Hindawi
Mathematical Problems in Engineering
Volume 2019, Article ID 2419579, 17 pages
https://doi.org/10.1155/2019/2419579

Research Article

Aerial Infrared Target Tracking in Complex Background Based on Combined Tracking and Detecting

Yangguang Hu,[1] **Mingqing Xiao**,[1] **Kai Zhang**,[2] **and Xiaotian Wang**[2]

[1]*School of Aeronautics Engineering, Air Force Engineering University, Xi'an 710038, China*
[2]*School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China*

Correspondence should be addressed to Yangguang Hu; sunshineflyhu@163.com

Received 28 October 2018; Revised 27 December 2018; Accepted 7 February 2019; Published 14 March 2019

Academic Editor: David González

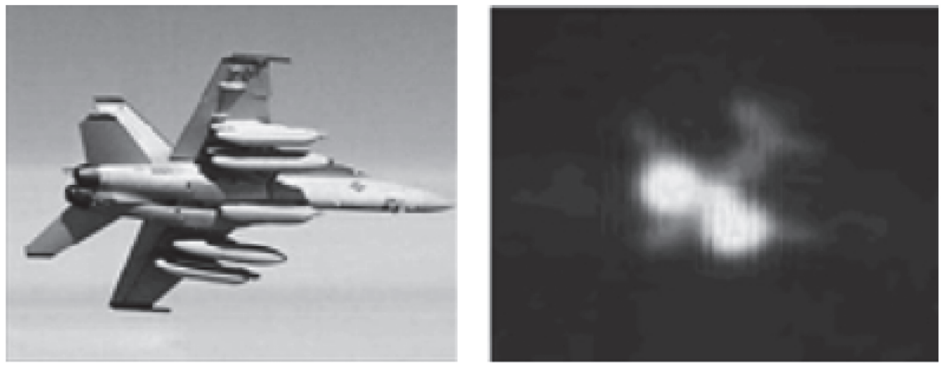
Aerial infrared target tracking is the basis of many weapon systems, especially the air-to-air missile. Till now, it is still challenging research to track the aircraft in the event of complex background. In this paper, we focus on developing an algorithm that could track the aircraft fast and accurately based on infrared image sequence. We proposed a framework composed of a tracker T based on correlation filter and a detector D based on deep learning, which we call combined tracking and detecting (CTAD). With such collaboration, the algorithm enjoys both the high efficiency provided by correlation filter and the strong discriminative power provided by deep learning. Finally, we performed experiments on three representative infrared image sequences and two sequences from VOT-TIR2016 dataset to quantitatively evaluate the performance of our algorithm. To evaluate our algorithm scientifically, we present the experiments performed on two sequences from AMCOM FLIR dataset of the proposed algorithm. The experimental results demonstrate that our algorithm could track the infrared target reliably, which shows comparable performance with the deep tracker, while running at a fast speed of about 18.1 fps.

1. Introduction

Infrared (IR) target tracking and detecting play a crucial role in military, video surveillance, and aeronautics applications [1, 2]. In military applications, IR imaging guidance is the development trend of IR guidance technology, which has attracted considerable attention due to its outstanding advantages of strong anti-interference ability, long detection distance, all-weather observations, and high guidance precision [3]. However, in contrast to visual images, IR images generally have low spatial resolution, poor signal-to-noise (SNR) ratios, and lack of textural information (see Figure 1) [4]. What is more, the platform used for tracking the aerial target moves so fast, which raises the problems of background motion and low target resolution [5]. Moreover, the development of various IR decoys makes it harder and harder to track the aerial target based on the IR image sequence. Despite significant progress in recent years, it remains a challenging task to find the useful information through the IR image sequence [6]. The tracking of an airborne IR target in a complex combat environment remains a challenging research field for IR imaging guidance.

Extensive work has been done in the area of aerial IR target tracking [3, 6–8]. Meanwhile, numerous superior trackers have been proposed in visual tracking, especially the deep tracker. When we use some classical algorithms in visible images to track the fighter in the complex background, it could not make us satisfied. Moreover, the infrared decoy develops fast in recent years, which presents the new challenge of tracking the fighter. The main challenge includes varied decoy, deformation, heavy occlusion, out-of-view, and illumination change from rapid maneuvering. The IR images are obtained by sensing the radiation in the IR spectrum, and due to this property, the signature of IR images is quite different from visual images.

