Design of bio-inspired binocular UAV detection system based on improved STC algorithm of scale transformation and occlusion detection

Huajun Song¹, Yanqi Wu¹ and Guangbing Zhou²

Abstract
With the rapid development of drones, many problems have arisen, such as invasion of privacy and endangering security. Inspired by biology, in order to achieve effective detection and robust tracking of small targets such as unmanned aerial vehicles, a binocular vision detection system is designed. The system is composed of long focus and wide-angle dual cameras, servo pan tilt, and dual processors for detecting and identifying targets. In view of the shortcomings of spatio-temporal context target tracking algorithm that cannot adapt to scale transformation and easy to track failure in complex scenes, the scale filter and loss criterion are introduced to make an improvement. Qualitative and quantitative experiments show that the designed system can adapt to the scale changes and partial occlusion conditions in the detection, and meets the real-time requirements. The hardware system and algorithm both have reference value for the application of anti-unmanned aerial vehicle systems.

Keywords
Anti-unmanned aerial vehicle, target detection, target tracking, spatio-temporal context tracking, bio-inspired

Introduction
In recent years, unmanned aerial vehicle (UAV) has been widely used in many fields, which has aroused great interest in military and civil fields. The military UAV undertakes the task of danger detection. The police and government’s UAV is used to do safety monitoring and environmental monitoring. Meanwhile, civil UAV is developing rapidly in the field of photography. UAVs have brought convenience in many aspects to our life. Under certain circumstances, UAVs violate personal privacy and affect normal social life. Important places need to restrict the entry of UAVs, such as airports and schools, etc. Therefore, it is necessary to locate and track the illegal and hidden UAVs in specific scenes. The mainstream detection methods include vision, radar, and electromagnetic wave. The vision system has the advantages of target visualization and active tracking, which can make up for the defect that radar cannot effectively identify the target and the defect of short electromagnetic detection range.

For the radar-based UAV positioning and tracking system, it relies on electromagnetic wave reception and reflection to determine the UAV’s position. It is less affected by light and can run in a dark environment. However, in a low-altitude environment, the electromagnetic signal emitted or reflected by the drone is easily submerged in the background noise, so it is difficult to distinguish. When the drone is far away from the radar detector, high hardware costs are required to achieve detection.

Meanwhile, for the computer vision-based system, the position of UAV is determined by visual sensors

¹College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, China
²Hangzhou Hikvision Digital Technology Co., Ltd, China

Corresponding author:
Huajun Song, College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China. Email: huajun.song@upc.edu.cn
(such as the camera), target detection algorithm, and target tracking algorithm. It has the advantages of the low cost of sensor and strong anti-jamming ability. However, due to the small size of the target, the accuracy is not enough in the case of detecting long-distance targets, and it is more vulnerable to the limitations of light and weather. In the actual application of UAV detection system, radar system and vision system always complement each other to meet the detection requirements in different scenarios.

The biological world often inspires industrial design. In recent years, the research of robot and intelligent system inspired by biological structure has been paid more and more attention. Wang et al. demonstrated the potential of wing-body interaction (WBI) in the design of flapping-wing micro aerial vehicle (MAV) that pursue higher performance. By analyzing the backward free flight of a dragonfly, Bode-Oke et al. found that wing-wing interaction could enhance the aerodynamic performance of the hindwings (HW) during backward flight. The system proposed in this paper is also inspired by the eyes of birds, and uses the focus of attention inspired by the biological visual system to model. Birds not only needs to look far in flight but also needs to see close scenery underwater. They can adjust the curvature of the lens in the eye to a great extent, which can have a larger field of view and can see distant objects clearly. To better solve the problem of small target detection, the system uses binocular cameras with a strong zoom ability and uses a servo system to simulate the head movement of birds for target tracking.

In recent years, more and more scholars have made great contributions in the field of UAV detection. Hoffmann et al. proposed a method to detect and track micro UAV by using multi-static radar NetRAD and combining time-domain signal with the micro-Doppler signal; Shi et al. integrated a variety of monitoring technologies to establish an anti-UAV system named ADS-ZJU to detect, locate, and radio frequency interfere with UAV; in the field of visual UAV, Wang et al. proposed a small flying target detection method based on Gaussian mixture background modeling in compressed sensing domain and low-rank sparse matrix decomposition of the local image, while Li et al. used the mobile camera to detect and track UAV through optical flow matching and Kalman filter. Dorudian et al. used an external RGB-D sensor, and added a blind update model to adapt to background sudden changes.

