electronics

MDPI

Article

Strong Interference UAV Motion Target Tracking Based on Target Consistency Algorithm

Li Tan [1,2,*], Xiaokai Huang [1], Xinyue Lv [1], Xujie Jiang [1] and He Liu [3]

[1] School of Computer Science and Engineering, Beijing Technology and Business University, Beijing 100048, China; huangxk@st.btbu.edu.cn (X.H.); lvxy@nercita.org.cn (X.L.); 2130072019@st.btbu.edu.cn (X.J.)
[2] Chongqing Institute of Microelectronics Industry Technology, University of Electronic Science and Technology of China, Chongqing 400031, China
[3] Chongqing Academy of Education Science, Chongqing 400015, China; 20221401016@stu.cqu.edu.cn
* Correspondence: tanli@th.btbu.edu.cn

Abstract: In recent years, unmanned aerial vehicle (UAV) image target tracking technology, which obtains motion parameters of moving targets and achieves a behavioral understanding of moving targets by identifying, detecting and tracking moving targets in UAV images, has been widely used in urban safety fields such as accident rescue, traffic monitoring and personnel detection. Due to the problems of complex backgrounds, small scale and a high density of targets, as well as mutual occlusion among targets in UAV images, this leads to inaccurate results of single object tracking (SOT). To solve the problem of tracking target loss caused by inaccurate tracking results, this paper proposes a strong interference motion target tracking method based on the target consistency algorithm for SOT based on an interframe fusion and trajectory confidence mechanism, fusing previous frames for the tracking trajectory correction of current frames, learning again from previous frames to update the model and adjusting the tracking trajectory according to the tracking duration. The experimental results can show that the accuracy of the proposed method in this paper is improved by 6.3% and the accuracy is improved by 2.6% compared with the benchmark method, which is more suitable for applications in the case of background clutter, camera motion and viewpoint change.

Keywords: UAVs; target tracking; interframe fusion; trajectory confidence

Citation: Tan, L.; Huang, X.; Lv, X.; Jiang, X.; Liu, H. Strong Interference UAV Motion Target Tracking Based on Target Consistency Algorithm. *Electronics* **2023**, *12*, 1773. https://doi.org/10.3390/electronics12081773

Academic Editor: Mahmut Reyhanoglu

Received: 6 March 2023
Revised: 31 March 2023
Accepted: 5 April 2023
Published: 8 April 2023

1. Introduction

In recent years, unmanned aerial vehicles (UAVs) have developed rapidly. Due to their small size, low cost and high mobility, UAVs are widely used in exploration, rescue, traffic monitoring, personnel detection and other urban safety fields [1]. For special missions such as disaster rescue, urban patrol and anti-terrorist investigation, UAVs are usually used to accomplish tasks due to the complexity of the environment and mission scenarios [2]; thus, they have great potential for application in the field of urban security. Meanwhile, they play an important role in emergency rescue work in several security fields such as emergency mapping, environmental monitoring, earthquake relief, etc., where UAVs play an important role due to their flexibility, remote operation and powerful scalability [3,4].

With the continuous development of computer vision technology, target tracking in complex scenes with UAVs has gradually become a challenging research direction and focus, attracting many experts and scholars to conduct in-depth research and exploration and promoting the rapid development and wide application of UAV target tracking technology on the basis of deep learning [5].

Currently, target tracking methods can be divided into three categories: correlation filter-based target tracking, multi-feature fusion-based target tracking and deep learning method-based target tracking. Among them, the correlation filter-based target tracking

method proposes a filtering template for performing operations on candidate target regions. The target position of the current frame is the position of its maximum output response. Correlation filtering-based target tracking methods are suitable for real-time applications, especially on embedded systems with limited computational resources. For example, Qin et al. [6] constructed a target tracking model based on Kalman filtering and the Camshift method of multi-feature fusion, which can effectively improve the tracking effectiveness. Zhang et al. [7] introduced the background information of a target's neighborhood into the similarity measurement between the target and the candidate, proposed a scale estimation mechanism that relies only on the Hellinger distance mean shift process and detected the size estimation to reduce the effect of background clutter.

To address the limitations of single features, researchers have investigated ways to improve the performance of target tracking by fusing different features. The Staple [8] algorithm uses a combination of the global color histogram and histogram of oriented gradient (HOG) methods to describe the target. First, in the global color template, the motion foreground and static background are further computed based on their pre-estimated positions, and then the score of each pixel is obtained to derive the color response map. Then, in the HOG template, the HOG features are extracted from the previously determined target regions, and, thus, the dense response template is obtained. Finally, the scores of the two templates are linearly combined and the location of the target is finally estimated. The spatially regularized discriminative correlation filters (SRDCF) [9] tracking algorithm uses spatially regularized components to address boundary effects, employing regularized weights to penalize the filter coefficients during training and generate a more discriminative model. The temporal regularized correlation filters (STRCF) [10] tracking algorithm introduces temporal regularization into the SRDCF tracking algorithm. The background-aware correlation filters (BACF) [11] tracking algorithm dynamically models the foreground and background of the target using HOG features, while an alternating direction method of multipliers (ADMM) [12] optimization method is designed to solve the filter. The aberrance repressed correlation filters (ARCF) [13] tracking algorithm suppresses the rate of change of the response map that occurs at the time of detection and then suppresses the aberration of the response map in cases such as target occlusion, as a way to improve the tracking accuracy.

The purpose of a target tracking method based on deep learning is to optimize the distance metric between detections. Due to the consideration of a variety of influencing factors that are not beneficial for target tracking, such as the generally low resolution of aerial UAV videos, more interfering targets, and faster viewpoint transformation, Bi et al. [14] proposed a context-based target tracking method for aerial UAV videos. The effect of regression is improved by connecting multiple convolutional layers with a residual module, which can effectively improve the tracking effect of the algorithm. Zha et al. [15] added the semantic space sub-module to the twin network-based model as an adaptation to track the target captured by the UAV in the middle of the temporal space, which can solve the problems of target occlusion and target disappearance and improve the accuracy of target tracking.

In summary, the general step of target tracking is to estimate the trajectory model (including position, the direction of motion, shape, etc.) of the tracked target in each scene of the captured video, and a powerful tracker can then assign consistent markers to the target object in successive scenes. Therefore, visual tracking is an operation designed to locate, detect and define a dynamic configuration of one or more targets in a video sequence from one or more cameras. With the rapid development of UAVs and the rapid increase in video material from aerial UAV photography, single target tracking for aerial UAV video is one of the key problems studied by scholars, which can provide fundamental support for practical applications in related UAV fields.

Therefore, to address the problem of inaccurate target tracking results in aerial UAV video due to complex backgrounds, a high density of small-scale targets and mutual occlusion between targets, this paper proposes a strong interference motion target tracking

method based on the target consistency algorithm for UAVs. The main contributions of the method are as follows:

(1) The interframe fusion method is introduced in the model to correct the model's tracking trajectory of the target by fusing the current frame with the previous frames, and to update the model's tracking trajectory by combining the tracking results of the previous frames and learning them again.
(2) The model introduces a trajectory confidence mechanism, which defines the confidence level of the trajectory according to the duration of the tracked trajectory, and corrects and updates the trajectory in multiple directions to ensure the accuracy of the tracking results.
(3) The model optimizes the objective function using the ADMM algorithm and solves the function by iteration to obtain the optimal tracking trajectory.

The remainder of this paper is organized as follows. Section 2 presents the related work. Section 3 details the proposed method. The experiments and results analyses are provided in Section 4 while introducing the selected dataset and evaluation indicators. Finally, conclusions are drawn in Section 5.