In essence, IR target tracking could also be treated as a visual object tracking issue, which plays a crucial role in computer vision and many other domains. Inspired by the excellent performance of the algorithms in the visual object tracking (VOT) challenge, we introduced the superior



(a) (b)

Figure 1: (a) Visual image and (b) IR image of the fighter.

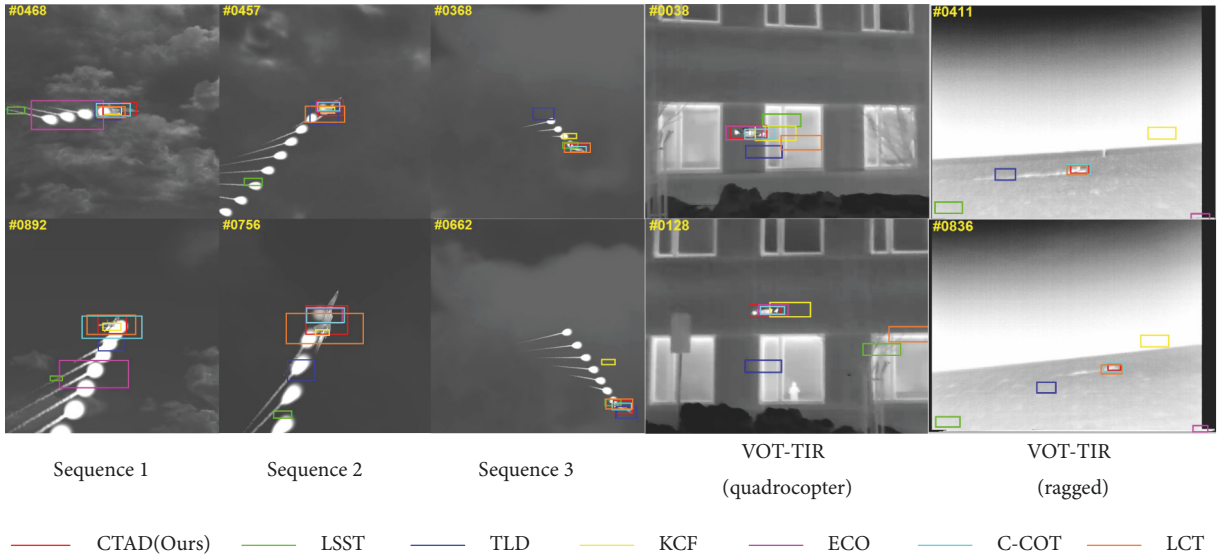


Figure 2: A comparison of the proposed CTAD with other 6 kinds of trackers on five representative sequences.

tracking theory to IR target tracking [9]. The VOT 2017 results show that the mainstream methods of tracking algorithms are mainly divided into three types: the first is the traditional correlation filtering method, the second is based on the convolutional neural network method, and the third is the combination of deep convolution features and traditional collaborative filtering. Among them, the method using deep convolution feature and collaborative filtering is the best. Despite the above-mentioned progress in either speed or accuracy, real-time, high-quality tracking algorithms remain scarce, especially for IR target tracking.

Our approach builds on two major observations based on prior work. The first one is deep learning-based techniques benefit from the expressive power of CNN features, which have outperformed previous low-level feature-based trackers [10]. Such algorithms, unfortunately, often suffer from high computational burden and hardly run in real-time. The other is the correlation filter-based trackers easily running at real-time, which also have an acceptable accuracy. To balance the tracking performance and the speed, we propose to build real-time high-accuracy tracker composed of a tracker T based on correlation filters (CF) and a detector D based on deep learning, which could track the target under the cluttered or even deceptive background. The main contributions of this work include the following:

As our first contribution, we explored the implication of excellent algorithm in IR target tracking, including LSST [11], TLD [12], KCF [13], LCT [14], C-COT [15], and ECO [16]. The experiment result is meaningful for the research of aerial target tracking based on IR image sequences. An example of the tracking result was shown in Figure 2.