Frame difference (FD), Gaussian mixture model (GMM), and ViBe algorithms are classic moving target detection algorithms. Scholars have made different improvements over the years. Sengar and Mukhopadhyay combined FD with W4 algorithm. Zong et al. proposed a deep auto-encoded GMM for unsupervised anomaly detection. Zhou et al. combined depth information and color information for foreground segmentation to improve ViBe algorithms. In target tracking, STC has been widely used. Cao et al. presented a hierarchical features-based tracker for spatio-temporal context (STC) learning, which devotes to enhancing tracking performance by constructing more robust model and designing more useful feature representations. A real-time updated learning rate and a fading factor are introduced to improve the occlusion loss problem of the STC algorithm in Yang et al.'s work. Xue et al. proposed multi-scale spatio-temporal context learning tracking, which formulates a low-dimensional representation named fast perceptual hash algorithm to update long-term historical targets and the medium-term stable scene dynamically with image similarity.

The system proposed in the above articles has achieved good results in different specific situations. However, for long-distance dim and small targets, the detection and tracking effects are still not ideal. The main problem lies in the inability to distinguish UAVs, birds, kites, and other objects, and it is easy to be disturbed by the background environment. The main contributions of this paper are as follows: (1) detecting dim and small targets based on spatiotemporal continuity to improve detection efficiency; (2) adding a scale filter and introducing loss criterion to optimize the tracking performance in the case of scale change and targets occluded of STC algorithm. In the test results, we qualitatively and quantitatively show that the tracking method proposed in this paper is superior to the old model in some evaluation indexes.

The remainder of this paper is organized as follows. Firstly, the hardware and software architecture of the system are introduced in the next section. The subsequent section discusses the current moving target detection method and our improvement based on spatiotemporal context. Then the current target tracking algorithm and our improvement on scale filter and loss criterion is analyzed. The penultimate section discusses the analysis of experimental test results. Finally conclusions and future work directions are drawn in the last section.

System architecture

The system is composed of optical zoom visible light camera and servo automatic tracking pan-tilt, as shown in Figure 1. Two HIKVISION DS-2ZCN3008 optical zoom network cameras are installed side by side on the YAAN YS3081 servo pan-tilt. YAAN YS3081 can continuously rotate 360° in the horizontal direction and –80° to 80° in the vertical direction. The maximum
speed under a load of 26 kg is \(30^\circ\) per second in the horizontal direction and \(15^\circ\) per second in the vertical direction. The main camera uses short-focus and wide-angle mode, which makes it easy to monitor large aerial areas and detect moving targets; the slave camera uses long focus mode, which has a narrow field of view but can enlarge the distant scene to display and identify the moving target. The system is a general-purpose device, which can be installed on high-rise buildings with a wide view. Our work conducts system testing in the wide view area of the floor where the laboratory is located.

At the software level, firstly, the moving object detection algorithm is run to detect suspicious moving targets. After the suspicious target is found, servo motor controls the camera to move to the target, the target is moved to the center of the field of view. The improved spatiotemporal context object tracking algorithm is used to track the target. Then the slave camera is controlled to adjust the focal length to enlarge it. Finally, we run the deep learning algorithm to recognize the enlarged moving target, and determine whether to continue tracking according to the results.

**Moving object detection algorithm**

**Analysis of current algorithm**

The software architecture is shown in Figure 2. For the traditional moving target detection algorithm, there are mainly frame difference methods, background difference method, the optical flow method, etc. The frame difference method obtains the object contour by difference operation of the adjacent frames of the image sequence. The background difference method uses the difference between the reference background and the video sequence to detect the object. The optical flow method uses the change of the pixel points in the image sequence in the time domain to extract the object motion information. The detection effect of these methods for weak and small UAV target is relatively limited.\(^{23,24}\) The frame difference method cannot effectively distinguish between noise and moving targets. During opening and closing operations, the background noise is removed, while the weak and small targets are filtered out. The common background difference methods include Gaussian modeling and background modeling, which have a large amount of calculation and cannot meet the requirements of real-time detection. The optical flow method also has a large amount of calculation, which is difficult to meet the real-time performance. Inspired by insect’s neurons, Wang et al.\(^{25}\) proposed a small target motion detectors neural network for small target detection in a cluttered background. In the aspect of spatial-temporal, based on sparse representation and Bayesian inference, a robust target tracking algorithm is proposed by Li et al.\(^{26}\) which can estimate and predict the temporal and spatial structure of targets. According to the motion characteristics of UAV, a first in first out (FIFO) denoising structure and an improved method based on spatiotemporal context is proposed.