2. Related Works

At present, scholars related to target tracking methods for aerial UAV video have conducted in-depth research and exploration, and many excellent results have been achieved; the details of some UAV target tracking algorithms are shown in Table 1. Among them, Liu et al. [16] constructed a target tracker TLD-KCF based on a conditional scale adaptive algorithm for aerial UAV video, and this method improved the tracking capability of quadrotor UAVs in complex outdoor scenes. Li et al. [17] designed a multi-vehicle tracking method for UAVs by combining SOT-based forward position prediction with results from intersection over union tracker (IOUT), which enhanced the detection results of the association phase. Chu et al. [18] used the results of target detection as SOT results and designed a multiple object tracking (MOT) network using multiple target interactions, which had a significant improvement in the performance of the MOT. Since a large number of targets overlap and obscure each other when performing UAV image multi-object tracking, this leads to identity-switch problems between targets and affects the performance of the algorithm. Feng et al. [19] used an SOT tracker and a reidentification network of the siamese region proposal network (SiameseRPN) [20] to extract short-term and long-term cues of targets, respectively. Then, a data association method with switcher-aware classification was used to improve the tracking results of the network while solving the identity-switch problem. However, in this method, the mutual independence of the SOT tracker and data association prevented the modules from collaborating well in the algorithm. For this reason, Zhu et al. [21] proposed a dual matching attention network to integrate single object tracking and data association into a unified framework to deal with intra-class interference and frequent interactions between targets. Wan et al. [22] designed a target tracking method based on sparse representation theory for aerial drone videos to solve the problem of partial occlusion between objects present in aerial drone videos that are used to localize the objects captured by UAVs, which contain pedestrians, vehicles, etc.

Table 1. Details of some UAV target tracking algorithms.

Proposer(s)	Dataset	Description	Characteristic
Liu et al. [16]	VOT 2014	A target tracker TLD-KCF based on a conditional scale adaptive algorithm for aerial UAV video.	Improved tracking capability of quadcopter UAVs in complex outdoor scenarios.
Li et al. [17]	UA-DETRAC	A multi-vehicle tracking method for UAV.	Combining SOT-based forward position prediction with results from IOUT enhances detection results in the association stage.

Table 1. *Cont.*

Proposer(s)	Dataset	Description	Characteristic
Chu et al. [18]	MOT15 MOT16	A CNN-based framework for online MOT.	The MOT network is designed through the interaction of multiple SOT results, so that the performance of MOT is significantly improved.
Feng et al. [19]	MOT16 MOT17	A unified MOT learning framework.	Makes full use of both long-term and short-term cues to deal with the complexities of MOT scenes, and considers potential identity-switch through switcher-aware classification.
Zhu et al. [21]	MOT16 MOT17	An MOT with Dual Matching attention networks.	Integrates the merits of single object tracking and data association methods in a unified framework to handle noisy detections and frequent interactions between targets.
Wan et al. [22]	VOT2015	A target tracking method for aerial UAV video based on sparse representation theory.	Solves the problem of partial occlusion between targets in aerial drone video.
Liu et al. [23]	MDMT	A multi-match authentication network MIA-net for multi-target tracking missions with multiple UAVs.	Solves cross-UAV association problems by constructing cross-UAV target topology relationships through local–global matching algorithms.
Yeom [24]	Practical scenarios	A long-range ground target tracking algorithm for small UAVs.	Selects the most suitable trajectory from multiple trajectories in a dense trajectory environment using nearest neighbor association rules.
Jiang et al. [25]	The 1st Anti-UAV Workshop and Challenge	An improved YOLOv5 UAV detection and tracking algorithm.	High-speed tracking performance by training low-resolution detectors combined with Kalman algorithms.
Lin et al. [26]	VisDrone2019	An improved UAV multi-target tracking model based on FairMOT.	Improves model tracking performance by sorting out temporal correlation structures and separating different functional heads.
Bhagat et al. [27]	Simulation experiments	A DQN-based persistent target tracking model for urban environments.	Enables UAVs to continuously track targets in different environments while avoiding obstacles in the environment.
Yang et al. [28]	ImageNet	A novel framework for hierarchical deep learning task assignments.	Performs tasks that require intensive computing with mobile edge computing servers that are rich in computing resources.
Bi et al. [14]	UAV123	A context-based remote sensing target tracking method for aerial UAV video MDnet.	Introduction of the RA-CACF module into the online tracking phase of the tracking network.
He et al. [29]	Visdrone-mot2020	A method for tracking different classes of multiple targets in different scenarios COFE model	Includes three main modules: multi-class target detection, coarse-class multi-target tracking and fine-grained trajectory refinement.

Liu et al. [23] proposed a multi-matching identity authentication network (MIA-Net) for a multi-target tracking task with multiple UAVs. The MIA-Net effectively solved the cross-UAV association problem by constructing cross-UAV target topology relationships through a local–global matching algorithm, and effectively complemented the obscured targets by taking advantage of multiple UAV viewpoint mapping. Yeom [24] studied ground target tracking algorithms at long distances (up to 1 km) using small UAVs and improved the association between trajectories by selecting the most suitable of multiple trajectories in a dense trajectory environment using nearest neighbor association rules. The detection of moving targets in the algorithm also includes frame-to-frame subtraction and thresholding, morphological operations and false alarm elimination based on object size and shape property, and the target's trajectory is initialized by the difference between the two nearest points in consecutive frames; then, the measurement statistically nearest to the state prediction updates the target's state. Jiang et al. [25] proposed an improved YOLOv5 UAV detection algorithm and tracking method to address the difficulties of poor imaging contrast, complex background and small target scale. The method improved UAV detection probability by adding a detection head and attention module, and achieved high-speed

tracking performance by training low-resolution detectors combined with the Kalman algorithm. Lin et al. [26] proposed an improved UAV multi-target tracking model based on FairMOT. The model contains a structure that separates the detection head and the ReID head to reduce the influence between each functional head. In addition, they developed a temporal embedding structure to enhance the characterization capability of the model. By combing the temporal association structure and separating the different functional heads, the performance of the model in UAV tracking tasks is improved.

Bhagat et al. [27] proposed a deep learning technique based on target-tracking DQN networks for persistent target tracking in urban environments. After experiments, it was shown that the algorithm enabled UAVs to persistently track targets in different environments while avoiding obstacles in both the training environment and the unseen environment. Since UAVs are generally severely limited in power supply and have a low computational power to perform tasks requiring intensive computation on their own, this poses a great challenge in terms of computational power, low latency and inference accuracy. Based on the above reasons, Yang et al. [28] proposed a novel hierarchical deep learning task assignment framework in which UAVs are embedded in the lower layers of the pre-trained CNN models, while mobile edge computing servers with abundant computational resources handle the higher layers of the CNN models; the effectiveness of the proposed offloading framework was demonstrated after experimental results. Bi et al. [14] proposed a context-based remote sensing target tracking method MDnet for aerial UAV video. In the network structure, residual connections are applied to fuse multiple convolutions, thus improving the network representation of remote sensing targets. In the pre-training phase, an enhancement strategy of rotating an adversarial autoencoder is used to generate enough negative samples to enhance the ability to distinguish between targets and background interference. In the online tracking phase, the RA-CACF module is introduced into the tracking network for remote sensing target tracking in aerial UAV video applications. He et al. [29] proposed a COFE method model for tracking different classes of multi-targets in different scenarios. The method contains three main modules: multi-class target detection, coarse-class multi-target tracking and fine-grained trajectory refinement.

With the development of UAV target tracking technology, we must at the same time be primarily aware of the risks involved. Especially when UAV target detection technology is applied in the field of urban security, it is important for professionals to be aware of the importance of UAV communication security, to understand possible threats, attacks and countermeasures related to UAV communication. It is also able to secure its communication using technologies such as blockchain technology, machine learning technology, fog computing and software-defined networking to guarantee the security and privacy of relevant data [30]. To deal with attacks and security threats such as jamming, information leakage and spoofing in UAV communication, Ko et al. [31] proposed a secure protocol after studying the security prerequisites of UAV communication protocols as a way to protect the communication between UAVs and between UAVs and ground control stations. The protocol can achieve perfect forward secrecy and non-repudiation, and is believed to have good applications in the field of urban security, where a high level of communication security is required. In summary, the correlation filter-based target tracking method can update the tracker at any time according to the diverse changes of the tracked targets, and it runs faster and is more suitable for target tracking in aerial UAV videos. Existing discriminative correlation filter-based trackers use predefined regularization terms to optimize learning for the target, such as to suppress the learning for the background or to adjust the change rate of the correlation filter. However, the predefined parameters not only require a lot of effort to adjust, but also cannot be adapted to new situations where no rules have been established.

