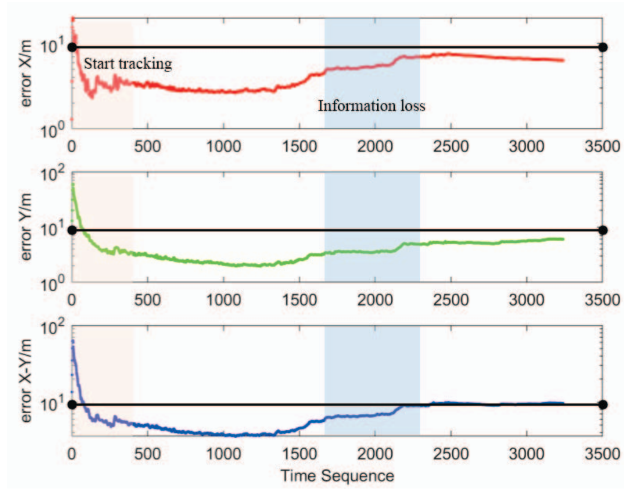


Figure 5. Average localization errors of long-term tracking. From top to bottom: X axis (red), Y axis (green) and X-Y planar (blue) localization errors.

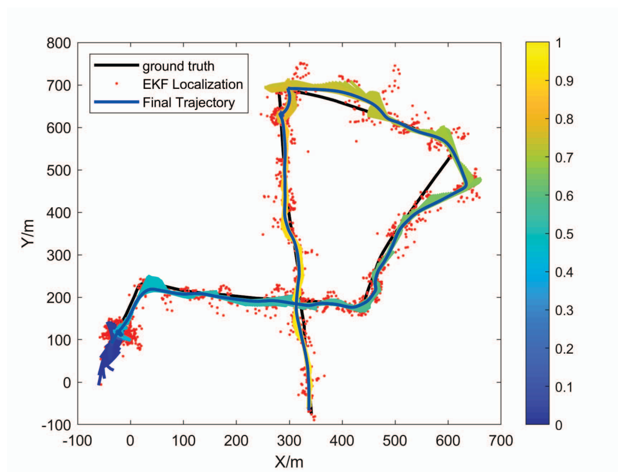


Figure 6. Ground target trajectory smoothing and optimization, including the actual trajectory, EKF localization results and smoothing results for every timestep.

The trajectories of ground target, including the actual trajectory, the estimated EKF localization measurements and the optimized smoothing trajectory, are shown in Figure 6. In order to make a quantitative analysis of the experiment results, all of the data are transformed into meter-scaled data with the UAV takeoff point as the origin, X pointing to the north and Y pointing to the east. In Figure 6, the gradient color is the result of each time step of online smoothing and the final optimized trajectory is indicated by blue. The smoothing trajectory is somewhat different from the actual trajectory, especially the upper right has a remarkable deviation. Three possible reasons are summarized as follows:

(1) Occasional false detection resulting in EKF localization and trajectory smoothing errors.

(2) Information loss: the upper rights deviation is the reason of the loss of data transmission.

(3) Delay of the obtained data used for online smoothing. Sensors such as onboard IMU, GPS and camera are not synchronized.

	X/m	Y/m	X-Y/m	Memory /KB	Time /ms
EKF	11.48	11.63	18.03	-	-
span=5	11.55	11.72	18.16	0.27	2.0
span=21	12.19	12.52	19.28	0.54	2.8
span=41	10.02	10.19	15.66	0.87	3.8
span=71	8.62	8.26	13.00	1.30	5.2
span=91	8.36	8.01	12.69	1.59	6.1
span=121	8.15	7.94	12.644	2.03	7.5
span=151	8.19	8.30	13.03	2.48	8.9
span=201	8.43	9.21	14.10	3.21	11.1

Table 3. Testing results of average absolute error, memory size and processing time of trajectory smoothing.

6. Concluding remarks

In this paper, we proposed a framework for a fixed-wing UAV to track a single mobile ground target. With the vision-based method, a lightweight target detection algorithm and its experimental results are presented. The detection results are used to localize the mobile targets in real time with a developed EKF-based localization method. To estimate the motion and position prediction, we provided an online trajectory smoothing algorithm. The results show good performance of tracking and estimation. In future work, we will focus on the scenarios and cooperation algorithm design multiple UAVs cooperative tracking on one or more moving targets.

References

[1] Erik P Anderson, Randal W Beard, and Timothy W McLain. Real-time dynamic trajectory smoothing for unmanned air vehicles. *IEEE Transactions on Control Systems Technology*, 13(3):471–477, 2005.

[2] D Blake Barber, Joshua D Redding, Timothy W McLain, Randal W Beard, and Clark N Taylor. Vision-based target geo-location using a fixed-wing miniature air vehicle. *Journal of Intelligent and Robotic Systems*, 47(4):361–382, 2006.

[3] Timothy D Barfoot. *State Estimation for Robotics*. Cambridge University Press, 2017.

[4] Gerald J Bierman. Sequential square root filtering and smoothing of discrete linear systems. *Automatica*, 10(2):147–158, 1974.

[5] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.

[6] William S Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, 74(368):829–836, 1979.