Our second contribution is a tracking algorithm called CTAD, which shows comparable performance with the deep tracker, while running at a fast speed of about 18.1 fps, as shown in Figure 3, for instance.

The third contribution is that we use a detecting algorithm to detect the target when it is lost, which is beneficial for the tracking of the aerial target. As a result, the algorithm is more applicative for the aerial infrared target tracking.

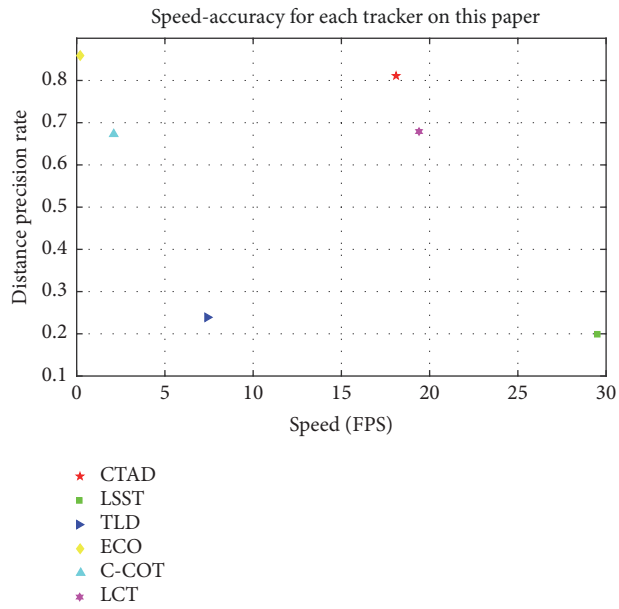


Figure 3: Comparison of computational efficiency, in terms of speed and accuracy, between the proposed tracker CTAD and the state-of-the-art visual trackers used in this paper using five IR sequences.

The remainder of the paper is organized as follows. In Section 2, we examine the existing literature and explain why we used a structure of combined tracking T and a detector D. In Section 3, we present the model structure of our approach, and then the details of the algorithm are introduced. In Section 4, we present experimental results and compare it with state-of-the-art algorithms in three representative IR image sequences and two sequences from VOT-TIR2016 dataset. We also present the experiments performed on two sequences from AMCOM dataset of the proposed algorithm. Comprehensive experiments clearly demonstrate that our approach concurrently has excellent performance in both tracking accuracy and speed.

2. Related Work and Problem Context

In this section, we first present a detailed review of the IR trackers and visual trackers based on the discriminative correlation filter and deep learning. Moreover, we introduced the idea of verification in tracking, which inspired us to design the structure of CTAD.

2.1. IR Trackers. IR sensors/cameras produce images with a low SNR, which causes differences between IR trackers and visual trackers to some extent. In [17], data received from the autopilot was used to improve the robustness and speed of the target tracking. When the size of the target does not change significantly and ego-motion is small, the exploiting morphological operators could benefit the tracking accuracy. In [6, 18], morphological filters are used in conjunction with multiscale decomposition and adaptive thresholding techniques to enhance SNR. Mean-shift based approaches are widely used in IR target tracking, as they provide a general optimization solution, which is independent of target features. However, it could not track the target properly when the target is affected by strong sensor ego-motion. In [19], they compensate the ego-motion using Gabor responses of two consecutive frames and proposed a pseudo-perspective motion model. In [20], they used a motion prediction metric to identify the occurrence of false alarms, and the tracking performance was improved by target detection based on efficient template matching. Moreover, consistency check and adaptive prediction are also used soon in IR target tracking. In [21], AM-FM consistency check was used for IR target tracking, which could dramatically improve the performance of challenging infrared data sequences and even the AMCOM forward-looking IR dataset. In [22], adaptive prediction of initial searching points was used to improve the efficiency and robustness of the tracker. In [23], histogram-based appearance learning was introduced to IR target tracking. Combined with the adaptive Kalman filter, it also got a good result in the AMCOM IR dataset.

As we can see from above, many scholars have done a lot of work on this problem and made brilliant achievements. At the same time, deep learning has got a state-of-the-art result in both target tracking and detecting. We introduce the research findings of visual trackers to improve the performance of the IR tracker.