**Moving object detection algorithm based on spatiotemporal context**

There is no abrupt change in the position of two adjacent frames in the image sequence, that is, there is a continuity in time and a special spatial relationship between the target and the surrounding background. The combination of the time and spatial information forms the spatiotemporal context information. When biological vision system focuses on the target, it will focus on a specific area. By the vision system, more attention will be paid on the points which are close to the target. Inspired by biological vision system, this method is used to calculate the prior probability of tracking target in STC algorithm.\(^{27,28}\)

If there is a small target at \((x, y)\) in the k-th frame image \(f_k\), the weak target will appear in the small neighborhood of \((x, y)\) in the next image \(f_{k+1}\). The noise is often randomly distributed in the image, as shown in Figures 3 and 4. Based on the characteristics of small and weak moving targets, the short-term trajectory of the target is constructed to distinguish the target from the noise, so as to improve the accuracy of target detection.

In Figure 3, the left side is the real scene, and the right side is the image after binarization by frame difference method. White pixels may be targets or noises. The white rectangular box is the target area which contains UAVs. The pink rectangular box is a noise area, which contains pedestrians or shaking trees.
In Figure 4(a) and (b) are the enlarged images in the white and pink rectangular frames in Figure 3. Figure 4(a) shows that the position of the target in the target area changes in a short time and remains in the initial window. Figure 4(b) shows the short-term position change of noise points in the noise area.

The following is the steps of moving target detection. The overall flow of the moving object detection algorithm is shown in Figure 5.

Step 1, get the candidate target image $D_k$. Save the last 10 frames of candidate target image results ($D_{k-9}, \ldots, D_k$) to establish FIFO (first in first out) channel. The new frame $D_{k+1}$ is processed by frame difference operation, and the appropriate threshold is selected for binarization. The binarization result after difference is put into the FIFO channel. It may contain moving targets and noise points.

Step 2, the binarization result is accumulated and the appropriate threshold is set to remove the noise. The gray value images of the first frame image in the channel are selected as the candidate target points. A $10 \times 10$ window is established with each candidate target as the center, and the gray values images of each window position of all images in the channel are accumulated respectively. Because the isolated noise points are randomly distributed and the real target has space-time continuity, we set the threshold of gray value. If the gray value accumulated result exceeds the threshold, it will be considered that there is a target in the window, and the location of the target will be updated. Through this operation, most of the isolated noise points can be removed and the real target can be retained.

Step 3, get the moving target. The noise in the window can be removed by “dilating” the gray value of the window area in the FIFO channel. Then performing the “and” operation on the two adjacent differential images. If the target does not move in certain frames, the “and” operation of adjacent image frames will lead to the loss of the real target. In this case, the window with the pixel value accumulated to 0 is
removed first, and then the two adjacent frames are operated with “and”. If the result exceeds a certain threshold, it is regarded as the target. In terms of FIFO size selection, a small-size FIFO cannot buffer the target for a short period of time, and it is easy to cause the target to be lost. Choosing a large size fifo will cause a significant delay in the tracking frame when the target is truly lost. Through testing, this article selects a FIFO with a size of 10, which can achieve better detection results.

Improved STC tracking algorithm

Analysis of current target tracking algorithm

UAV imaging is small in the distance from the camera, and there are almost no texture features. At the same time, when the UAV moves from near to far or from far to near, the scale of the target changes significantly, which requires the tracking algorithm to have scale adaptability. Besides, UAV may be similar to the background color or be covered by the background during flight, so the tracking algorithm needs to consider the influence of complex background. The mainstream target tracking algorithms are compared in different scene datasets, including Staple (sum of template and pixel-wise learners29,30), STC, ECOHC (efficient convolutional network for online video understanding31), DSST (discriminative scale space tracking32), BACF (boot angle compensation filter33), CN (color names), CSK (exploiting the circulant structure of tracking-by-detection with kernels34), KCFv2 (kernel correlation filter35). All of them can approach the real-time detection (STC algorithm reaches 350 FPS in Intel Core i7 platform). The performance of the tracking algorithm is evaluated by the success rate curve and the area under curve (AUC) of the area surrounded by the coordinate axis.

The accuracy can accurately reflect the performance of the tracker. The success rate is a reflection of the overlap area between the output box of the tracking algorithm and the annotation box. Some tracking algorithms are sensitive to the selection of the initial tracking frame. Giving initial boxes in different initial frames will lead to different tracking results. Therefore, the initial target frame is changed in space and time to test its SRE (spatial robustness assessment) and TRE (temporal robustness assessment).

