RESEARCH　　　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

Target tracking algorithm combined part-based and redetection for UAV

Qiusheng He[1,2], Weifeng Zhang[1,2], Wei Chen[1,3*], Gang Xie[1,2] and Yanxin Yao[4]

* Correspondence: chenwdavior@
163.com
[1]School of Electronic and
Information Engineering, Taiyuan
University of Science and
Technology, Taiyuan 030024, Shanxi,
China
[3]School of Computer Science and
Technology, China University of
Mining and Technology Xuzhou,
Jiangsu 221116, China
Full list of author information is
available at the end of the article

Abstract

In the process of target tracking for UAV video images, the performance of the tracking algorithm declines or even the tracking fails due to target occlusion and scale variation. This paper proposes an improved target tracking algorithm based on the analysis of the tracking framework of the kernel correlation filter. First, four subblocks around the center of the target center are divided. A correlation filter fusing Histogram of Oriented Gradient (HOG) feature and Color Name (CN) feature tracks separately each target subblocks. According to the spatial structure characteristics in the subblocks, the center location and scale of the target are estimated. Secondly, the correct center location of target is determined by the global filter. Then, a tracking fault detection method is proposed. When tracking fails, the target redetection module which uses the normalized cross-correlation algorithm (NCC) to obtain the candidate target set in the re-detection area is started. Besides, this algorithm uses the global filter to obtain real target from the candidate set. In the meanwhile, this algorithm adjusts sectionally the learning rate of the classifiers according to detection results. Lastly, the performance of this algorithm is verified on the UAV123 dataset. The results show that compared with several mainstream methods, that of this algorithm is significantly improved when dealing with target scale variation and occlusion.

Keywords: Target tracking, Kernel correlation filter, Part-based tracking, Target redetection, UAV video image

1 Introduction

An unmanned aerial vehicle (UAV) is an aircraft without a human pilot on board, which exploits radio remote control equipment or self-provided program control devices. In recent years, the Internet of things has been developing rapidly. With the integration of information, communication, and network technology, the UAV derived from consumer leisure and entertainment toys can be applied to high-value commercial, agricultural, and defense fields, thus turning into a killer product. Chen et al. [1] design a traceable and privacy protection protocol to conduct the UAVs' application in sensitive control area. Chen et al. [2] propose a novel multi-hop clustering algorithm, called DMCNF. Lin et al. [3] present a novel moving-zone-based architecture and a corresponding routing protocol for message dissemination in VANETs by using

vehicle-to-vehicle communications only (i.e., without using vehicle-to-infrastructure communications). Chen proposes a CP-based method to analyze cellular network signals (i.e., NLUs, HOs, and CAs) for estimating vehicle speeds [4]. In addition, developed deep learning was a developed method to learn the potentially complex and irregular probability distributions [5]. Cheng et al. [6] introduce a novel Markov Random Field (MRF) model to describe the data correlation among sensor nodes. Owing to the research results in the UAV and communication fields, the UAV has great potential in different domains and under various missions, as well as greater flexibility in application.

At present, the UAV has become a more powerful and reliable task performer in terms of aerial photography, investigation, search, and rescue. Since the video image-related target tracking technology can provide autonomous navigation information for the UAV to settle many thorny problems, visual processing technology is of great significance to the UAV system. Liu et al. [7] propose a deformable convolution layer to enrich the target appearance representations in the tracking-by-detection framework. Huang et al. [8] design a bidirectional tracking scheme to solve the problem of model drift in online tracking. The tracking algorithm in line with deep learning has ideal performance but is generally time-consuming in [7–13]. However, the tracking algorithm based on the correlation filter also can well adapt to the variation of the target appearance, considering its extremely fast computation speed and good positioning performance in the Fourier domain. Therefore, the primary real-time tracking algorithm is the improvement of the target tracking algorithm according to correlation filters. Bolme et al. [14] propose the MOSSE tracking algorithm, which first applies the correlation filter to visual tracking. Henriques et al. [15] put forward the CSK method, an algorithm with a fairly good performance and high calculation speed. It is worth noting that those two trackers both feature a single-channel Gray function. Danelljan et al. [16] improve the CSK methods by using the CN feature that is a multiple channel feature. Bolme et al. [17] propose the KCF method, which further enhanced the efficiency of the CSK tracker by the use of HOG features. Besides, the ridge regression problem in linear space is mapped to nonlinear space by the kernel method. For the scale evaluation problem, the DSST tracker in [18] exploits the HOG feature to learn an adaptive multi-scale correlation filter, which aims to evaluate the scale variation of the object target. Apart from that, a series of modified algorithms, such as SAMF and RPT [19–24], have been proposed successively.

Due to flight altitude change, the wide flight range and the intricate background, the UAV target tracking is apt to generate tracking challenges, such as target occlusion or target movement out of view. Hence, how to devise a more steady and accurate tracking algorithm is a challenging problem in UAV target tracking. This design should help the tracking algorithm accurately trace the target when occluded and recapture the target when it reappears.

This paper proposes an improved KCF algorithm in combination with abilities based on parts and redetection. In an attempt to improve the tracking performance of the algorithm under occlusion, part-based tracking strategy fused with the multi-feature is employed on the basis of the traditional KCF algorithm. For the sake of solving the issue of target scale variation, the relative position variation of corresponding subblocks in two adjacent frames is used for the calculation of the target scale step. Last, the

target tracking failure is detected by calculating the value of target tracking confidence FPRM, through which the learning rate is sectionally adjusted. When the tracking fails, the target redetection module starts to retrieve the real target. Overall, this algorithm is more robust.

The remainder of this paper is organized as follows. Section 2 reviews the basic theoretical knowledge of the KCF algorithm, and Section 3 introduces the framework of this proposed method. Experimental results and analysis are shown in Section 4, with the conclusion given in Section 5.

2 The KCF tracker

In the KCF algorithm [17], the target tracking problem is transformed into solving ridge regression problem under minimum mean square error. The goal of training is to find a function $f(z) = \omega^T z$ that minimizes the squared error over samples x_i and their regression targets y_i:

$$\min_{\omega} \sum_{i}^{n} (f(x_i) - y_i)^2 + \lambda \|\omega\| \tag{1}$$

where Eq. (1) is solved to obtain optimal solution, which makes the cost function minimized. Where the λ is a regularization parameter to control overfitting, x_i represents the training sample generated by the shift of the base sample through circulation. $y_i \in [0, 1]$ is the training label, which is the expected output of the training sample x_i. y is the Gaussian distribution. The function f can be written as the linear combination of basis samples: $f(x) = \omega^T x$. ω is the classifier parameter.

To deal with the case of nonlinear regression, we propose $f(z) = \omega^T z = \sum_{i=1}^{n} \alpha_i \kappa(z, x_i)$ in which the kernel function $\kappa(z, x_i)$ is introduced, and the cyclic matrix and the Fourier transform are used. For the most commonly used kernel functions, the circulant matrix trick can also be used [17]. For dual space coefficients α can be learnt as below:

$$\hat{\alpha}^* = \frac{\hat{y}}{\hat{\kappa}^{xx} + \lambda} \tag{2}$$

where $\wedge$ denotes Fourier transform. $\hat{\alpha}^*$ denotes the complex-conjugate of $\hat{\alpha}$. We adopt the Gaussian kernel which can be applied in the circulant matrix trick as below:

$$\kappa^{xx'} = \exp\left(-\frac{1}{\sigma^2}\left(\|x\|^2 + \|x'\|^2\right) - 2F^{-1}\left(\hat{x} \otimes \hat{x}'^*\right)\right) \tag{3}$$

where $\otimes$ denotes the dot product of the corresponding position elements among the matrixes.