Therefore, the automatic spatio-temporal regularization tracker (AutoTrack) [32] tracking algorithm improves on the STRCF algorithm, which uses the connection of the responses of two adjacent frames as an adaptive spatio-temporal regularization term and uses the global response change to determine its update rate, thus improving the tracking perfor-

mance. The spatio-temporal regularization term proposed in this algorithm can make full use of local and global response variations to achieve both spatial and temporal regularization, as well as automatic and adaptive hyperparameter optimization, based on the local and global information hidden in the response graph. The algorithm uses response variation to achieve regularization because the information hidden in the response graph is crucial in the detection process, and its quality somehow reflects the similarity between the target appearance model learned in the previous frames and the actual target detected in the current frame. Additionally, the reason why the algorithm utilizes both local and global response changes is that, if only global response map changes are used, then local response changes in the plausibility of different locations in the target image are ignored, and drastic local changes will lead to low plausibility, and vice versa.

Existing target tracking algorithms use a frame-by-frame approach to update the model, which can easily ignore the issue of whether the tracking effect of the current frame is accurate or not and update the tracker blindly, which can lead to tracker learning errors. Therefore, this paper proposes a strong interference motion target tracking method based on the target consistency algorithm for the problem of losing the tracked target due to the inaccurate tracking result of the current frame.

3. Method

3.1. Overall Structure

We propose in this paper a strong interference motion target tracking method based on the target consistency algorithm for aerial UAV video, and the general framework is shown in Figure 1. In the tracking model, the current frame is fused with the previous frame for tracking trajectory correction, and the previous frame is combined to update the model. Secondly, a trajectory confidence mechanism is proposed in the tracking model. The longer a trajectory is tracked, the more reliable this trajectory is, as a way to enhance the accuracy of subsequent tracking. Finally, the objective function is optimized using ADMM, and the problem is decomposed into multiple sub-problems to iteratively solve the problem and finally obtain the global optimal solution.

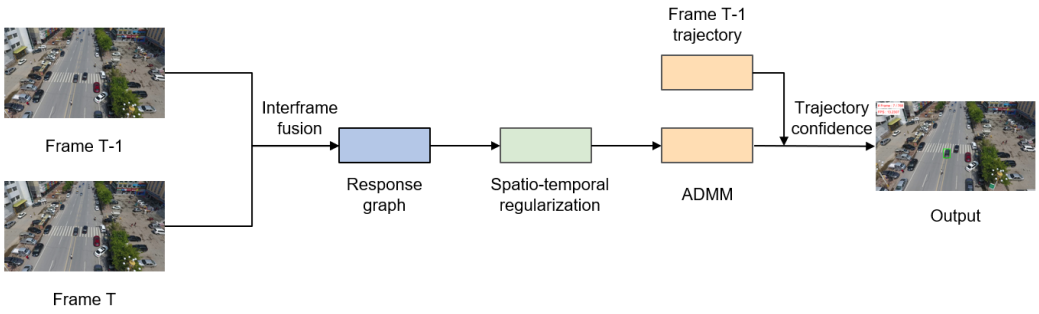

Figure 1. The general framework diagram.

3.2. Interframe Fusion

Since the tracking effect of the AutoTrack algorithm on the target only depends on the response map linkage between adjacent frames, when there is wrong tracking, it will cause the model to lose the effective tracking target information. Therefore, this paper improves the method based on the AutoTrack algorithm to enhance the interframe fusion capability of the model.

The method learns online and updates the relevant parameters automatically, using spatial local response variation as spatial regularization, allowing the filter to focus on learning the plausible places while using global response variation to determine the update rate of the filter and ensure its stability. This method adaptively learns and continuously adjusts the predefined parameters, which also use local as well as global response maps, with local variation indicating local plausibility in the target bounding box and global variation indicating global plausibility in the target bounding box, where severe illumination

changes and partial occlusion reduce the plausibility of the appearance, to dynamically adjust the spatial as well as temporal weights so that it is possible to make better use of the local and global information implied in the response map.

When the problem of losing the tracking target due to inaccurate tracking results of the current frame occurs, the method fuses the previous frames on the tracking results of the current frame for tracking trajectory correction, and updates the model by combining the previous frames to learn again to avoid tracker learning errors, thus enhancing the accuracy of subsequent tracking.

3.3. Trajectory Confidence Mechanism

To integrate the decomposability of the pairwise ascent method with the excellent convergence properties of the augmented Lagrange multiplier method, an improved form of the optimal alternating direction method of multipliers (ADMM) has been proposed. The aim is to be able to decompose the original function and the augmented function to facilitate parallel optimization under more general assumptions.

The core of the correlation filter-based target tracking problem is the solution of filters. With the advent of advanced algorithms, the models of filters are becoming more and more complex and computationally slow, making the advantage of correlation filtering in terms of computational speed less and less obvious. For example, the AutoTrack algorithm we improved in Section 3.2 uses spatio-temporal regularization to solve the boundary effect, and this measure to solve the boundary effect will make the tracking of correlation filtering face the challenge of real-time. Therefore, introducing the ADMM algorithm in this context can be a good way to divide a large optimization problem into multiple subproblems that can be solved simultaneously in a distributed manner, so that the objective function of the filter can be quickly minimized by iterating over the subproblems to obtain the global optimal solution.

The ADMM algorithm provides a framework for solving optimization problems with linear equation constraints, allowing us to break down the original optimization problem into several relatively well-solved suboptimization problems for iterative solving. This "disassembly" function is the core of the ADMM algorithm. The algorithm takes the following form.

$$\begin{aligned} &\min_{x,z} f(x) + g(z) \\ &s.t. Ax + Bz = c \end{aligned} \tag{1}$$

Here, both $f(x)$ and $g(z)$ are convex functions. At this point, their corresponding augmented Lagrangian functions are:

$$L_\rho(x, z, y) = f(x) + g(z) + y^T(Ax + Bz - c) + \left(\frac{\rho}{2}\right)\|Ax + Bz - c\|_2^2 \tag{2}$$

Additionally, its optimization steps are:

$$\begin{aligned} x^{k+1} &:= arg\min_x L_\rho\left(x, z^k, y^k\right) \\ z^{k+1} &:= arg\ min L_\rho\left(x^{k+1}, z, y^k\right) \\ y^{k+1} &:= y^k + \rho\left(Ax^{k+1} + Bz^{k+1} - c\right) \end{aligned} \tag{3}$$

This is a combination of the pairwise ascent method and the multiplicative Lagrange multiplier method. Theoretically, the optimization variables can be further split into more blocks, such as x, z, z_1, ... If we express the optimal solution of the original problem as:

$$p^* = inf\{f(x) + g(z) | Ax + Bz - c\} \tag{4}$$

then the ADMM algorithm, satisfying the basic assumptions, ensures that:

$$f\left(x^k\right) + g\left(z^k\right) \to p^* ask \to \infty \tag{5}$$

This also reflects the convergence of the algorithm, i.e., the final global optimal solution is obtained.

In summary, the trajectory confidence mechanism in this paper refers to the process of fusing the previous frames, which is not just a simple additive relationship, but is adjusted according to the tracking duration during the tracking process. The longer the tracking duration indicates that the tracking is more stable and therefore this trajectory is more credible, the higher the weight occupied by the current frame, as shown in Equation (6). This method performs fusion in a cumulative manner, not only with a particular frame, as a way to solve the tracking problem that is overly dependent on two adjacent frames.

$$S_i = \alpha * S_{i-1} + \beta * R_i \tag{6}$$

where i is the current frame, R_i is the detection position of the current frame, S_i is the correct position of the current frame and α and β are the weighting coefficients.

4. Experiment

4.1. Experimental Environment

The operating system of this experimental platform: Memory 16GB, GPU: NVIDIA GeForce RTX 2060, Graphic memory: 8GB.

4.2. Dataset

This experiment used the VisDrone-SOT [33] UAV image single target tracking dataset. This dataset was collected by the AISKYEYE team at the Lab of Machine Learning and Data Mining, Tianjin University, China. The benchmark dataset consists of 400 video clips formed by 265,228 frames and 10,209 static images captured by various drone-mounted cameras, covering a wide range of aspects including location (taken from 14 different cities separated by thousands of kilometers in China), environment (urban and country), objects and density (sparse and crowded scenes). The dataset was collected in different scenarios and under various weather and lighting conditions. These frames were manually annotated by more than 2.6 million bounding boxes or frequent target points of interest, and contain a total of 10 categories of targets for bus, car, van, truck, awning-tricycle, tricycle, motor, bicycle, pedestrian and people. To better utilize the data, some important attributes are also provided, including aspect ratio change, background clutter, camera motion, full occlusion, illumination variation, low resolution, partial occlusion, scale variation, similar objects, viewpoint change and several other cases. Based on the above, we believe that the dataset contains geographic factors, scene factors, weather and lighting factors and common target types in urban security, and can represent a real urban security environment to some extent.