In the analysis of the success rate, we mainly consider the tracking success rate when the overlap rate threshold is above 0.4. In the accuracy analysis, the accuracy of the target center positioning error threshold below 20 is mainly considered. Run different tracking algorithms to track all sequences, including weak and small target sequences, scale transformation sequences, and complex scene sequences. Examples of tracking results are shown in Figure 6, and the results are analyzed as shown in Tables 1 and 2. The figures in the table represent the results, while the overlap rate threshold is 0.4 or the target center positioning error threshold is 20.

In Tables 1 and 2, generally, in the SRE and TRE test results of all sequences in the dataset, STC, Staple, and ECOHC algorithms have higher success rates. Among the accuracy test results of all sequences in the data set, the STC, Staple, BACF, and ECOHC algorithms have higher accuracy. In the subdivision dataset, the STC, Staple, ECOHC algorithm tracking
success rates and accuracy are relatively high in the weak and small target sequence test. Compared with other algorithms, the STC algorithm can track small targets stably and accurately compared to other algorithms. In the success rate and accuracy test of the scale conversion sequence, STC, Staple, and DSST algorithms have high success rates, and STC, CN, Staple and other algorithms have high accuracy. In the tracking success rate and accuracy test results of complex scene sequences, the accuracy and success rate of various tracking algorithms are generally low. The accuracy and precision of the BACF algorithm are higher than other algorithms. The STC algorithm still updates the model when the target is occluded, which leads to model deviation and tracking error.

In the UAV detection scenario, first is to deal with the change of the target from far to near. The algorithm should have high accuracy for the weak and small target and scale change scene. Secondly, because the above tracking algorithms cannot track accurately in complex scenes, such as target occlusion, the tracking algorithm is required to have the retrieval ability. At the same time, when it is judged that the target cannot be retrieved for a long time, re-running the target detection algorithm is used to deal with it. Therefore, this paper selects STC algorithm which has outstanding tracking effect on weak and small targets and can still track stably under the condition of target scale change. Meanwhile, the scale adaptation of STC algorithm and tracking performance in complex environment will be improved.

**Improved STC tracking algorithm based on scale filter and loss criterion**

Aiming at the problem that the STC algorithm cannot effectively adapt to scale transformation, inspired by the idea of the DSST target tracking algorithm, an improved scale filter is added to the original STC position filter to replace the simple scale adaptation method in STC algorithm. Assuming that the size of the current target in the image is $P \times R$, with the target as the center, $S$ target image blocks are intercepted. The size of the image block is

$$a^n P \times a^n R, n = \left\lfloor \frac{S - 1}{2} \right\rfloor, \ldots, \left\lfloor \frac{S - 1}{2} \right\rfloor$$

In the above formula, $a$ is the scale factor, set as 1.02, and the scale number $S$ is 33, so as to find the optimal solution of the image block adapted to the target. For 33 image blocks, the feature $f$ is obtained by solving the feature descriptors respectively, and the

| Algorithm | Success rate of SRE | Precision of SRE |
|-----------|---------------------|------------------|
|           | Dim | Multi-scale | Complex | All  | Dim | Multi-scale | Complex | All  |
| Staple    | 0.533 | 0.467 | 0.320 | 0.443 | 0.846 | 0.993 | 0.527 | 0.809 |
| STC       | 0.574 | 0.380 | 0.184 | 0.380 | 0.988 | 1.000 | 0.413 | 0.814 |
| ECOHC     | 0.432 | 0.453 | 0.277 | 0.394 | 0.829 | 1.000 | 0.339 | 0.750 |
| DSST      | 0.265 | 0.433 | 0.234 | 0.323 | 0.408 | 0.950 | 0.406 | 0.624 |
| BACF      | 0.406 | 0.174 | 0.414 | 0.315 | 0.685 | 0.918 | 0.709 | 0.788 |
| CN        | 0.184 | 0.342 | 0.230 | 0.261 | 0.402 | 1.000 | 0.429 | 0.649 |
| KCFv2     | 0.085 | 0.155 | 0.160 | 0.129 | 0.470 | 0.329 | 0.305 | 0.364 |
| CSK       | 0.056 | 0.201 | 0.211 | 0.160 | 0.086 | 0.482 | 0.371 | 0.330 |