In the rapid detection process, we define that the image patch z at the same position in the next frame is the reference sample. The response map of the input image in time domain can be obtained by fast target detection:

$$F(z) = F^{-1}\left(\left(\hat{\kappa}^{\tilde{x}z}\right)^* \otimes \hat{\alpha}\right) \tag{4}$$

where $\tilde{x}$ denotes data learned from the model. F^{-1} denotes inverse discrete Fourier transform. The position corresponding to the maximum response is the location of the target being tracked.

$$P_t = \ \arg \ \max_{P \in R_t} F(z) \tag{5}$$

where P_t denotes the center position of the target in the current frame. R_t denotes search area. P represents all possible target positions in the search area, and the search area is set to 2.4 times the target size in this paper.

3 The improvement of KCF algorithm

The traditional KCF algorithm fails to consider the scale problem and lacks the deficiency of tracking failure detection. This paper proposes two improvements of the KCF algorithm as follow. First, the multi-feature fusion block tracking strategy is used, and the s-type function is used to optimize the scale step, so the ability of the algorithm to deal with occlusion and scale variation is improved. Second, a method of tracking failure detection is proposed, and then target is detected again after tracking failure. According to the results of tracking failure detection, the learning rate is adjusted in sections. A more robust tracking algorithm is obtained.

3.1 Part-based tracking strategy based on multi-feature fusion

When the target is occluded, the improvement of KCF algorithm can still use the unblocked part to provide accurate positioning information for tracking in [24]. When the target scales change, the tracking results of each subblock can overlap and separate accordingly. As a result, the target scale is deduced by using this theorem. HOG feature describes the distribution of gradient intensity and gradient direction in the local area of the images. CN feature describes the global surface properties of the images. The two features are fused which the performance of the tracker is improved by using the complementary advantages among different features in [19]. Furthermore, the high-speed processing capability of correlation filtering algorithm provides the possibility of real time part-based tracking [20].

The blocking method in this paper, four equal-sized subblocks, are generated around the target center locations, as shown in Fig. 1. To be more exact, the height and width of the four subblocks are 0.5 times that of the overall target box respectively. The central coordinates of the upper and lower subblocks are separately moved up and down
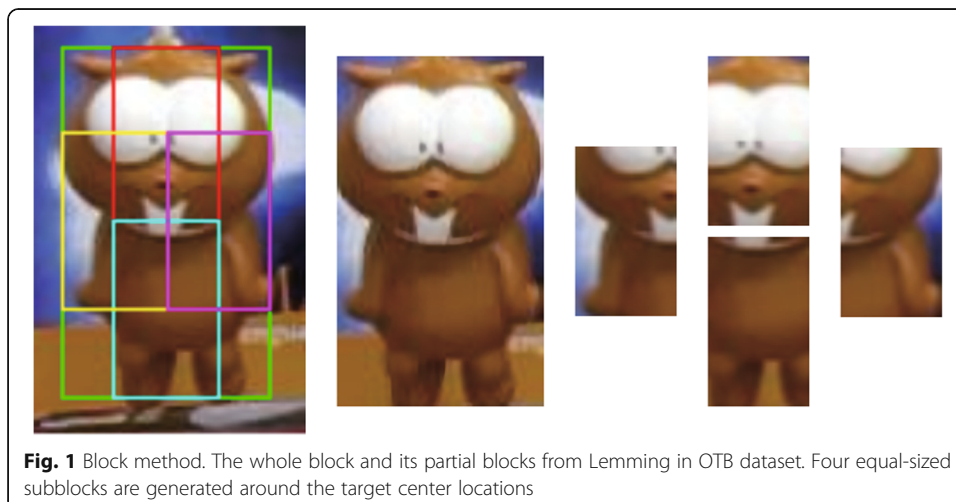

Fig. 1 Block method. The whole block and its partial blocks from Lemming in OTB dataset. Four equal-sized subblocks are generated around the target center locations

0.25h from the target center location. The center coordinates of the left and right sub-blocks are moved to the left and right 0.25w from target center location, individually. The function is given as below:

$$B_k = \begin{cases} [P_t(1,1){-}m & P_t(1,2)] & k=1 \\ [P_t(1,1){+}m & P_t(1,2)] & k=2 \\ [P_t(1,1) & P_t(1,2){-}n] & k=3 \\ [P_t(1,1) & P_t(1,2){+}n] & k=4 \end{cases} \tag{6}$$

where B_k denotes the central location of all target subblocks. $k \in (1,2,3,4)$ denotes the index of the target subblocks. $m = 0.25h$ and $n = 0.25w$, ω and h are respectively represent the height and width of the overall target box. P_t represents the center position of the global target box.

This paper proposes the concatenation of Histogram of Oriented Gradient (HOG) and Color Name (CN) as the feature representation to balance discriminative power and efficiency. The KCF tracking algorithm and fusion feature described above are used to track each target subblock independently to obtain the response map:
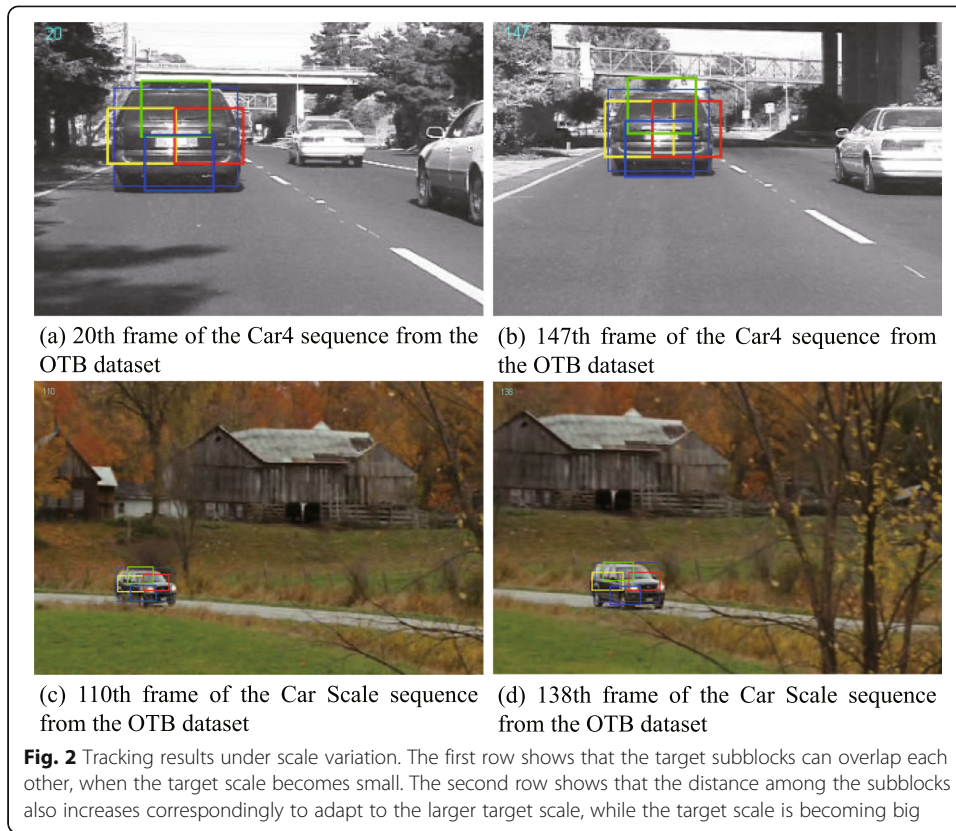
$$k^{'} = \underset{k \in (1,2,3,4)}{\arg \max} (\max(F_k)) \tag{7}$$

where $k^{'}$ denotes the index of the selected response. The selected response is then used to calculate the location $\hat{B}_{k'}$ of the corresponding subblocks. The rough tracking results $\tilde{P}_t$ are estimated by combining the position constraint relationship between the part and the whole of the target. And then uses the global filter to determine the exact location of the target center P.