4.3. Evaluation Metrics

To verify the effectiveness of the proposed method, a comparison is made using precision and success rates.

(1) Precision Plot

The accuracy graph mainly measures the percentage of successful frames of the target rectangular bounding box predicted by the tracker within a given threshold distance, and the distance between the predicted target position and the center point between the actual positions was calculated to obtain the accuracy value. The number of video frames whose distance between the predicted position and the center point of the actual position was smaller than the set threshold varies for different thresholds, and their percentage is different, so a curve can be obtained.

(2) Accuracy Plot

The accuracy rate plot shows the proportion of bounding boxes predicted by the tracker with a coincidence rate score greater than a given threshold. The overlap rate is defined as:

$$OS = \frac{|a \cap b|}{|a \cup b|} \tag{7}$$

where OS is the coincidence score, which takes values from 0 to 1, a is the rectangular bounding box of the target predicted by the tracker, b is the rectangular bounding box of the real position of the target and $|\bullet|$ denotes the number of pixels in the region. A frame is a successful frame if its coincidence score is greater than a given threshold. The accuracy rate is the number of all successful frames as a percentage of the number of all frames.

4.4. Experimental Results and Analysis of Target Tracking Algorithms for UAVs

Since the algorithm proposed in this paper is based on a UAV's target tracking task in the urban security domain, we will validate and analyze the experimental results of the algorithm proposed in this paper and some of the UAV tracking algorithms mentioned in Section 2 on the VisDrone-SOT dataset in this section.

The results in Table 2 show that our proposed algorithm is in the leading position in terms of accuracy compared with other UAV target tracking algorithms at 59.8%. However, in terms of precision, the SO-MOT algorithm [34] is the best at 91.7% and our algorithm is 91.5%, with a difference of 0.2%. This is due to the presence of a strong detector based on Cascade RCNN and an embedding model based on a multi-grain network in the SO-MOT algorithm and the creation of a simple online multi-target tracker. The model initializes some tracklets based on the estimated bounding box in the first frame, and in subsequent frames, associates the bounding box with the existing tracklets based on the distance measured by the embedding features, making it possible to update the appearance features of trackers at each time step to handle appearance changes.

Table 2. Experimental results of target tracking algorithms for UAVs.

Methods	Precision (%)	Accuracy (%)
FairMOT + ReID [26]	90.4	57.8
TF-DQN [27]	88.4	52.6
Mdnet [14]	90.2	54.9
COFE [29]	91.1	58.7
SO-MOT [34]	**91.7**	59.6
Ours	91.5	**59.8**

In summary, we believe that the algorithm proposed in this paper can satisfy a UAV's target tracking task in urban security in terms of precision and accuracy. Although the present algorithm is for single-target tracking, we believe that it can still be improved and applied to multi-target tracking tasks, which will be the next step of our research.

4.5. Visualization of Experimental Results

Figure 2 shows a visualization of the experimental results of the proposed method on the dataset, where the leftmost image is a screenshot of the original video, and the three images on the right are screenshots of the visualization of the algorithm tracking a single target on the original video, where the target labeled by the bounding box is the target we need to track.

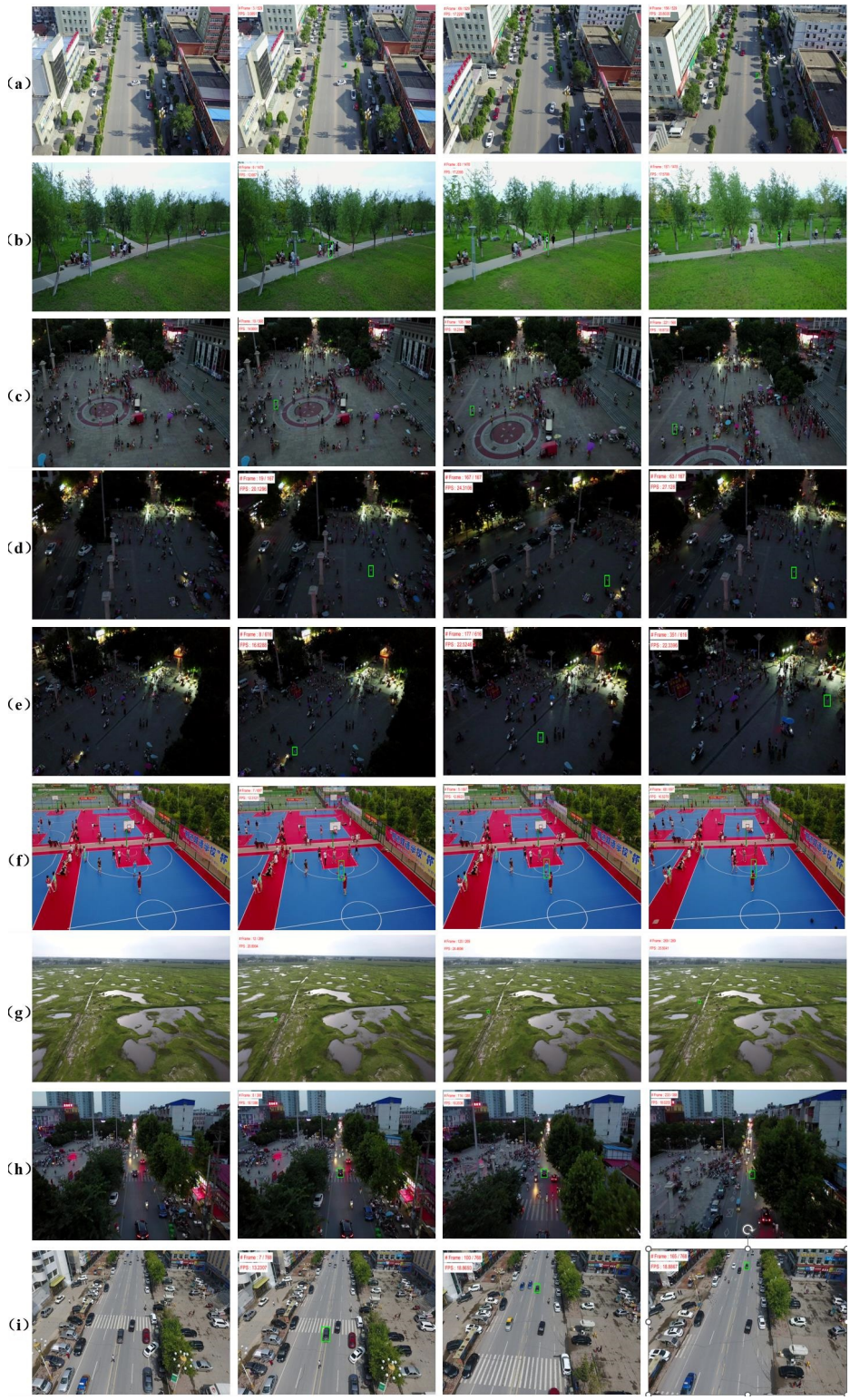

Figure 2. Visualization results of the algorithm under different conditions. The leftmost figure shows the original image in the dataset, and the right three figures show the resultant video frames of the tracking model. (**a**,**i**) shows the tracking effect when the light is sufficient and the targets are small in scale; (**b**,**f**) shows the tracking effect when the light is sufficient and the targets are obscured; (**c**,**h**) shows the tracking effect when the light is insufficient and the targets are dense; (**d**,**e**) shows the tracking effect when it is dark and the targets are small in scale; (**g**) shows the tracking effect when the targets are sparse and too small in scale.

4.6. Analysis of Trajectory Confidence Parameters

For the weight coefficients α and β in Equation (6), α should be less than 0.5, β should be greater than 0.5 and the sum of the weight coefficients should be 1, which means $\alpha + \beta = 1$, since the detection position of the current frame should account for a larger percentage. Since the method proposed in this paper is an improvement of the AutoTrack algorithm, we used its experimental results as a benchmark to find the optimal weighting values by comparing the experimental results of α and β of the grid search taking values.