| Table 1. Success rate's and precision's evaluation results of SRE. |

| Algorithm | Success rate of TRE | Precision of TRE |
|-----------|---------------------|------------------|
|           | Dim | Multi-scale | Complex | All  | Dim | Multi-scale | Complex | All  |
| Staple    | 0.747 | 0.664 | 0.484 | 0.635 | 0.842 | 0.995 | 0.761 | 0.879 |
| STC       | 0.807 | 0.568 | 0.343 | 0.588 | 0.999 | 0.988 | 0.683 | 0.900 |
| ECOHC     | 0.683 | 0.502 | 0.573 | 0.577 | 0.968 | 0.857 | 0.718 | 0.849 |
| DSST      | 0.413 | 0.623 | 0.434 | 0.503 | 0.487 | 0.965 | 0.707 | 0.744 |
| BACF      | 0.502 | 0.326 | 0.648 | 0.476 | 0.787 | 0.901 | 0.843 | 0.849 |
| CN        | 0.339 | 0.596 | 0.392 | 0.458 | 0.422 | 1.000 | 0.713 | 0.740 |
| KCFv2     | 0.193 | 0.273 | 0.408 | 0.290 | 0.360 | 0.496 | 0.728 | 0.525 |
| CSK       | 0.164 | 0.338 | 0.377 | 0.298 | 0.239 | 0.525 | 0.662 | 0.480 |

| Table 2. Success rate's and precision's evaluation results of TRE. |
response output $g$ is constructed as a vector with a dimension of $1 \times S$. The middle value of the vector is the largest and gradually decreases toward both ends. $h$ is the scale filter template, and the minimum loss function is as follows:

$$
\varepsilon = \left\| \sum_{l=1}^{d} H_l \ast h^l - g \right\|^2 + \lambda \sum_{l=1}^{d} \|H_l\|^2
$$

where $f^l$ is the $l$-th dimension of feature $l$, $h^l$ is the $l$-th dimension of feature $h$, $l \in \{1, \ldots, d\}$, and lambda is a regularization term.

Perform DFT transformation on each dimension of the feature to obtain $F^l$, and perform DFT transformation on the output response $g$ to obtain $G$. Next, solve the scale filter template $H_{scale}$ of each dimension after the transformation:

$$
H_{scale}^{l+1} = \begin{cases} 
(1 - \rho) H_{scale}^l + \rho h_r, & \text{if } \max(c_l(x)) > 0 \text{ and } APCE > T \\
H_{scale}^l, & \text{if } \max(c_l(x)) < 0 \text{ and } APCE < T
\end{cases}
$$

where $h_r$ is the spatial context model, $H_{scale}^l$ is the spatio-temporal context model, $\rho$ is the learning rate, $APCE$ is the loss criterion, $T$ is the threshold of the set loss criterion, and $\max(c_l(x))$ is the maximum value of the confidence map. $APCE$ is defined as:

$$
APCE = \frac{|F_{max} - F_{min}|^2}{\text{mean} \left( \sum_{w,k} (F_{w,h} - F_{min})^2 \right)}
$$

where $F_{max}$, $F_{min}$ and $F_{WH}$ represent the maximum value, minimum value and the value of $h$ col and $w$ row in the confidence map respectively. At this time, if the $APCE$ is occluded, the tracking of multiple targets will be accurately reflected if the $APCE$ is occluded. When the $APCE$ value is higher than the threshold value, the position filter model is updated; when the $APCE$ value is lower than the threshold value, it is judged that the target is blocked or lost, then the model update is stopped to prevent tracking errors.

As shown in Figure 8, when the target is occluded (about frame 70), the $APCE$ value of the sequence decreases significantly. $APCE$ is added to the criterion. When the $APCE$ value is lower than the threshold value, the model updating is stopped.

**Analysis of test results**

*Analysis of moving target detection results based on spatio-temporal continuity*

The frame difference method (FD), Gaussian background modeling (GMM), vibe algorithm (VIBE),
and the target detection method based on the Spatio-temporal context (OURS) proposed in this paper are tested on 6831 image sequences dataset, which contains weak small UAV targets. The detection results are shown in Figure 9.

The first line of Figure 9 is the detection result diagram of frame difference method, and the blue box is the detection result box. Due to the slight camera jitter, the results of frame 60, 150, and 300 all contain the noise caused by leaf shaking, and there is a large area of false detection around the 150th frame, which shows that the original frame difference method cannot effectively distinguish leaf shaking and real moving objects. Figure 9 shows the GMM background modeling result in the second line, and the green box is the detection result box. Because the background model is not stable, GMM algorithm has a large number of false detection in about 60 frames. The third line is the detection result of vibe algorithm, and the pink box is the detection result box. Due to the need to model each pixel, the speed of vibe algorithm cannot meet the real-time requirements, and the probability of missed detection is high.

The fourth line of Figure 9 shows the results image of the proposed target detection algorithm based on spatiotemporal continuity proposed in this paper. The red box is the detection result box. The results show that the proposed algorithm can effectively deal with random noise, and can accurately detect UAV and pedestrian targets.