The general solution to solve the scale problem is to calculate the confidence level at different scales by scaling the original trace window. And then, the target size is determined by selecting the maximum scale of confidence level as the target scale. As shown in Fig. 2, this paper uses a part-based tracking method. Although the tracking box size of each target subblock is constant, the relative location among each block varies with the target scales. In this paper, the character as above in part-based tracking are quantified to calculate the scale of the target. First, preliminary estimation of target scale steps is obtained by calculating the average Euclidean distance ratio among the center locations of all the target subblocks in the current frame and that of the corresponding subblocks in the previous frame:

$$\tilde{step} = sqrt\left(\frac{1}{K(K{-}1)} \sum_{i=1}^{K} \sum_{i=1}^{K} \frac{dist\left(\hat{B}_i^t, \hat{B}_j^t \right)}{dist\left(\hat{B}_i^{t-1}, \hat{B}_j^{t-1} \right)} \right) \quad (i \neq j) \tag{8}$$

where $\hat{B}_i^t$ is the location obtained by the independent tracking of the ith target subblock in the tth frame. *dist* denotes Euclidean distance. In order to avoid the error caused by the poor tracking quality of individual subblocks, an s-type function is used to constrain the scale step size based on the properties of scale variation:

(a) 20th frame of the Car4 sequence from the OTB dataset

(b) 147th frame of the Car4 sequence from the OTB dataset

(c) 110th frame of the Car Scale sequence from the OTB dataset

(d) 138th frame of the Car Scale sequence from the OTB dataset

Fig. 2 Tracking results under scale variation. The first row shows that the target subblocks can overlap each other, when the target scale becomes small. The second row shows that the distance among the subblocks also increases correspondingly to adapt to the larger target scale, while the target scale is becoming big

$$\text{step} = (0.12/(1 + \exp(-20\tilde{\text{step}} + 20))) + 0.94 \tag{9}$$

where *step* denotes the step size of the target scale, which means change of target size compared with the previous frame. The scale parameter is given by Eq. (10):

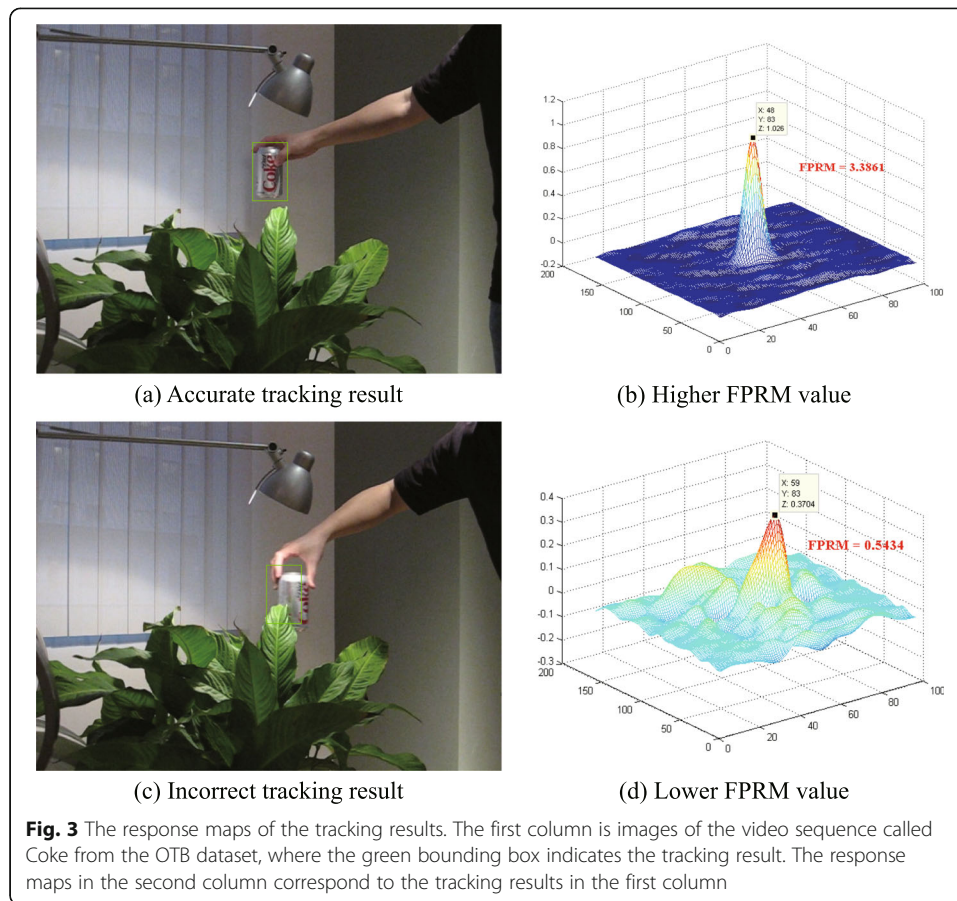$$S_t = S_{t-1} \cdot \text{step} \tag{10}$$

where S_t is the scale parameters of the target in the current frame. The scale of the target in the current frame can be obtained through the scale of the previous frame multiplied by the scale step size calculated in the current frame.

3.2 Reliability estimation

To some degree, the peaks and fluctuations of response can reveal the confidence of the tracking results [25]. The ideal response map has only one peak and is smooth in other areas when the detected trace frame matches the real target, and the response fluctuates dramatically and the peaks become smaller, when the target tracking fails, as can be seen in Fig. 3. Therefore, FPRM (fluctuations and peaks of response map) can be calculated by using this property:

$$\text{FPRM} = F_{\max} \cdot \ln\left(\frac{|F_{\max} - F_{\text{mean}}|}{|F_{\min} - F_{\text{mean}}|}\right) + e \tag{11}$$

Where $F_{\max}$ denotes maximum value in response. $F_{\min}$ denotes minimum value in response. F_{mean} denotes the average of all the elements in the response. The number e is

(a) Accurate tracking result (b) Higher FPRM value

(c) Incorrect tracking result (d) Lower FPRM value

Fig. 3 The response maps of the tracking results. The first column is images of the video sequence called Coke from the OTB dataset, where the green bounding box indicates the tracking result. The response maps in the second column correspond to the tracking results in the first column

a mathematical constant that is the base of the natural logarithm: the unique number whose natural logarithm is equal to one. It is approximately equal to 2.71828. When the target tracking fails, the response fluctuation increases, the peak decreases, and the value of FPRM decreases.

This FPRM distribution is shown in Fig. 4. When the serious tracking drift and tracking faint appear, the corresponding FPRM value becomes extremely low, which is 0.2302 and 0.1487 in frames 26 and 47, respectively. Besides, those two values are around 0.3 and 0.26 times of the $FPRM_{mean}$ values of the corresponding positions, respectively. From the 74th frame to the 77th frame, the target starts to enter occluded areas. Accordingly, the value of FPRM decreases rapidly to somewhere near and below the value of $FPRM_{mean}$ in the corresponding position. In the 104th frame, the target box perfectly matches the actual location of the target, with its corresponding FPRM value also reaching the maximum that far exceeds the $FPRM_{mean}$ of the relevant position. Therefore, this character can be used to determine whether the tracking fails and attest the effectiveness of this tracking failure detection method proposed as well.