2.2. Visual Trackers

2.2.1. Discriminative Correlation Filter. Correlation filters originated from the field of signal processing and was later applied to image classification and other aspects [24]. Correlation is a measure of the similarity of two signals. If the two signals are similar, the correlation value is higher. In the tracking application, it is necessary to design a filter template so that when it acts on the tracking target, it gets the most maximal response. It has shown dominant and impressive results on many object tracking benchmarks [9]. The most impressive advantage of correlation filters is their high speed. In [25], the correlation filters could deal with 669 figures per second, which raises the speed of tracking algorithm from the real-time level to the high-speed level. Impressed by the state-of-the-art performance of the correlation filters, Henriques et al. [26] improved the tracking performance by extending the correlation filter to multichannel inputs and kernel-based training. It could perform 320 fps using the Histogram of Oriented Gradient. In [13], it has been extended with kernels and multichannel features, which could get a higher precision. Nevertheless, those methods are limited to only estimating the target translation; they could not track the target properly when the target has significant scale variations. In [27], they tackled the challenging problem of scale estimation for visual tracking with discriminative correlation filters. The advancement in DCF tracking performance is predominantly attributed to powerful features and sophisticated learning formulations.

For an IR imaging missile, the tracking precision of the final term is much more crucial than the beginning and the middle term. The tracking of the aerial IR target

tracking is a long-term tracking issue, so we must consider the changes of appearance. However, long-term tracking is much more difficult than short-term one. With the change of size, characteristics, and appearance of the aerial IR target, the effectiveness of correlation filter could not fit the requirement of aerial IR target tracking completely. Therefore, the deep learning method may give us a new solution.

2.2.2. Deep Trackers. The performance of visual tracking has vastly improved with the advances of deep learning research [28]. In [29–31], the conventional network was used as a feature extractor, and they all adopt correlation filter as their base tracker. MDNet [32] trains a small-scale network by separating domain-specific layers, which shows that the CNN depth feature is indeed used to improve the tracking results. C-COT [15] employs the implicit interpolation method to solve the learning problem in the continuous spatial domain. ECO [16] is an improved version of C-COT in terms of both performance and speed. While these trackers result in high accuracy and robustness, their computational speed could not fulfill the real-time requirement of online tracking [15, 29].

The state-of-the-art performance of deep trackers benefits from the expressive power of CNN features. However, unfortunately, it often suffers from high computational burden. What is more, the computing power of the missile platform is not comparable with the high-performance GPU, which makes it even impossible to use deep trackers in our application. As a result, the trackers based on deep learning are not a good solution in our application by now.

2.3. YOLO: You Only Look Once. Deep learning model provided a method for learning representations of data with multiple levels of abstraction. It has dramatically improved the state-of-the-art in visual object recognition and many other domains. With the development of the deep learning, the new architectures accelerate the progress of visual object recognition [33, 34]. YOLO is a new approach for object detection which contains a single network, it can be optimized end-to-end directly on detection performance [35]. Unlike the traditional deep learning object detection algorithms [34], YOLO deals with the object detection as a regression problem. The experiment of the YOLO shows it could process images in a much higher speed. Moreover, the author made modifications to the prior version of YOLO, which made it more and more superior for detecting task.

To improve the accuracy of object positioning and recall rate, the author of YOLO proposes various improvements to the YOLO detection method called YOLO9000 [36]. In this paper, they employ the idea of anchor box in the fast R-CNN to modify the architecture. The output layer was replaced by the fully connected layer, and the ImageNet object classification data was used for training the model. Compared to YOLO, YOLO9000 has greatly improved in terms of recognition type, accuracy, speed, and positioning accuracy.

In [37], the author makes some changes to the YOLO9000 to make it better. It is a little bigger than last time but more accurate. They produced the multiscale prediction and a better classifier rather than Soft-max. Compared with the two previous versions, the YOLOv3 is more accurate and faster.