The precision is defined as the ratio of the number of correct targets detected and the number of all targets detected. The recall is defined as the ratio of the number of correct targets detected and the number of real targets. Data analysis is shown in Figure 10.
STC, STC with scale variation (STCSCALE), and the improved STC algorithm proposed in this paper are used to track and test the UAV data set sequence. The test results are as follows:

In Tables 3 and 4, compare the SRE and TRE tests of the proposed algorithm on the weak and small target data set and the scale change data set. Compared with the classic STC algorithm, the proposed algorithm has a slight improvement in the success rate of weak and small target tracking, but it has a significant improvement in the calculation results on the scale change data set sequence.

![Figure 10. Quantitative analysis of result of detection.](image)

**Table 3.** SRE and TRE test result in dim target datasets.

| Algorithm | Success rate | Precision |
|-----------|--------------|-----------|
|           | SRE  | TRE  | SRE  | TRE  |
| STC       | 0.574 | 0.807 | 0.968 | 0.999 |
| OURS      | 0.593 | 0.773 | 0.968 | 0.999 |

**Table 4.** SRE and TRE test result in scale change datasets.

| Algorithm | Success rate | Precision |
|-----------|--------------|-----------|
|           | SRE  | TRE  | SRE  | TRE  |
| STC       | 0.236 | 0.448 | 0.802 | 0.987 |
| STC_Scale | 0.328 | 0.568 | 0.805 | 0.995 |
| OURS      | 0.515 | 0.586 | 0.976 | 0.988 |

**Figure 11.** Examples of tracking results of STC, STCSCALE and OURS.

**Analysis of improved STC tracking algorithm test result**

STC, STC with scale variation (STCSCALE), and the improved STC algorithm proposed in this paper are used to track and test the UAV data set sequence. The test results are as follows:

In Tables 3 and 4, compare the SRE and TRE tests of the proposed algorithm on the weak and small target data set and the scale change data set. Compared with the classic STC algorithm, the proposed algorithm has a slight improvement in the success rate of weak and small target tracking, but it has a significant improvement in the calculation results on the scale change data set sequence.

Figure 11(a) shows the tracking test result of small and dim target; Figure 11(b) shows the tracking test result of scale transformation sequence 1; Figure 11(c) shows the tracking test result of scale transformation sequence 2. The red box represents the STC algorithm, the green box represents STCSCALE, and the blue box represents the improved STC algorithm proposed in this article.

Besides, the low computational complexity is one prime characteristic of the STC algorithm in which only six FFT operations are involved for processing one frame. For the local context region of $M \times N$ pixels, the time complexity of each FFT calculation is only $O(MN\log_2(MN))^{28}$ For the APCE criteria and proposed FIFO moving object detection algorithm, since there are no high-order calculations and loop operations, the time complexity is $O(1)$, which has a low impact on the time complexity of the entire algorithm.

In Table 5, the calculation speed of the proposed algorithm is compared with the classical algorithms and the recent algorithms. The algorithms are run on the Intel Core-i7 3.4 GHz platform (GTX1050ti is used in recognition, not in tracking). When they are not integrated into the entire test UAV monitoring system, the proposed algorithm has a good performance on processing time. After being integrated into the system, the proposed algorithm can still meet real-time requirements.

**Analysis of dataset**

YOLOv4 is trained to recognize the moving target and choose an initial target. Eleven video sequences in the anti-drone scene were shot and collected, totaling 14,705 pictures, and the true value was marked, that is, the position coordinates of the drone in each picture were recorded. At the same time, in order to evaluate the tracking performance of different algorithms under
different environmental factors, and comprehensively select real-time fast algorithms suitable for anti-UAV scenarios, our work refers to Professor Wu Yi’s Benchmark to add weak targets, scale changes, and complexity to the video sequence. Several attributes such as scenes facilitate the specific analysis of the adaptability of the tracking algorithm to different scenes. The meaning of each attribute is as follows:

Normal target: Image sequences with target size greater than 100 pixels;

Weak targets: Image sequences whose target size is less than 100 pixels;

Scale change: There are image sequences in which the ratio of the target frame scales of the two images exceeds the interval $[1/3, 3]$;

Complex background: The target is completely occluded or partially occluded in the scene, or is similar to the background color;

The examples of our dataset are shown in Figure 12. The composition of our dataset is shown in Table 6.