3.3 Target redetection and model update

Traditional correlation filter tracking algorithm lacks tracking failure detection [26, 27]. Once target tracking is lost, it is difficult to recover. This section shows that aiming at
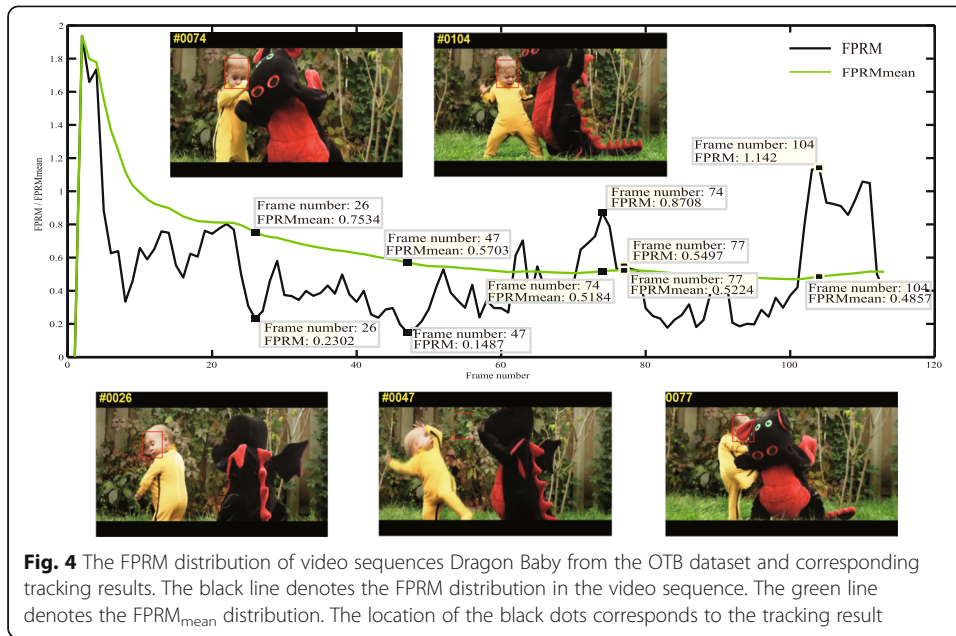
Fig. 4 The FPRM distribution of video sequences Dragon Baby from the OTB dataset and corresponding tracking results. The black line denotes the FPRM distribution in the video sequence. The green line denotes the FPRM$_{mean}$ distribution. The location of the black dots corresponds to the tracking result

the shortcoming of the lack of target faint detection in the kernel correlation filter; the improvement of this algorithm is achieved. When tracking failure is detected, the target redetection module is started to retrieve the real target.

When a trace failure is detected, that is FPRM≤0.3FPRM$_{mean}$. Meanwhile, when target tracking fails, the peak of the response is recorded as F_θ. And the target redetection module needs to be started. First, the target redetection module gets the set ϕ_t of candidate target locations p in the recheck area via normalized cross correlation algorithm (NCC), as shown in Fig. 5. The normalized correlation altogether between matrix x and matrix z, which have the same size, is defined as follows [28]:

$$\lambda(x,z) = \frac{\sum_{i,j}\left(x_{i,j}-\overline{x}\right)\left(z_{i,j}-\overline{z}\right)}{\sqrt{\sum_{i,j}\left(x_{i,j}-\overline{x}\right)^2\sum_{i,j}\left(z_{i,j}-\overline{z}\right)^2}} \tag{12}$$

where $x_{i,j}$ is the value of the elements in row i and column j of the matrix x. The larger $\lambda(x,z)$ is, the higher the correlation between matrix x and matrix z is.

For each candidate target in the set ϕ_t, the correlation filtering model of fusion feature is used to detect whether it is a real target. This paper uses Eq. (11) to calculate its corresponding FPRM$_{t,p_i}$, where p_i denotes the ith element in the candidate target set. A candidate target p with a maximum value of FPRM$_{t,\ p}$ is found in all alternative target. And, when FPRM$_{t,\ p} \geq 1.2F_\theta$, the candidate target p is defined as a real target obtained from the current frame by this target redetection module. When the current frame fails to detect a real target, the results of the part-based tracking are maintained. In the next frame, the target redetection module continues to work until a reliable target is detected.

The traditional correlation filter tracking algorithm updates the features of target appearance and classifier parameters in each frame, without considering whether the
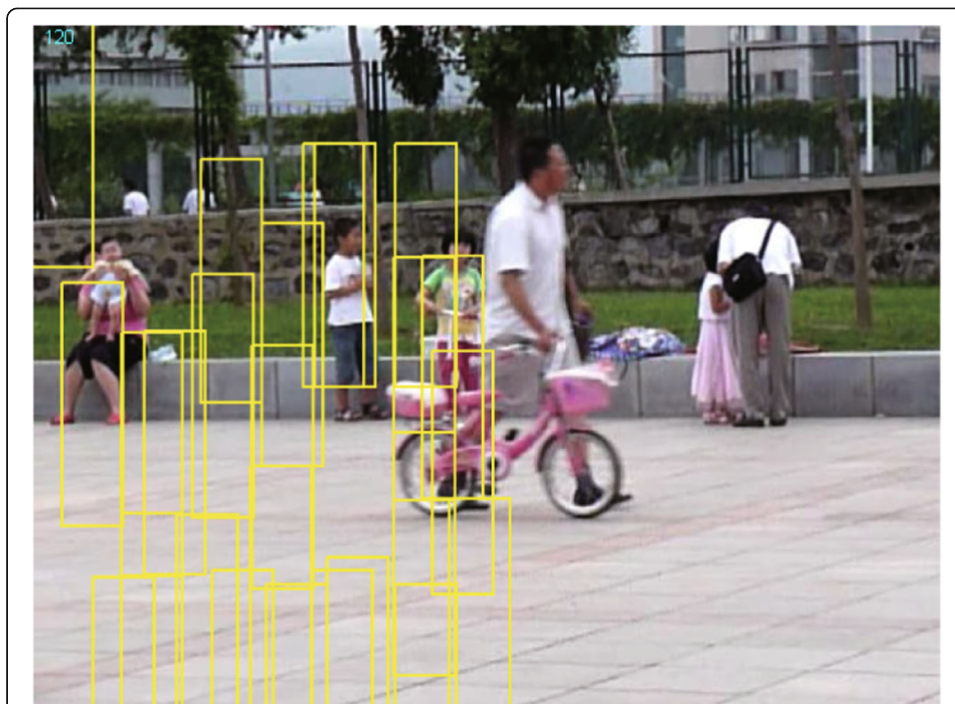
Fig. 5 Candidate target set. The redetection module generates a candidate target set in the redetection region via normalized cross correlation algorithm (NCC), and then the correlation filtering model of fusion feature is used to detect whether it is a real target

tracking is reliable. This update method will introduce errors into the model, resulting in tracking drift and even tracking failure. This paper proposes a method to adjust the classifier learning rate according to the result of tracking failure detection to reduce the error into the model. Through series of simulation experiments, the learning rate γ_t is:

$$\gamma_t = \begin{cases} \gamma & \text{FPRM}_{\text{mean}} \leq \text{FPRM} \\ 0.8\gamma & 0.3\text{FPRM}_{\text{mean}} \leq \text{FPRM} \leq \text{FPRM}_{\text{mean}} \\ 0 & \text{FPRM} \leq 0.3\text{FPRM}_{\text{mean}} \end{cases} \tag{13}$$

where $\text{FPRM}_{\text{mean}}$ denotes the average of FPRM calculated from 1st to $(t\text{-}1)$ th frames during the tracking process. $\gamma = 0.018$.

On the one hand, the problem of model drift caused by occlusion, deformation, etc. is avoided by adjusting the learning rate in sections during tracking process. On the other hand, when the target is lost, the target redetection module is started to retrieve the real target.

3.4 Algorithmic process

The tracking algorithmic in this paper uses the multi-feature fusion block tracking strategy and introduces a target recheck module to retrieve the target after it is lost. Meanwhile, the learning rate of the classifier is adjusted according to the tracking quality.