The precision and total precision of the proposed method in various scenarios when the weights α and β were taken as different values are shown in Table 3 and Figure 3. The results of the experiments show that the improved method proposed in this paper does not have the best accuracy in all cases. Among them, in the cases of complete occlusion, illumination change and partial occlusion, the detection precision is higher for the cases of occlusion and the complex environment because the AutoTrack algorithm itself introduces a temporal regularization term to locate similar targets between different video frames. Similarly, the improved method with $\alpha = 0.4$, $\beta = 0.6$ is better in the case of aspect ratio variation and low resolution, because the key information of the target in the previous frames is more helpful for the model to achieve better interframe fusion to track the target correctly in the case of aspect ratio variation and low resolution. The improved method of $\alpha = 0.1$, $\beta = 0.9$ works better when the background is cluttered, the camera is moving and the viewpoint is changing, also because the target position information of the current frame is more important than the previous frame in the above case, which can guarantee the precision better. Finally, according to the results of total precision, it can be seen that the precision of all four weight distributions is higher than that of the original AutoTrack algorithm, among which the precision is highest when $\alpha = 0.1$, $\beta = 0.9$, which is 6.3% better than that of the AutoTrack algorithm.

The accuracy and total accuracy of the proposed method in various scenarios when the weights α and β take different values are also shown in Table 3 and Figure 3. The results of the experiments also show that the improved method proposed in this paper does not have the best accuracy in all cases. Among them, in the cases of aspect ratio change, complete occlusion, illumination change and low resolution, the accuracy of video frame detection in the case of large front-to-back changes of the target and a complex environment will be higher because the AutoTrack algorithm itself has a temporal regularization term. Meanwhile, the improved method of $\alpha = 0.2$, $\beta = 0.8$ works better in the case of proportional change because in this case, the key information of the target in the previous frame needed to be balanced with the target information of the current frame to ensure the detection accuracy of video frames. The improved method of $\alpha = 0.1$, $\beta = 0.9$ works better when the background is cluttered, the camera is moving and the viewpoint is changing, again because in this case, the target position information of the current frame was more important than that of the previous frame, which can better guarantee the accuracy. Finally, according to the results of the total accuracy, it can be seen that the accuracy of all four weight distributions is higher than that of the original AutoTrack algorithm, and the accuracy is still the highest when $\alpha = 0.1$, $\beta = 0.9$, which is 2.6% higher than that of the AutoTrack algorithm. Therefore, this paper adopts the improved algorithm with $\alpha = 0.1$, and $\beta = 0.9$ weight distribution.

Table 3. Comparison of precision and accuracy under different weights.

Weights		Aspect Ratio Variation (%)	Background Clutter (%)	Camera Motion (%)	Completely Obscured (%)	Illumination Variation (%)	Low Resolution (%)	Partial Occlusion (%)	Proportional Changes (%)	Similar Objects (%)	Viewpoint Changes (%)	Total (%)
AutoTrack	precision	80.7	78.8	83.5	**88.0**	**94.9**	91.9	**94.0**	92.8	92.8	75.2	85.2
	accuracy	**50.5**	50.6	56.4	**61.3**	**64.9**	**57.9**	65.5	63.4	58.6	48.1	57.2
A = 0.4 B = 0.6	precision	**82.4**	86.7	89.7	87.5	94.5	**94.1**	93.8	92.8	92.8	84.9	90.7
	accuracy	49.9	53.3	58.5	60.5	63.1	57.7	65.4	63.6	59.2	52.5	58.8
A = 0.3 B = 0.7	precision	80.6	81.8	89.8	87.5	89.6	80.0	93.8	92.8	80.9	86.5	87.3
	accuracy	50.2	50.3	58.6	60.5	60.2	50.6	65.4	63.6	52.1	52.5	56.7
A = 0.2 B = 0.8	precision	80.3	87.3	90.1	87.5	94.6	91.9	93.8	92.8	92.8	87.0	91.1
	accuracy	50.3	54.2	59.3	60.8	63.5	57.6	65.5	**63.7**	59.2	53.7	59.5
A = 0.1 B = 0.9	precision	80.0	**87.9**	**90.6**	87.5	94.4	91.9	93.8	92.8	92.8	**87.9**	**91.5**
	accuracy	49.8	**54.7**	**59.6**	60.8	63.2	57.3	65.5	63.5	59.2	**54.6**	**59.8**

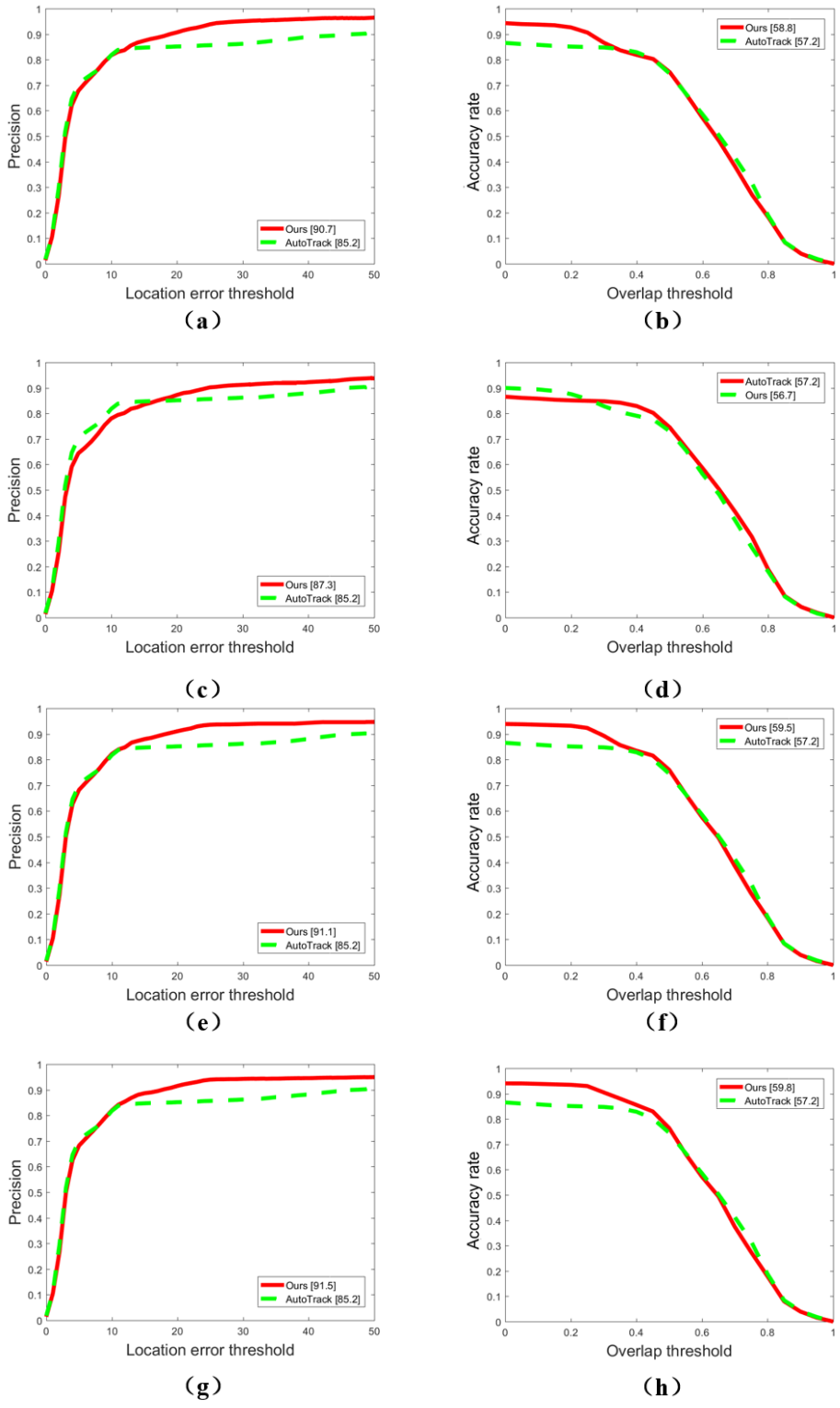

Figure 3. Comparison of total precision and total accuracy under different weights. (**a**) is the precision when $\alpha = 0.4$, $\beta = 0.6$, (**b**) is the accuracy when $\alpha = 0.4$, $\beta = 0.6$, (**c**) is the precision when $\alpha = 0.3$, $\beta = 0.7$, (**d**) is the accuracy when $\alpha = 0.3$, $\beta = 0.7$, (**e**) is the precision when $\alpha = 0.2$, $\beta = 0.8$, (**f**) is the accuracy when $\alpha = 0.2$, $\beta = 0.8$, (**g**) is the precision when $\alpha = 0.1$, $\beta = 0.9$ and (**h**) is the accuracy when $\alpha = 0.1$, $\beta = 0.9$.