### Conclusions

A set of binocular vision system for tracking UAV is designed and built. The binocular camera is inspired by biology and modeled by biological vision focusing, which has good real-time and practicability. The detection efficiency is improved by detecting dim and small targets based on spatiotemporal continuity. In this paper, the comprehensive performance of the mainstream tracking algorithm is tested, and the evaluation based on different quality indicators is given. The STC algorithm, which has poor performance in scale adaptation, is improved. Inspired by DSST, a scale filter is added and a loss criterion is introduced to optimize the tracking performance of STC for scale transformed and occluded targets. The improved STC algorithm can achieve a rate of 58fps. On these datasets, we qualitatively and quantitatively show that the tracking method proposed in this paper is superior to the old model in some evaluation indexes.

At present, some work still needs to be improved. First of all, the current use of dual visible light cameras, one is used to detect and track weak and small targets, the other is used to zoom the near-sighted field to observe the target, which not only ensures the tracking robustness but also causes a waste of resources. It is a better scheme to use only one camera to complete target detection, recognition, and tracking. Secondly, although the improved STC algorithm has a significant effect on tracking dim and small targets and scale transformation scenes, the tracking accuracy still needs to be improved in complex scenes. Due to the strong adaptability of BACF to complex scenes, the state machine tracking algorithm combined with BACF is the future research direction.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this
article: This work was supported by The Fundamental Research Funds for the Central Universities (NO.18CX02109A); Science and Technology on Electronic Test & Measurement Laboratory (6142001180514).

**ORCID iD**

Yanqi Wu [https://orcid.org/0000-0001-6830-1297](https://orcid.org/0000-0001-6830-1297)

**References**

1. Sheu BH, Chiu CC, Lu WT, et al. Development of UAV tracking and coordinate detection method using a dual-axis rotary platform for an anti-UAV system. *Applied Sciences*, 2019; 9: 2583.

2. Patel JS, Fioranelli F and Anderson D. Review of radar classification and RCS characterisation techniques for small UAVs or drones. *IET Radar Sonar Navig* 2018; 12: 911–919.

3. Rozantsev A, Lepetit V and Fua P. Detecting flying objects using a single moving camera. *IEEE Transac Pattern Anal Mach Intell* 2016; 39: 879–892.

4. Wang C, Wang T, Wang E, et al. Flying small target detection for anti-UAV based on a Gaussian mixture model in a compressive sensing domain. *Sensors* 2019; 19: 2168.

5. Wang J, Ren Y, Li C, et al. Computational investigation of wing-body interaction and its lift enhancement effect in hummingbird forward flight. *Bioinspirat Biomimet* 2019; 14: 046010.

6. Bochkovskiy A, Wang CY and Liao HYM. Yolov4: Optimal speed and accuracy of object detection *arXiv preprint arXiv* 2020; 2004.10934.

7. Bode-Oke AT, Zeyghami S and Dong H. Flying in reverse: kinematics and aerodynamics of a dragonfly in backward free flight. *J R Soc Interf* 2018; 15: 20180102.

8. Katzir G and Howland HC. Corneal power and under-water accommodation in great cormorants (phalacrocorax carbo sinensis). *J Exp Biol* 2003; 206: 833–841.

9. Hoffmann F, Ritchie M, Fioranelli F, et al. Micro-Doppler based detection and tracking of UAVs with multistatic radar. In: *2016 IEEE radar conference (RadarConf)*, Philadelphia, PA, USA, 02-06 May 2016, pp. 1–6. Piscataway, NJ: IEEE.

10. Shi X, Yang C, Xie W, et al. Anti-drone system with multiple surveillance technologies: architecture, implementation, and challenges. *IEEE Commun Mag* 2018; 56: 68–74.

11. Li J, Ye DH, Chung T, et al. Multi-target detection and tracking from a single camera in unmanned aerial vehicles (UAVs). In: *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Daejeon, 9-14 October 2016, pp. 4992–4997. Piscataway, NJ: IEEE.

12. Dorudian N, Lauria S and Swift S. Nonparametric background modelling and segmentation to detect micro air vehicles using RGB-D sensor. *Int J Micro Air Vehic* 2019; 11: 1–20.

13. Singla N. Motion detection based on frame difference method. *Int J Inform Comput Technol* 2014; 4: 1559–1565.

14. Reynolds DA. Gaussian mixture models. *Encycop Biometrics* 2009; 741: 659–663.

15. Sengar SS and Mukhopadhyay S. Moving object detection based on frame difference and w4. *Signal Image Video Process* 2017; 11: 1357–1364.

16. Karton A, Rabinovich E, Martin JM, et al. W4 theory for computational thermochemistry: in pursuit of confident sub-KJ/mol predictions. *J Chem Phys* 2006; 125: 144108.