Input：

Input image patch I_t

The last frame position P_{t-1} and scale S_{t-1}

$FPRM_{mean}$ of the front $t-1$ frames

Apparent model x_{t-1} and classifier coefficient α_{t-1}

Output：

Target position P_t and scale S_t

$FPRM_{mean}$ of the front t frame

The updated apparent model x_t and classifier coefficient α_t

---

(1) Divide the input image patch I_t into four subblocks.

(2) Track each subblock individually by using the tracking algorithm and fusion feature in the previous article and generate the response maps of each subblock.

(3) Estimate roughly the tracking result $\widetilde{P}_t$, according to the position constraint relationship between the target subblock and the target overall.

(4) Obtain the accurate tracking results P_t based on the rough estimation results from the global filter, and then calculate the target scale S_t .

(5) Calculate $FPRM$ in order to determine whether the target redetection module needs to be started. When this module needs to be started, it should work until the real target is detected. Otherwise, directly execute step (6).

(6) Obtain the apparent model x_t and the classifier coefficient α_t based on the learning rate adjusted by the tracking quality.

(7) Start to the step (2) when the scale S_t is between 0.8 and 1.2. Otherwise, start to the step (1).

---

4 Results and discussion

4.1 Experiment setup

In order to comprehensively evaluate the effectiveness of the proposed algorithm in this paper, the proposed algorithm is compared with five correlation filtering algorithms with excellent comprehensive performance on the UAV123 benchmark dataset [29]. The trackers used for our comparison are CSK [15], CN [16], KCF [17], DSST [18], and SAMF [19]. The UAV123 dataset contains 12 different attributes such as illumination variation (IV), scale variation (SV), partial occlusion (POC), full occlusion (FOC), out-of-view (OV), fast motion (FM), camera motion (CM), background clutter (BC), similar object (SOB), aspect ratio change (ARC), viewpoint change (VC), and low resolution (LR).

In this experiment, the experimental platform is Matlab2014a, and all the experiments are conducted on an Intel i3-3110M CPU (2.40 GHz) PC with 6 GB memory. The parameters of the algorithm in this paper are set as follows: the σ used in Gaussian function is set to 0.5, the cell size of HOG is 4×4, and the orientation bin number of HOG is 9. The parameters of five other algorithms used for comparison are the default parameters of the source program published by the authors [15–19].

In all the experiments, the performance is measured with three evaluation criteria which are overlap region (OR), distance precision (DP), and overlap precision (OP) [30]. For the OR, it reflects the overlap between the tracking output bounding box r_t and the ground truth bounding box g_t. The OR is $o = |r_t \cap g_t|/|r_t \cup g_t|$, where $|\cdot|$ denotes

the area. The DP refers to the percentage of frames whose center location error (CLE) are less than a certain threshold (20 px) accounts for the total frames during the tracking process. In addition, the OP is used which is the percentage of frames with OR greater than a certain threshold (0.5) in the total frames in the tracking process. As an evaluation criterion, larger value of OR, DP, and OP indicates better performance of algorithms.

4.2 Qualitative experimental results and analysis

This section mainly analyzes the effectiveness of this algorithm from two perspectives of scale variation and occlusion. The occlusion includes partial occlusion (POC) and full occlusion (FOC). The algorithm in this paper and five algorithms mentioned above, are analyzed and contrasted. In the qualitative analysis experiment, six video sequences with scale variation and occlusion attributes, including Boat6, Car9, Group1-4, Person4-1, Person1-s, and Person19-2, are selected from the UAV123 dataset. Table 1 shows the specific attributes of these six sets of video sequences.
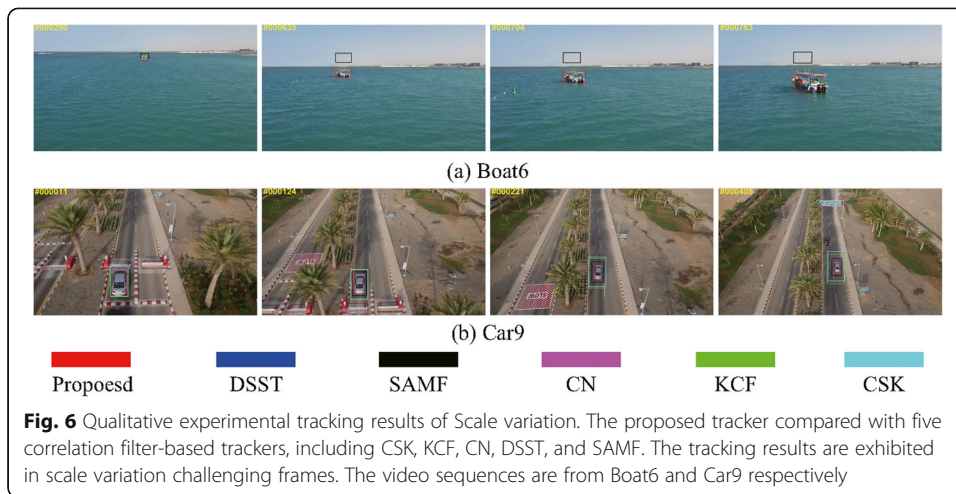
4.2.1 Experimental results and analysis of scale variation

The scale variation is the only one attribute in Fig. 6a. This video describes a scene of a ship moving in the water towards a drone hovering in the air. The distance between the target and the UAV image equipment becomes smaller. The original target size grows up from 27×16 to 434×264, which means the ultimate target size is 16 times as large as the original one. From the experimental results, although, DSST, SAMF, and the algorithm in this paper, have the ability of scale estimation. However, the scale estimation capabilities of the DSST and SAMF are difficult to cope with the drastic variation of the target scale during the tracking process. The tracking boxes of these two algorithms fail to change as the change of target size. In addition, these algorithms except for that in this paper have shift to some degree and cannot accurately track the target because of the relatively distant initial position, the relatively small target size, and relatively low resolution.

There are target scale changes in Fig. 6b. This video illustrates a walking car, as a target, is gradually moving away from the drone in the air. The target size keeps a downward trend, declining from 99×169 to 41×40. From the experimental results, all the algorithms can capture the moving targets, while only the algorithm in this paper and DSST algorithm can estimate the target scale variation more accurately.

Table 1 The attributes of these six sets of video sequences

Video sequences	Attributes
Boat6	SV
Car9	SV, ARC, LR, FM, POC, CM, SOB
Group1–4	SV, ARC, POC, IV, VC, CM, SOB
Person20	SV, ARC, POC, IV, VC, CM, SOB
Person1-s	SV, FM, POC, IV, VC, CM
Person19-2	SV, ARC, FM, FOC, POC, OV, CM

Fig. 6 Qualitative experimental tracking results of Scale variation. The proposed tracker compared with five correlation filter-based trackers, including CSK, KCF, CN, DSST, and SAMF. The tracking results are exhibited in scale variation challenging frames. The video sequences are from Boat6 and Car9 respectively

4.2.2 Experimental results and analysis of occlusion

Figure 7 a shows partial occlusion and slight-scale variation. There are three people walking in a park together. As can be seen, three people walk together, while one of them can be captured as the target. They occlude each other during the walking. As the shooting angle of the UAV changes, so does the target scale. According to the experimental results, DSST, CN, and CSK fail to track due to interference factors such as occlusion and scale variation. Although SAMF and KCF can keep up with the target's movement, their results also exist shift to some extent. Overall, the algorithm in this paper can accurately track the target.