2.4. Verification in Tracking. For a long-term tracker, the idea of verification is a crucial and useful approach for improving the performance of the tracker. The TLD tracker is a good example, in which tracking results are validated per frame to decide how learning and/or detection shall progress. Unlike in previous studies, the verification in our algorithm should have the ability to detect the IR target in the event of complex background, especially the infrared decoy. If the tracker could not track the target properly, we need to detect the target again until the target could be tracked rightly.

Despite deep learning trackers obtaining superior performance, they suffer from the heavy computation for extracting deep features from every figure. Motivated by the verification approach and the application of the aerial target, we need to design algorithm considering two issues. On the one hand, the effective verification mechanism is quite necessary. On the other hand, the tracker needs to track the target properly in most cases.

In summary, we proposed an algorithm composed of a tracker based on the structure of the LCT tracker and a deep learning detecting method based on the YOLOv3. As a result, it could take advantage of both deep tracker and high-speed correlation filter.

3. Methodology

In the following section, we would like to introduce the details of our algorithm called CTAD: a tracker based on the LCT and a detector YOLOv3 used for verifying the tracking result. First, we would like to introduce the framework of our algorithm. Second, we will provide details about the tracker T and the detector D.

3.1. Framework. The algorithm contains two parts: one is the tracker T, and the other is the detector D. The two components work together toward real-time and high-accuracy tracking. Illustration of the CTAD is shown in Figure 4.

(i) The tracker T is the core of the algorithm, which is responsible for tracking the target in most of the process. It is responsible for the "real-time" requirement of our application. This tracker was chosen with vast tests using our IR image sequences. The structure of the T is based on an algorithm called LCT, which was proposed to address the problem of long-term visual tracking where the target undergoes significant appearance variation and heavy occlusion. Through decomposing the task of tracking into translation and scale estimation of objects, the algorithm improves the accuracy and reliability under complex environment.

(ii) The detector D, based on deep learning called YOLOv3, is used as a verification in tracking. The result of YOLOv3 shows that, in most cases, the

```
1 Initialize the tracking thread for tracker T;
2 Initialize the detecting thread for detector D;
3 Run tracker T;
4   if index of the figure > Δn or track failure then
5      Run detector D
6      Verify the position of the target with the result of detector D;
7   else
8       Run tracker T;
9   End.
```

Algorithm 1: Combined tracking and detecting (CTAD).