[7] Oliver M Cliff, Robert Fitch, Salah Sukkarieh, Debra L Saunders, and Robert Heinsohn. Online localization of radio-tagged wildlife with an autonomous aerial robot system. In *Robotics: Science and Systems*, 2015.

[8] Milan Erdelj, Enrico Natalizio, Kaushik R Chowdhury, and Ian F Akyildiz. Help from the sky: Leveraging uavs for disaster management. *IEEE Pervasive Computing*, 16(1):24–32, 2017.

[9] Jurgen Everaerts et al. The use of unmanned aerial vehicles (uavs) for remote sensing and mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(2008):1187–1192, 2008.

[10] Eric W Frew, Dale A Lawrence, and Steve Morris. Coordinated standoff tracking of moving targets using lyapunov guidance vector fields. *Journal of guidance, control, and dynamics*, 31(2):290–306, 2008.

[11] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[12] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[13] Jose-Ernesto Gomez-Balderas, Gerardo Flores, LR Garcia Carrillo, and Rogelio Lozano. Tracking a ground moving target with a quadrotor using switching control. *Journal of Intelligent & Robotic Systems*, 70(1-4):65–78, 2013.

[14] Chenlei Guo, Qi Ma, and Liming Zhang. Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[15] Jarrod C Hodgson, Rowan Mott, Shane M Baylis, Trung T Pham, Simon Wotherspoon, Adam D Kilpatrick, Ramesh Raja Segaran, Ian Reid, Aleks Terauds, and Lian Pin Koh. Drones count wildlife more accurately and precisely than humans. *Methods in Ecology and Evolution*, 9(5):1160–1167, 2018.

[16] Xiaodi Hou and Liqing Zhang. Saliency detection: A spectral residual approach. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. Ieee, 2007.

[17] Khaled Kaaniche, Benjamin Champion, Claude Pégard, and Pascal Vasseur. A vision algorithm for dynamic detection of moving vehicles with a uav. In *proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 1878–1883. IEEE, 2005.

[18] Michail Kontitsis and Kimon Valavanis. A cost effective tracking system for small unmanned aerial systems. In

Selected papers from the 2nd International Symposium on UAVs, Reno, Nevada, USA June 8–10, 2009, pages 171–191. Springer, 2009.

[19] Jian Li, Martin D Levine, Xiangjing An, Xin Xu, and Hangen He. Visual saliency based on scale-space analysis in the frequency domain. *IEEE transactions on pattern analysis and machine intelligence*, 35(4):996–1010, 2012.

[20] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.

[21] Sara Minaeian, Jian Liu, and Young-Jun Son. Vision-based target detection and localization via a team of cooperative uav and ugvs. *IEEE Transactions on systems, man, and cybernetics: systems*, 46(7):1005–1016, 2015.

[22] Hyondong Oh, Seungkeun Kim, Hyo-Sang Shin, Brian A White, Antonios Tsourdos, and Camille Alain Rabbath. Rendezvous and standoff target tracking guidance using differential geometry. *Journal of Intelligent & Robotic Systems*, 69(1-4):389–405, 2013.

[23] Herbert E Rauch, CT Striebel, and F Tung. Maximum likelihood estimates of linear dynamic systems. *AIAA journal*, 3(8):1445–1450, 1965.

[24] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[25] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271, 2017.

[26] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[27] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[28] Xun Wang, Huayong Zhu, Daibing Zhang, Dianle Zhou, and Xiangke Wang. Vision-based detection and tracking of a mobile ground target using a fixed-wing uav. *International Journal of Advanced Robotic Systems*, 11(9):156, 2014.

[29] Yu Wang, Shuo Wang, Min Tan, Chao Zhou, and Qingping Wei. Real-time dynamic dubins-helix method for 3-d trajectory smoothing. *IEEE Transactions on Control Systems Technology*, 23(2):730–736, 2014.

[30] Lele Zhang, Fang Deng, Jie Chen, Yingcai Bi, Swee King Phang, Xudong Chen, and Ben M Chen. Vision-based target three-dimensional geolocation using unmanned aerial vehicles. *IEEE Transactions on Industrial Electronics*, 65(10):8052–8061, 2018.

[31] Kuang Zhao, Yong Zhou, Zhengyuan Zhou, Dengqing Tang, Yuan Chang, Qiang Fang, Han Zhou, and Tianjiang Hu. Onboard smart surveillance for micro-uav swarm: an experimental study. In *2018 IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, pages 67–72. IEEE, 2018.

[32] Yong Zhou, Qiang Fang, Kuang Zhao, Dengqing Tang, Han Zhou, Guoqi Li, Xiaojia Xiang, and Tianjiang Hu. Robust task-oriented markerless extrinsic calibration for robotic pick-and-place scenarios. *IEEE Access*, 2019.

[33] Pengfei Zhu, Longyin Wen, Xiao Bian, Haibin Ling, and Qinghua Hu. Vision meets drones: A challenge. *arXiv preprint arXiv:1804.07437*, 2018.

[34] Senqiang Zhu, Danwei Wang, and Chang Boon Low. Ground target tracking using uav with input constraints. *Journal of Intelligent & Robotic Systems*, 69(1-4):417–429, 2013.