Fig. 7 Qualitative experimental tracking results of occlusion. The proposed tracker compared with five correlation filter-based trackers, including CSK, KCF, CN, DSST, and SAMF. The tracking results are exhibited in challenging frames, i.e., partial occlusion and full occlusion. The video sequences are from the Group1–4, Person20, Person1-s, and Person19-2 respectively
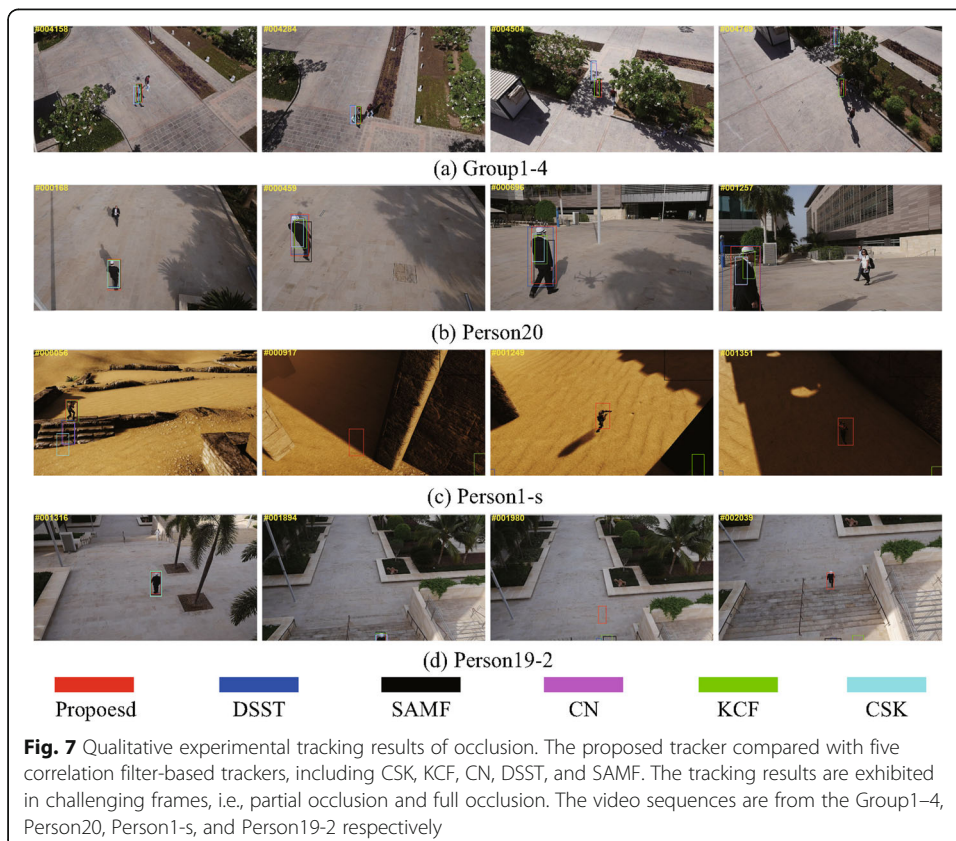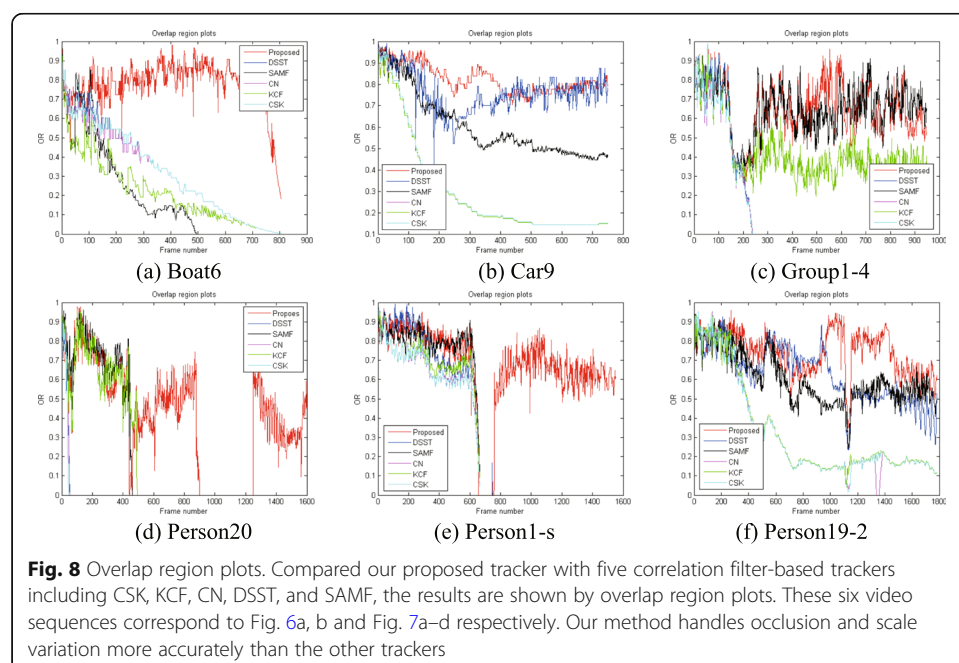
Figure 7 b illustrates partial occlusion and large-scale variation. The video exhibits the process of pedestrians walking in a certain scenario. The target walks from far to near, while the target scale is increasing. The target is partially occluded due to the change of shooting angles, which means the SAMF, CN, KCF, and CSK in this video is severely shifted. Although the DSST can still adapt to this target's scale change, it still exists shift to some extent.

Figure 7 c exhibits partial occlusion and slight scale variation. As can be seen, a scene where the character in the game moves quickly. This video includes the process from this game character's disappearance to reappearance. In the 56th frame of the video, the DSST, CN, and CSK fail to track due to the target's rapid movement. The target enters the obstruction, and then disappears completely in the 917th frame. When the target reappears, the proposed algorithm recaptures the target in the 1249th frame. Therefore, this algorithm can track the target steadily during the subsequent process.

Figure 7 d shows partial occlusion, full occlusion, and slight scale variation. There is a pedestrian walking on the stairs. The target is completely occluded due to the change of shooting angle in the pictures. The target completely disappears from the video in the 1980th frame. When the target reappears, it can be retracked and continue to be followed due to this algorithm with target redetection mechanism.

4.2.3 Experimental results and analysis based on overlap region

High overlap region requires not only lower tracking center location error, but also higher accuracy of scale elevation. So, the OR can better represent the tracking performance of the algorithm. As can be seen from Fig. 8, when the target scale varies significantly (such as Boat6, Car9, and Person20), this algorithm's scale estimation method can also successfully estimate the target's scale variation. When the target is occluded to some extent (such as Group1−4, Person20), this algorithm can maintain a good