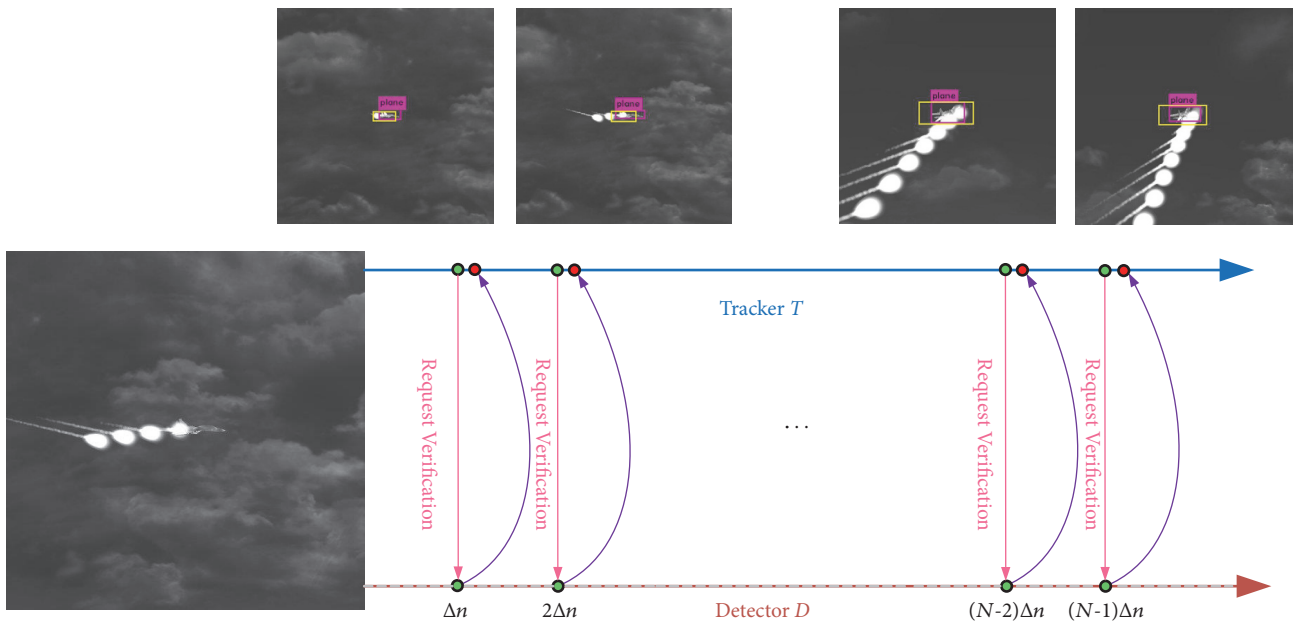


Figure 4: Illustration of the CTAD framework in which detector D is used to modify the track result.

method could detect the target even if it contains complex background. In our approach, the detector is used for supporting a new position of the target in a frequency of a still value. To avoid heavy computation, detector D only verifies the tracking result in a certain frequency, which means that in most cases the tracker T works independently.

In our research, the tracker and the detector work together with necessary interactions to keep tracking the target in acceptable precision and speed. When the tracker could not find any target in the view or it has dealt with more than Δn figures, the tracker T sends a request for verification (red solid in Figure 4.). The detector D detects the target in the view at that time and sends the detection result to the tracker T. Then, the tracker T continues to track the target again.

The tracker T and detector D were initialized in the first frame. When we have used N images in the image sequence, the detector D starts to work. With the implication of the YOLOv3, the tracker could get a verification. Then, we still use tracker to track the target until the end of the process.

It is worth noting that the CTAD combines two excellent algorithms in tracking and detecting field. The tracking algorithm T has been evaluated in our IR image sequence, which could track the target properly even in a complex background. The YOLOv3 is a modified version of the YOLO. Algorithm 1 summarizes the general CTAD framework.

3.2. CTAD Implementation

3.2.1. Tracking T: Two Regression Models Based on Correlation Tracking. We choose the LCT tracker [14] as the base of the tracker T in our algorithm based on its superior performance in long-term tracking and numerous implications in our IR image sequences. As a long-term tracking algorithm, the LCT could fit the need for the IR image guidance missile. LCT is a tracker based on DSST [27], which added the third filter responsible for detecting target confidence. Tracking confidence is quite crucial for aerial target tracking, which needs to be able to reflect the reliability of each tracking result. When the tracking confidence is low to a certain number, we need to detect the target in the view again.

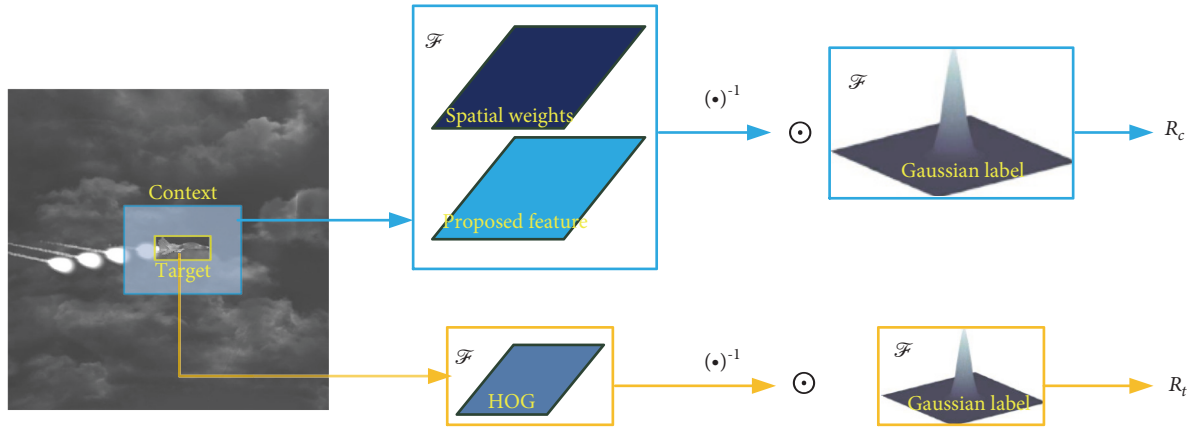


Figure 5: Two regression models learned from a single frame.