4.7. Experiment Results and Analysis of Target Tracking Algorithm Based on Correlation Filtering

To test the effectiveness of the tracking algorithms proposed in this paper, we selected three excellent target tracking algorithms with correlation filter-based performance, Au-

totrack, Staple and ARCF, mentioned in Sections 1 and 2, and compared them with the algorithm proposed in this paper using the VisDrone-SOT dataset. Among them, Autotrack is the benchmark model of the algorithm proposed in this paper, which achieves both spatial and temporal regularization by making full use of local and global response variations through the spatio-temporal regularization term to achieve target localization. Additionally, Staple, based on the color response map derived from the global color histogram, uses the HOG method to extract the HOG features to obtain the dense response template and linearly combines the scores of the two templates to estimate the target location. In contrast, ARCF suppresses the rate of change of the response map at the time of detection, thus suppressing the distortion of the response map in the case of target occlusion and improving the tracking accuracy. Through experimental comparison with these three methods, the effectiveness of the proposed algorithm in this paper can be verified from three perspectives: the baseline model, the target tracking in complex cases and the target tracking in occlusion cases. The evaluation was performed using One-Pass Evaluation (OPE), which initializes the first frame of the image with the position of the actual labeled target, and the average accuracy and precision were obtained by calculation.

4.7.1. Precision and Accuracy Comparison of Different Correlation Filtering Algorithms

The comparison results of the precision and accuracy of different correlation filtering algorithms for various scenarios are shown in Table 4 and Figure 4.

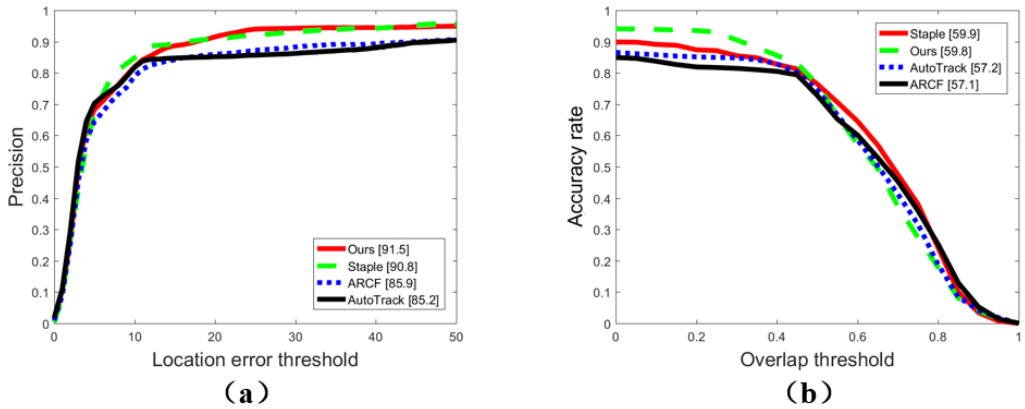

Figure 4. Comparison of the total precision and total accuracy under different correlation filtering algorithms. (**a**) is the total precision of different algorithms, (**b**) is the total accuracy of different algorithms.

It can be seen that the precision of Staple is higher in the cases of aspect ratio change, camera motion and viewpoint change, because Staple can derive the target location by combining both color response maps and HOG dense response templates, and thus has the best target precision performance for the case of target scale change. Meanwhile, the precision of AutoTrack is higher in cases of complete occlusion, illumination change and partial occlusion, also because it has its spatio-temporal regularization term, which can guarantee the detection precision in the case of target occlusion change. Our proposed algorithm, on the other hand, can still guarantee detection precision in background clutter due to the introduction of the trajectory confidence mechanism. Finally, the comparison of the total precision of different correlation filtering algorithms shows that the precision of the proposed method is as high as 91.5%, which is not only higher than the precision of the AutoTrack algorithm, but also higher than the precision of the Staple and ARCF algorithms. Therefore, the method proposed in this paper is better under the comprehensive consideration of multiple cases.

Table 4. Comparison of the precision and accuracy under different correlation filtering algorithms.

Algorithm		Aspect Ratio Variation (%)	Background Clutter (%)	Camera Motion (%)	Completely Obscured (%)	Illumination Variation (%)	Low Resolution (%)	Partial Occlusion (%)	Proportional Changes (%)	Similar Objects (%)	Viewpoint Changes (%)	Total (%)
AutoTrack	precision	80.7	78.8	83.5	**88.0**	**94.9**	**91.9**	**94.0**	**92.8**	**92.8**	75.2	85.2
	accuracy	50.5	50.6	56.4	61.3	**64.9**	**57.9**	**65.5**	63.4	58.6	48.1	57.2
Staple	precision	**82.5**	86.9	**94.0**	85.0	90.4	79.5	92.5	90.6	78.2	**94.0**	90.8
	accuracy	**53.1**	**57.7**	**62.8**	**62.7**	61.1	42.2	64.6	56.7	47.1	**64.1**	**59.9**
ARCF	precision	66.6	79.9	88.6	62.8	83.4	79.2	81.4	76.0	63.4	84.3	85.9
	accuracy	39.8	49.8	61.2	45.5	54.5	43.9	57.7	53.4	32.5	55.8	57.1
Ours	precision	80.0	**87.9**	90.6	87.5	94.4	**91.9**	93.8	**92.8**	**92.8**	87.9	**91.5**
	accuracy	49.8	54.7	59.6	60.8	63.2	57.3	**65.5**	**63.5**	**59.2**	54.6	59.8

Again, it can be seen that the accuracy of Staple is higher in the cases of aspect ratio change, background clutter, camera motion, complete occlusion and viewpoint change, again because Staple can derive the location of the target from the color response and dense response templates, and thus has the best accuracy for detecting targets in video frames in cases such as target scale change. At the same time, the accuracy of AutoTrack is higher in the case of illumination changes and low resolution, also because it has its spatio-temporal regularization term, which can guarantee detection accuracy in the case of large target changes. Our proposed algorithm, on the other hand, is able to guarantee detection accuracy in the case of scale changes and similar objects due to the introduction of an interframe fusion mechanism. Finally, the comparison of the total accuracy of different correlation filtering algorithms shows that the accuracy of the proposed method in this paper reaches 59.8%, which is higher than the accuracy of AutoTrack and ARCF algorithms and approaches the 59.9% accuracy of Staple. Therefore, the method proposed in this paper still has an excellent performance in a variety of situations.

4.7.2. Robustness Testing of Different Correlation Filtering Algorithms

To verify the robustness of the algorithms, the data set was disrupted in time and space and then evaluated using two evaluation metrics, namely Temporal Robustness Evaluation (TRE) based on disrupted time and Spatial Robustness Evaluation (SRE) based on disrupted space. SRE evaluates whether the algorithm is sensitive to initialization by slightly translating and scaling up or down the real labeled target position to produce an initialized position for target tracking, and finally obtains the average accuracy and precision.

Comparisons of the precision and accuracy of different correlation filtering algorithms under TRE or SRE are shown in Table 5 and Figure 5. In the case of TRE metrics, it can be seen that the precision of ARCF is higher in the case of aspect ratio change, background clutter, camera motion, complete occlusion, partial occlusion and viewpoint change; AutoTrack is more precise in the case of illumination change, low resolution and similar objects; and Staple is more precise in the case of scale change. AutoTrack is more accurate in the case of low resolution and similar objects, and ARCF is more accurate in several other cases. In terms of overall performance, the total precision and total accuracy of the ARCF algorithm under TRE are higher than the other algorithms compared with the total precision and total accuracy, which are both low under OPE. Meanwhile, it can be seen in the case of SRE metrics that AutoTrack has higher precision in cases of aspect ratio change, complete occlusion, partial occlusion, scale change and similar objects, while Staple has higher precision in several other cases. The accuracy of AutoTrack is higher in cases of complete occlusion, partial occlusion, scale change and similar objects, while the accuracy of Staple is higher in cases of aspect ratio change, background clutter, camera motion, illumination change and viewpoint change, and the accuracy of the proposed method in this paper is higher in the case of low resolution. In terms of overall performance, the Staple algorithm has the highest total precision and total accuracy.