Fig. 8 Overlap region plots. Compared our proposed tracker with five correlation filter-based trackers including CSK, KCF, CN, DSST, and SAMF, the results are shown by overlap region plots. These six video sequences correspond to Fig. 6a, b and Fig. 7a–d respectively. Our method handles occlusion and scale variation more accurately than the other trackers
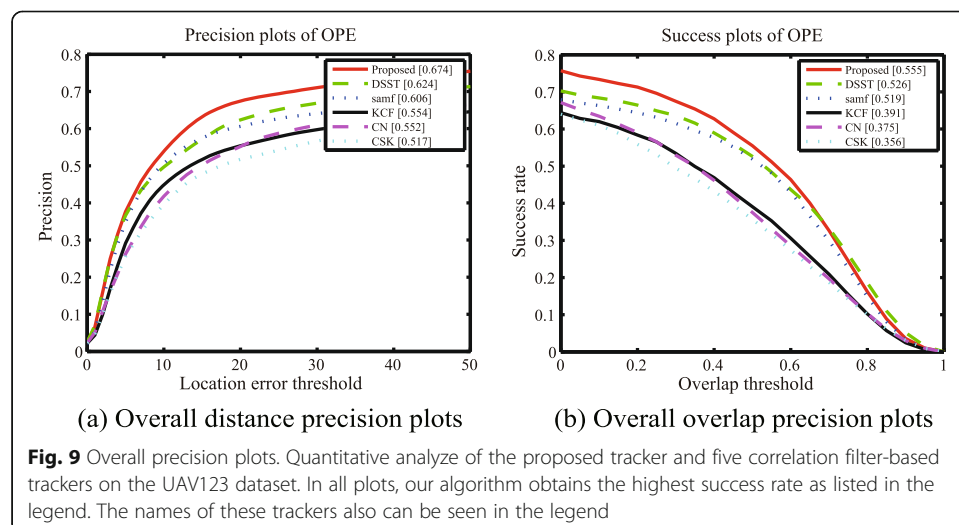
tracking effect compared with other algorithms. Additionally, due to the additional target re-detection module, the algorithm's ability to process long-term tracking has been greatly improved. When the target is completely occluded, disappears, and then reappears (such as 900th frame~1249th frame in Person1-s, 1907th frame~2010th frame in Person19-2), this algorithm can retrack the target in the tracking process.
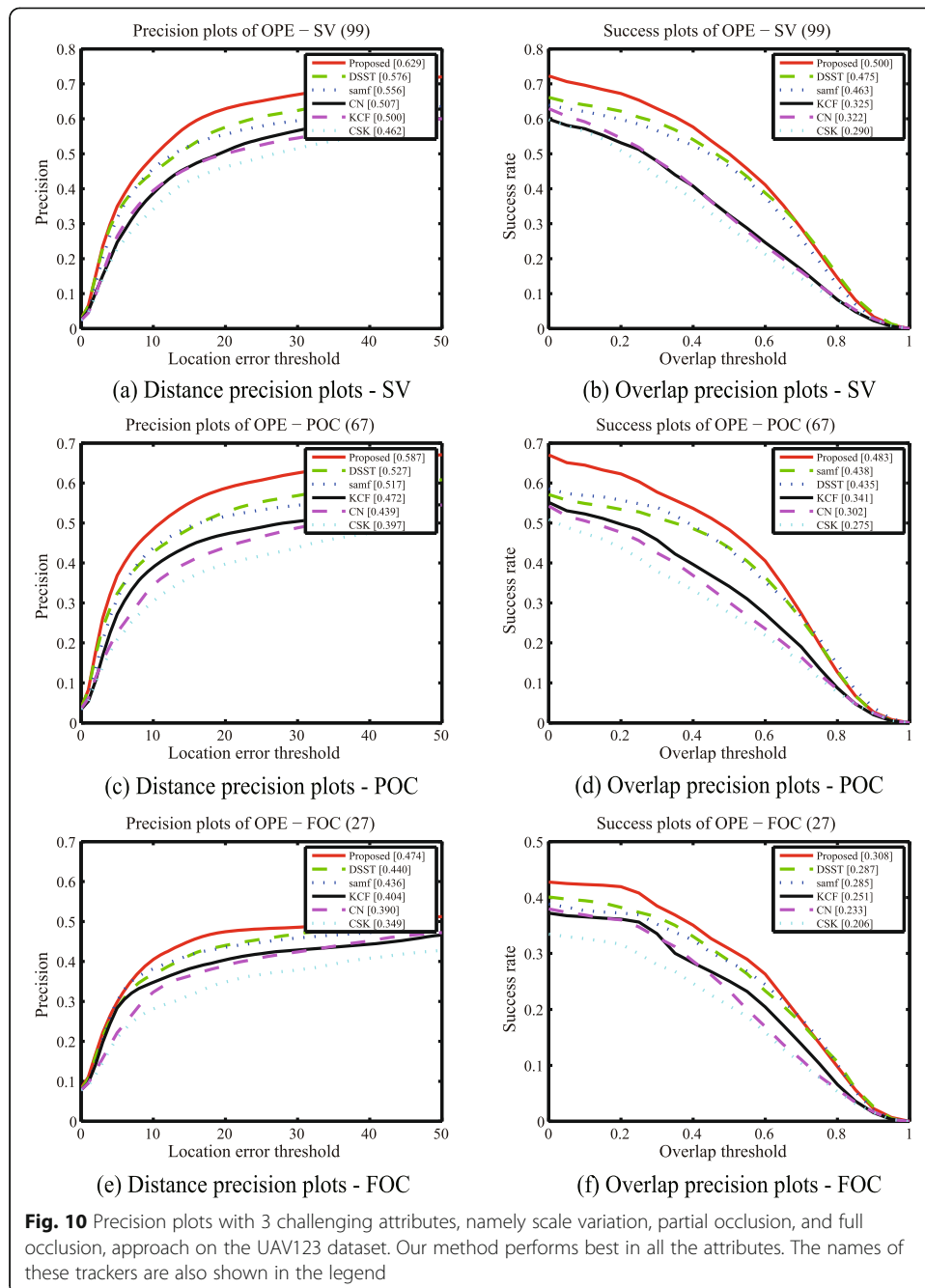
4.3 Quantitative experimental results and analysis

In order to more intuitively reflect the overall performance of this algorithm, this paper uses the UAV123 benchmark dataset to verify the proposed algorithm and compares it with other five correlation filtering algorithms with better performance. Distance precision (DP) and overlap precision (OP) are used to represent the overall performance of the tracking algorithm.

Figure 9 a and b respectively shows the distance precision plots and overlap precision plots under different thresholds. As can be seen from the figure, the tracking effect of the algorithm in this paper is the best compared with the other five algorithms. The DP and the OP of the algorithm in this paper are 0.674 and 0.555, respectively. Compared with the original KCF algorithm, the DP and the OP increase by 12% and 16.4% individually. The main reason is that this algorithm uses the part-based tracking method based on KCF, which improves its anti-occlusion and processing scale variation performance. In addition, compared with DSST and SAMF algorithms with scale estimation capabilities, this algorithm adds a target redetection module to make it obtain the ability of long-term tracking, which makes it more suitable for application scenarios for UAV video image tracking.

In order to verify the performance of the proposed algorithm on scale variation and occlusion scenarios, Fig. 10 shows the distance precision and overlap precision plots of these six kinds of algorithms under three attributes: scale change (SV), partial occlusion (POC), and full occlusion (FOC). It can be seen from Fig. 10 that the algorithm proposed in this paper performs excellently in the face of scale changes and occlusions and ranks first under these three attributes.

(a) Overall distance precision plots (b) Overall overlap precision plots

Fig. 9 Overall precision plots. Quantitative analyze of the proposed tracker and five correlation filter-based trackers on the UAV123 dataset. In all plots, our algorithm obtains the highest success rate as listed in the legend. The names of these trackers also can be seen in the legend

Fig. 10 Precision plots with 3 challenging attributes, namely scale variation, partial occlusion, and full occlusion, approach on the UAV123 dataset. Our method performs best in all the attributes. The names of these trackers are also shown in the legend

5 Conclusion

Based on the research achievements in the field of target tracking, UAV technology, representing the future of many industries, will have a wider application prospect in the future. For example, home delivery by the UAV will become one of the most common applications; the miniature UAV will be used to pollinate plants. Advanced technology can bring a better and convenient living environment for mankind.

This paper studies the problem of target tracking for UAV video images. Based on the KCF algorithm, a tracking algorithm combined with abilities based on parts and redetection is proposed. This algorithm is conducive to dealing with the tracking failure

resulted from the situation when the target is briefly occluded and re-enters the field of view after leaving the camera for a period of time. Part-based tracking strategy in accordance with multi-feature fusion improves the anti-occlusion performance of the algorithm and solving the problem of target scale variation. Concerning target tracking confidence FPRM, it is used to determine whether the tracking failed. When it does, the target redetection module initiates the retrieval of the real target to enable the algorithm to achieve long-term tracking. Meanwhile, the learning rate is partly adjusted through the value of target tracking confidence FPRM to reduce the introduction of errors. Seen from the quantitative experiment results, the algorithm in this paper has obvious edges in dealing with occlusion and scale change problems. In addition, qualitative experimental results show that the distance accuracy (DP) and overlap rate accuracy (OP) of the proposed algorithm perform best in both the overall and challenge tests. In general, our method outperforms other mainstream algorithms in the literature on the UAV123 dataset. Notably, stabilized tracking for the ground moving target can be realized when the algorithm processes video sequences with scale variation and occlusion. The fundamental purpose of the algorithm proposed in this paper is to track a single target. In the future, UAV detection and tracking of multiple targets will be explored.