17. Zong B, Song Q, Min MR, et al. Deep autoencoding Gaussian mixture model for unsupervised anomaly detection. In: *International conference on learning representations*, Vancouver, BC, 30 April – 3 May, 2018.

18. Zhou X, Liu X, Jiang A, et al. Improving video segmentation by fusing depth cues and the visual background extractor (vibe) algorithm. *Sensors* 2017; 17: 1177.

19. Cao Y, Ji H, Zhang W, et al. Learning spatio-temporal context via hierarchical features for visual tracking. *Signal Process* 2018; 66: 50–65.

20. Yang X, Zhu S, Zhou D, et al. An improved target tracking algorithm based on spatio-temporal context under occlusions. *Multidimensional Syst Signal Process* 2020; 31: 329–344.

21. Wu Y, Lim J and Yang MH. Online object tracking: a benchmark. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Portland, Oregon, USA, 23–28 June, 2013, pp. 2411–2418. IEEE.

22. Xue W, Xu C and Feng Z. Robust visual tracking via multi-scale spatio-temporal context learning. *IEEE Trans Circuits Syst Video Technol* 2017; 28: 2849–2860.

23. Dosovitskiy A, Fischer P, Ilg E, et al. Flownet: learning optical flow with convolutional networks. In: *Proceedings of the IEEE international conference on computer vision*, Santiago, Chile, 7–13 December 2015, pp. 2758–2766. IEEE.

24. Piccardi M. Background subtraction techniques: a review. In: *2004 IEEE international conference on systems, man and cybernetics (IEEE Cat. No. 04CH37583)*, Vol. 4, The Hague, Netherlands, 10-13 October 2004, pp. 3099–3104. Piscataway, NJ: IEEE.

25. Wang H, Peng J and Yue S. A directionally selective small target motion detecting visual neural network in cluttered backgrounds. *IEEE Transac Cybernet* 2018; 50: 1541–1555.

26. Li Z, Chen C, Liu D, et al. Infrared small target tracking algorithm based on temporal-spatial structure sparse Bayesian estimation. *Infrared Phys Technol* 2020; 105: 103160.

27. Torralba A. Contextual priming for object detection. *Int J Comput Vis* 2003; 53: 169–191.

28. Zhang K, Zhang L, Liu Q, et al. Fast visual tracking via dense spatio-temporal context learning. In: *European conference on computer vision*, Zurich, Switzerland, 6-12 September 2014, pp. 127–141. Berlin: Springer.

29. Barnich O and Van Droogenbroeck M. Vibe: a universal background subtraction algorithm for video sequences. *IEEE Trans Image Proc* 2010; 20: 1709–1724.

30. Bertinetto L, Valmadre J, Golodetz S, et al. Staple: complementary learners for real-time tracking. In:
Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, 27-30 June 2016, pp. 1401–1409. IEEE.

31. Zolfaghari M, Singh K and Brox T. Eco: efficient convolutional network for online video understanding. In: Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018, pp. 695–712.

32. Danelljan M, Häger G, Khan FS, et al. Discriminative scale space tracking. IEEE Trans Pattern Anal Mach Intell 2016; 39: 1561–1575.

33. Kiani Galoogahi H, Fagg A and Lucey S. Learning background-aware correlation filters for visual tracking. In: Proceedings of the IEEE international conference on computer vision, Venice, Italy, 22-29 October 2017, pp. 1135–1143.

34. Henriques JF, Caseiro R, Martins P, et al. Exploiting the circulant structure of tracking-by-detection with kernels. In: European conference on computer vision, Firenze, Italy, 7–13 October 2012, pp. 702–715. Berlin: Springer.

35. Li Y and Zhu J. A scale adaptive kernel correlation filter tracker with feature integration. In: European conference on computer vision, Zurich, Switzerland, 6-12 September 2014, pp. 254–265. Berlin: Springer.

36. Danelljan M, Häger G, Khan F, et al. Accurate scale estimation for robust visual tracking. In: British machine vision conference, Nottingham, 1–5 September 2014. BMVA Press.

37. Ma H, Acton ST and Lin Z. Situp: scale invariant tracking using average peak-to-correlation energy. IEEE Trans Image Process 2020; 29: 3546–3557.

38. Wang M, Liu Y and Huang Z. Large margin object tracking with circulant feature maps. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, 21-26 July 2017, pp. 4021–4029.

39. Yu T, Mo B, Liu F, et al. Robust thermal infrared object tracking with continuous correlation filters and adaptive feature fusion. Infrared Phys Technol 2019; 98: 69–81.