In conclusion, after the results of comparison experiments and robustness experiments are evaluated, the method proposed in this paper outperforms AutoTrack in the case of background clutter, camera motion and viewpoint change, the method proposed in this paper has higher precision than other algorithms in the case of background clutter and the method proposed in this paper has higher accuracy than other algorithms in the case of scale change and similar objects. In terms of robustness, the proposed method in this paper is sensitive to the initial position given in the first frame, which will cause a relatively large impact at different positions or frame initials, and is also sensitive to the bounding box given during initialization. Therefore, the generalizability of the algorithm proposed in this paper in detecting both the temporal and spatial aspects of the video needs to be improved by further research. However, at the same time, this algorithm is intended to be applicable to UAV target tracking tasks targeting the urban security field, so a high accuracy and precision in the case of background clutter, camera motion, viewpoint change and similar objects can meet the task requirements well. Additionally, the generalizability of the algorithm for time and spatial location can be compensated to some extent by means of human intervention during the task. Therefore, this method can be well-applied to tracking tasks related to urban security.

Table 5. Comparison of the precision and accuracy of different correlation filtering algorithms under different metrics.

Algorithm	Evaluation Metrics		Aspect Ratio Variation (%)	Background Clutter (%)	Camera Motion (%)	Completely Obscured (%)	Illumination Variation (%)	Low Resolution (%)	Partial Occlusion (%)	Proportional Changes (%)	Similar Objects (%)	Viewpoint Changes (%)	Total (%)
AutoTrack	TRE	precision	71.7	81.9	87.8	78.2	**89.8**	**81.5**	89.1	87.8	**81.9**	85.7	87.3
		accuracy	46.6	54.7	61.1	55.0	62.2	**47.3**	65.2	60.8	**55.2**	60.1	59.3
	SRE	precision	**83.6**	77.7	87.3	**91.3**	90.0	78.3	**95.6**	**95.1**	**81.3**	81.8	84.4
		accuracy	49.2	44.9	55.0	**61.6**	55.9	43.2	**62.5**	**62.7**	**44.5**	49.0	51.8
Staple	TRE	precision	70.5	81.7	88.2	78.2	88.7	79.3	89.1	**88.7**	81.5	86.8	87.2
		accuracy	47.0	56.7	60.8	57.7	61.9	43.9	66.0	55.5	54.9	62.3	59.1
	SRE	precision	81.6	**87.0**	**93.6**	85.5	**90.5**	**80.6**	92.7	91.1	79.9	**93.4**	**90.9**
		accuracy	**49.5**	**54.4**	**59.8**	60.1	**57.6**	40.0	61.8	54.4	44.2	**60.7**	**56.8**
ARCF	TRE	precision	**74.5**	**83.1**	**90.0**	**79.4**	89.6	78.1	**89.7**	88.4	79.1	**89.6**	**88.2**
		accuracy	**50.0**	**57.9**	**64.4**	**60.2**	**64.1**	46.4	**68.4**	**61.8**	54.7	**65.1**	**61.9**
	SRE	precision	66.5	**77.7**	86.9	65.4	83.0	79.6	82.7	77.7	64.9	80.9	84.4
		accuracy	38.4	45.8	57.0	46.6	51.3	40.8	55.5	52.2	31.2	51.7	52.8
Ours	TRE	precision	64.8	78.1	86.1	69.9	85.1	78.0	85.0	82.5	73.0	82.4	84.6
		accuracy	42.2	51.3	59.1	46.7	58.1	45.6	60.8	56.9	49.1	56.4	56.9
	SRE	precision	71.1	72.6	83.0	74.2	85.0	79.0	87.1	83.9	70.9	74.2	80.8
		accuracy	40.7	42.3	52.2	48.7	52.7	**45.4**	56.6	55.2	40.3	44.5	49.9

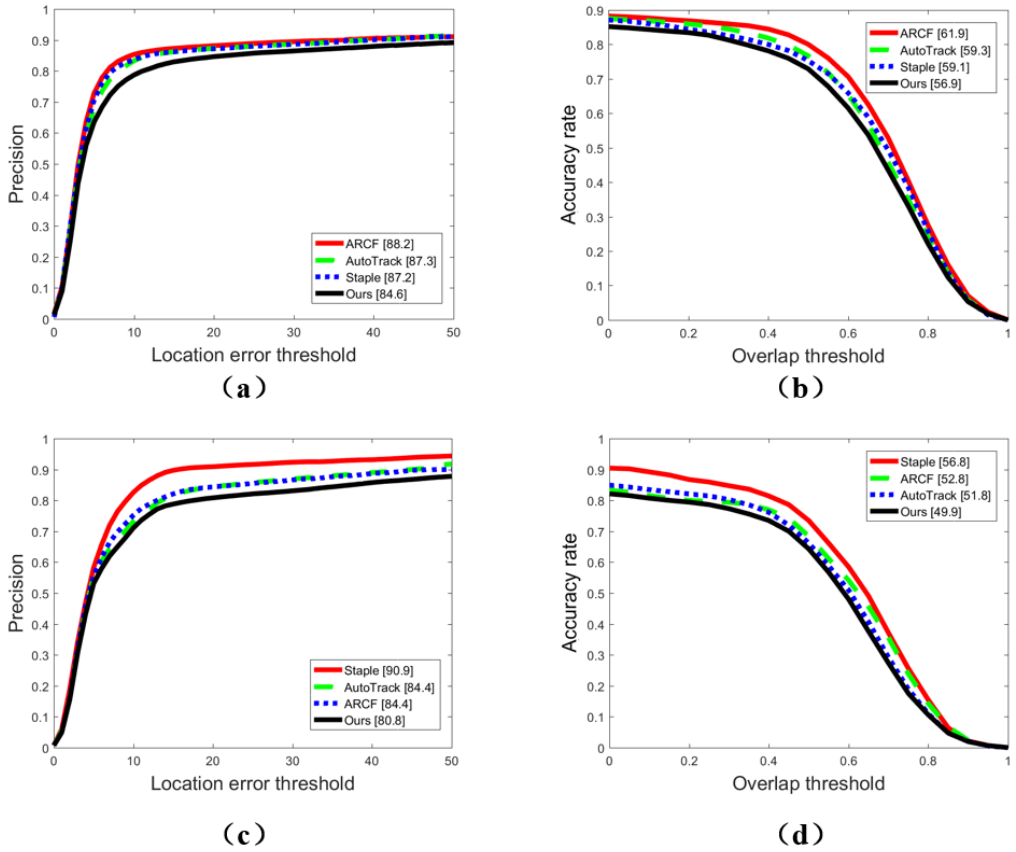

Figure 5. Comparison of the total precision and accuracy of different correlation filtering algorithms under different metrics. (**a**) is the total precision of different correlation filtering algorithms under TRE, (**b**) is the total accuracy of different correlation filtering algorithms under TRE, (**c**) is the total precision of different correlation filtering algorithms under SRE and (**d**) is the total accuracy of different correlation filtering algorithms under SRE.

5. Conclusions

To address the problem of losing the tracked target due to inaccurate tracking results of the current frame, this paper proposes a strong interference motion target tracking method based on the target consistency algorithm. When there is a tracking problem in the current frame, the tracking accuracy of the subsequent tracking is enhanced by combining the previous trajectories to learn again and updating the model according to the trajectory confidence mechanism to avoid tracker learning errors. The experimental results prove that the proposed method of this paper improves 0.2% of the accuracy compared with the current advanced UAV target tracking algorithm SO-MOT on the basis of guaranteed tracking precision, and also improves 6.3% of the total accuracy and 2.6% of the total accuracy compared with the benchmark model AutoTrack, which proves the effectiveness of the method. In particular, the high precision and accuracy in the case of background clutter, camera movement, viewpoint change and similar objects can well-meet the needs of target tracking tasks in aerial UAV video in the field of urban security. In the future, we expect that this method can be better applied in the field of urban security drone inspection to ensure the safety and stability of urban environments.

Author Contributions: Conceptualization: L.T., X.H. and X.L.; methodology: X.H. and X.L.; formal analysis and investigation: X.H. and X.L.; writing—original draft preparation: X.H. and X.L.; writing—review and editing: L.T., X.H., X.L., X.J. and H.L.; supervision: L.T. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Chongqing (CSTB2022NSCQ-MSX1415).

Data Availability Statement: The dataset used during the current study is available in https://github.com/VisDrone/VisDrone-Dataset (accessed on 11 November 2022).