Abbreviations
UAV: Unmanned aerial vehicle; HOG: Histogram of Oriented Gradient; CN: Color Name; NCC: Normalized cross-correlation; KCF: Kernelized correlation filter; SAMF: Scale adaptive with multiple features; DSST: Discriminative Scale Space Tracker; CSK: Circulant structure of tracking-by-detection with kernels; RPT: Reliable patch trackers; CLE: Center location error; DP: Distance precision; OP: Overlap precision; OR: Overlap rate; IV: Illumination variation; SV: Scale variation; POC: Partial occlusion; FOC: Full occlusion; OV: Out-of-view; FM: Fast motion; CM: Camera motion; BC: Background clutter; SOB: Similar object; ARC: Aspect ratio change; VC: Viewpoint change; LR: Low resolution; OTB: Online object tracking benchmark; FPRM: Fluctuations and peaks of response map

Authors' contributions
Qiusheng He and Weifeng Zhang proposed the idea and had extensively participated in all of the paper writing. Wei Chen mainly provided the guidance for deriving these expressions. Gang Xie and Yanxin Yao mainly added to and revised the related works. All of the authors equally contributed to reviewing the manuscript. The authors read and approved the final manuscript.

Availability of data and materials
Not applicable.

Competing interest
The authors declare that they have no competing interests.

Author details
[1]School of Electronic and Information Engineering, Taiyuan University of Science and Technology, Taiyuan 030024, Shanxi, China. [2]Shanxi Key Laboratory of Advanced Control and Intelligent Information System, Taiyuan 030024, Shanxi, China. [3]School of Computer Science and Technology, China University of Mining and Technology Xuzhou, Jiangsu 221116, China. [4]Beijing Information Science & Technology University, Beijing 100101, China.

References

1. C.L. Chen, Y.Y. Deng, W. Weng, et al., A traceable and privacy-preserving authentication for UAV communication control system. Electronics 9(1), 62 (2020)
2. Y. Chen, M. Fang, S. Shi, et al., Distributed multi-hop clustering algorithm for VANETs based on neighborhood follow. EURASIP J. Wirel. Commun. Netw.. 2015(1), 98 (2015)
3. D. Lin, J. Kang, A. Squicciarini, et al., MoZo: a moving zone based routing protocol using pure V2V communication in VANETs. IEEE Trans. Mob. Comput. 16(5), 1357–1370 (2016)
4. C.H. CHEN, A cell probe-based method for vehicle speed estimation[J]. IEICE TRANSACTIONS on Fundamentals of Electronics. Communications and Computer Sciences. 103lgorithmic process(1), 265–267 (2020)
5. C.H. Chen, F. Song, F.J. Hwang, et al., A probability density function generator based on neural networks. Physica A: Statistical Mechanics and its Applications. 541, 123344 (2020)
6. H. Cheng, Z. Su, N. Xiong, et al., Energy-efficient node scheduling algorithms for wireless sensor networks using Markov Random Field model. Inf. Sci. 329, 461–477 (2016)
7. W. Liu, Y. Song, D. Chen, et al., Deformable object tracking with gated fusion. IEEE Trans. Image Process. 28(8), 3766–3777 (2019)
8. Z. Huang, Y. Yu, M. Xu, Bidirectional tracking scheme for visual object tracking based on recursive orthogonal least squares. IEEE Access. 7, 159199–159213 (2019)
9. M. Danelljan, G. Bhat, F.S. Khan, et al., ECO: efficient convolution operators for tracking. Paper Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017), pp. 6931-6939
10. J. Choi, H. Chang, S. Yun, et al., Attentional correlation filter network for adaptive visual tracking. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017), pp. 4828-4837
11. L. Bertinetto, J. Valmadre, J. Henriques, et al., Fully-convolutional Siamese networks for object tracking. Springer Verlag. 9914, 850–865 (2016)
12. B. Li, J. Yan, W. Wu, et al., High performance visual tracking with Siamese region proposal network. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018), pp. 8971-8980
13. Q. Wang, L. Zhang, L. Bertinetto, et al., Fast online object tracking and segmentation: a unifying approach. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2019). pp. 1328-1338
14. D.S. Bolme, J.R. Beveridge, B.A. Draper, et al., Visual object tracking using adaptive correlation filters. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13-18 June 2010
15. J.F. Henriques, R. Caseiro, et al., Exploiting the circulant structure of tracking-by-detection with kernels. Proceedings of the European conference on Computer Vision, Berlin, Heidelberg, October 2012
16. M. Danelljan, F.S. Khan, M. Felsberg, et al., Adaptive color attributes for real-time visual tracking. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Los Alamitos, CA, USA, June 2014
17. J.F. Henriques, R. Caseiro, P. Martins, et al., High-speed tracking with kernelized correlation filters. IEEE Trans. Pattern Anal. Mach. Intell. 37(3), 583–596 (2015)
18. D. Martin, H. Gustav, S.K. Fahad, F. Michael. Accurate scale estimation for robust visual tracking. Proceedings of the British Machine Vision Conference, Nottingham, 1-5 September 2014
19. Y. Li, J. Zhu. A scale adaptive kernel correlation filter tracker with feature integration. Proceedings of the European Conference on Computer Vision, September 2014
20. L. Yang, J. Zhu, S. Hoi, Reliable patch trackers: robust visual tracking by exploiting reliable patches. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015), pp. 353-361
21. A.S. Montero, J. Lang, R. Laganiere, Scalable kernel correlation filter with sparse feature integration. Proceedings of the IEEE International Conference on Computer Vision Workshop (2015), pp. 587-594
22. O. Akin, E. Erdem, A. Erdem, et al., Deformable part-based tracking by coupled global and local correlation filters. J. Vis. Commun. Image Represent. 38(C), 763–774 (2016)
23. A. Lukezic, L.C. Zajc, M. Kristan, Deformable parts correlation filters for robust visual tracking. Proceedings of the IEEE International Conference on Computer Vision Workshop (2015), pp. 587–594
24. T. Liu, G. Wang, Q. Yang. Real-time part-based visual tracking via adaptive correlation filters. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015), pp 4902-4912
25. M. Wang, Y. Liu, Z. Huang, Large margin object tracking with circulant feature maps. Proceedings of the International Conference on Computer Vision and Pattern Recognition (2017). pp. 4021-4029
26. C. Liu, P. Liu, W. Zhao, et al., Robust tracking and re-detection: collaboratively modeling the target and its context. IEEE Transactions on Multimedia. 20(4), 889–902 (2017)
27. N. Wang, W. Zhou, H. Li, Reliable re-detection for long-term tracking. IEEE Transactions on Circuits and Systems for Video Technology 29(3), 730–743 (2018)
28. U.D. Hanebeck, K. Briechle, Template matching using fast normalized cross correlation. Proceeding of Spie on Optical Pattern Recognition XII 4387, 95–102 (2001)
29. M. Mueller, N. Smith, B. Ghanem, A benchmark and simulator for UAV tracking. Far East Journal of Mathematical Sciences 2(2), 445–461 (2016)
30. Y. Wu, J. Lim, M. Yang, Online object tracking: a benchmark. Paper presented at IEEE Conference on Computer Vision and Pattern Recognition, Los Alamitos, CA, USA, June 2013

Publisher's Note