Acknowledgments: The authors would like to thank the anonymous reviewers for their suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lu, H.; Li, Y.; Mu, S.; Wang, D.; Kim, H. Motor Anomaly Detection for Unmanned Aerial Vehicles using Reinforcement Learning. *IEEE Internet Things J.* **2017**, *5*, 2315–2322. [CrossRef]
2. Zhao, X.; Zhang, Q.; Wang, L.; Xie, F.; Zhang, B. A Special Operation UAV in Urban Space. In *AOPC 2021: Optical Sensing and Imaging Technology*; SPIE: Bellingham, WA, USA, 2021; Volume 12065, pp. 433–442.
3. Xu, Z. Application Research of Tethered UAV Platform in Marine Emergency Communication Network. *J. Web Eng.* **2021**, *20*, 491–511. [CrossRef]
4. Waleed, E.; Arslan, A.; Aliza, M.; Mohamed, I. Energy-Efficient Task Scheduling and Physiological Assessment in Disaster Management using UAV-Assisted Networks. *Comput. Commun.* **2020**, *155*, 150–157.
5. Han, Y.; Liu, H.; Wang, Y.; Liu, C. A Comprehensive Review for Typical Applications Based Upon Unmanned Aerial Vehicle Platform. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 9654–9666. [CrossRef]
6. Qin, X.; Wang, T. Visual-based Tracking and Control Algorithm Design for Quadcopter UAV. In Proceedings of the 2019 Chinese Control and Decision Conference (CCDC), Nanchang, China, 3–5 June 2019.
7. Zhang, R.; Sun, S.; Li, Y.; Li, Z.; Tian, K. An Adaptive Scale Estimation Target Tracking Algorithm Based on UAV. In Proceedings of the 2020 International Conference on Robots & Intelligent System (ICRIS), Sanya, China, 7–8 November 2020; pp. 545–551.
8. Bertinetto, L.; Valmadre, J.; Golodetz, S.; Miksik, O.; Torr, P.H. Staple: Complementary Learners for Real-Time Tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27 June–1 July 2016; pp. 1401–1409.
9. Danelljan, M.; Häger, G.; Khan, F.; Felsberg, M. Learning Spatially Regularized Correlation Filters for Visual Tracking. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 4310–4318.
10. Li, F.; Tian, C.; Zuo, W.; Zhang, L.; Yang, M.-H. Learning Spatial-Temporal Regularized Correlation Filters for Visual Tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4904–4913.
11. Galoogahi, H.; Fagg, A.; 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.
12. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Found. Trends Mach. Learn.* **2011**, *3*, 1–122. [CrossRef]
13. Huang, Z.; Fu, C.; Li, Y.; Lin, F.; Lu, P. Learning Aberrance Repressed Correlation Filters for Real-Time UAV Tracking. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 2891–2900.
14. Bi, F.; Lei, M.; Wang, Y. Context-Aware MDNet for Target Tracking in UAV Remote Sensing Videos. *Int. J. Remote Sens.* **2020**, *41*, 3784–3797. [CrossRef]
15. Zha, Y.; Wu, M.; Qiu, Z.; Sun, J.; Zhang, P.; Huang, W. Online Semantic Subspace Learning with Siamese Network for UAV Tracking. *Remote Sens.* **2020**, *12*, 325. [CrossRef]
16. Liu, Y.; Wang, Q.; Hu, H.; He, Y. A Novel Real-Time Moving Target Tracking and Path Planning System for A Quadrotor UAV in Unknown Unstructured Outdoor Scenes. *IEEE Trans. Syst. Man Cybern. Syst.* **2019**, *49*, 2362–2372. [CrossRef]
17. Li, A.; Luo, L.; Tang, S. Real-Time Tracking of Vehicles with Siamese Network and Backward Prediction. In Proceedings of the IEEE ICME, London, UK, 6–10 July 2020; pp. 1–6.
18. Chu, Q.; Ouyang, W.; Li, H.; Wang, X.; Liu, B.; Yu, N. Online Multi-Object Tracking using Cnn-Based Single Object Tracker with Spatial-Temporal Attention Mechanism. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 4836–4845.
19. Feng, W.; Hu, Z.; Wu, W.; Yan, J.; Ouyang, W. Multi-Object Tracking with Multiple Cues and Switcher-Aware Classification. *arXiv* **2019**, arXiv:1901.06129.
20. Li, B.; Yan, J.; Wu, W.; Zhu, Z.; Hu, X. High-Performance Visual Tracking with Siamese Region Proposal Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8971–8980.
21. Zhu, J.; Yang, H.; Liu, N.; Kim, M.; Zhang, W.; Yang, M.-H. Online Multi-Object Tracking with Dual Matching Attention Networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 366–382.
22. Wan, M.; Gu, G.; Qian, W.; Ren, K.; Maldague, X.; Chen, Q. Unmanned Aerial Vehicle Video-Based Target Tracking Algorithm. *IEEE Internet Things J.* **2019**, *6*, 9689–9706. [CrossRef]
23. Liu, Z.; Shang, Y.; Li, T.; Chen, G.; Wang, Y.; Hu, Q.; Zhu, P. Robust Multi-Drone Multi-Target Tracking to Resolve Target Occlusion: A Benchmark. *IEEE Trans. Multimed.* **2023**, 1–16. [CrossRef]
24. Yeom, S. Long Distance Ground Target Tracking with Aerial Image-to-Position Conversion and Improved Track Association. *Drones* **2022**, *6*, 55. [CrossRef]

25. Jiang, Y.; Jingliang, G.; Yanqing, Z.; Min, W.; Jianwei, W. Detection and Tracking Method of Small-Sized UAV Based on YOLOv5. In Proceedings of the 2022 19th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China, 16–18 December 2022; pp. 1–5.
26. Lin, Y.; Wang, M.; Chen, W.; Gao, W.; Li, L.; Liu, W. Multiple Object Tracking of Drone Videos by A Temporal-Association Network with Separated-Tasks Structure. *Remote Sens.* **2022**, *14*, 3862. [CrossRef]
27. Bhagat, S.; Sujit, P.B. UAV Target Tracking in Urban Environments using Deep Reinforcement Learning. In Proceedings of the 2020 International Conference on Unmanned Aircraft Systems (ICUAS), Athens, Greece, 1–4 September 2020; pp. 694–701.
28. Yang, B.; Cao, X.; Yuen, C.; Qian, L. Offloading Optimization in Edge Computing for Deeplearning-Enabled Target Tracking by Internet of UAVs. *IEEE Internet Things J.* **2020**, *8*, 9878–9893. [CrossRef]
29. Fan, H.; Du, D.; Wen, L.; Zhu, P.; Hu, Q.; Ling, H.; Shah, M.; Pan, J. Visdrone-MOT2020: The Vision Meets Drone Multiple Object Tracking Challenge Results. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Cham, Switzerland, 2020; pp. 713–727.
30. Krichen, M.; Adoni, W.Y.H.; Mihoub, A.; Alzahrani, M.Y.; Nahhal, T. Security Challenges for Drone Communications: Possible Threats, Attacks and Countermeasures. In Proceedings of the 2022 2nd International Conference of Smart Systems and Emerging Technologies, Riyadh, Saudi Arabia, 9–11 May 2022; pp. 184–189.
31. Ko, Y.; Kim, J.; Duguma, D.G.; Astillo, P.V.; You, L.; Pau, G. Drone Secure Communication Protocol for Future Sensitive Applications in Military Zone. *Sensors* **2021**, *21*, 2057. [CrossRef] [PubMed]
32. Li, Y.; Fu, C.; Ding, F.; Huang, Z.; Lu, G. AutoTrack: Towards high-performance visual tracking for UAV with automatic spatio-temporal regularization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11920–11929.
33. Zhu, P.; Wen, L.; Du, D.; Bian, X.; Fan, H.; Hu, Q.; Ling, H. Detection and Tracking Meet Drones Challenge. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*; IEEE: Piscataway, NJ, USA, 2021; Volume 44, pp. 7380–7399.
34. Chen, G.; Wang, W.; He, Z.; Wang, L.; Yuan, Y.; Zhang, D.; Zhang, J.; Zhu, P.; Gool, L.V.; Han, J.; et al. VisDrone-MOT2021: The Vision Meets Drone Multiple Object Tracking Challenge Results. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 2839–2846.