In the general formulation, a correlation filter contains a set of training patterns and labels $(x_1, y_1), \ldots, (x_m, y_m)$, a classifier $f(x)$ is trained by searching the parameters that minimize the regularized risk. When the classifier has a form like $f(x) = \mathrm{w}x + b$, the issue could be described like this

$$\min \sum_{i=1}^{m} L\left(y_i, f\left(x_i\right)\right) + \lambda \left\|\mathrm{w}\right\|^2. \tag{1}$$

Here $L(y, f(x))$ is a loss function, λ is a parameter used to control the amount of regularization, and w refers to learning rate, which contains the coefficients of a Gaussian ridge regression [38]. We use $L(y, f(x)) = (y - f(x))^2$ as the loss function. We need to find the w, which could get the goal of (1). It could be written like this

$$\mathrm{w} = \operatorname{argmin} \sum_{m,n} \left|\phi\left(X_{m,n}\right) \cdot w - y\left(m.n\right)\right|^2 + \lambda \left|\mathrm{w}\right|^2. \tag{2}$$

Here ϕ denotes the mapping to a kernel space. Fast Fourier transformation (FFT) was used to compute the correlation. Then, the object function could be minimized as follows.

$$\mathrm{w} = \sum_{m,n} a\left(m, n\right) \phi\left(\mathrm{x}_{m,n}\right) \tag{3}$$

The coefficient a is defined by the following equation:

$$A = \mathcal{F}\left(a\right) = \frac{\mathcal{F}\left(y\right)}{\mathcal{F}\left(\phi\left(\mathrm{x}\right) \cdot \phi\left(\mathrm{x}\right)\right) + \lambda} \tag{4}$$

where $\mathcal{F}$ denotes the discrete Fourier operator. For a new image with the search window size $M \times N$, the response map could be obtained from the following function:

$$\hat{y} = \mathcal{F}^{-1}\left(A \odot \mathcal{F}\left(\phi\left(\mathrm{z}\right) \cdot \phi\left(\hat{\mathrm{x}}\right)\right)\right) \tag{5}$$

where $\hat{\mathrm{x}}$ denotes the learned target appearance model and $\odot$ is the Hadamard product. Therefore, the new position of target is detected by searching for the location of the maximal value of $\hat{\mathrm{y}}$.

As shown in Figure 5, there are two regression models based on correlation filters from one single frame. The context model R_c was calculated based both the target and surrounding context into account. As a temporally stable and useful information, it is quite useful to discriminate the target from the background in the case of occlusion. The regression model R_c needs to be capable of dealing with the issue of occlusion, deformation, and abrupt motion. So, the model is updated with a learning rate α frame by frame as

$$\hat{x}^t = (1 - \alpha)\, \hat{x}^{t-1} + \alpha x^t \tag{6a}$$

$$\widehat{A}^t = (1 - \alpha)\, \widehat{A}^{t-1} + \alpha A^t \tag{6b}$$

where t denotes the index of the current frame.

The other discriminative regression model R_t used for estimating the scale of the target from the most reliable tracked targets. Based on the observation of the scale of little change between two consecutive frames, the maximal value of $\hat{\mathrm{y}}$ was used for describing the confidence of the tracking result. To improve the stability of the model, only update the model when $\max(\hat{y}) \geq T_a$. T_a is a predefined threshold.

$$\hat{s} = \arg\max_{s} \left(\max\left(\hat{y}_1\right), \max\left(\hat{y}_2\right), \ldots, \max\left(\hat{y}_s\right)\right) \tag{7}$$

Accordingly, the regression model R_t is updated by (6a) and (6b) when $\max(\hat{y}_s) \geq T_a$.

3.2.2. Detector D: YOLOv3. As we know, deep learning could act as state-of-the-art in the detecting task. It could learn very general representations of objects, which is crucial for the tracking of the fighter. Unlike the traditional detection systems that repurpose classifiers to perform detection, where the classifier is run at evenly spaced locations over the entire image with sliding window and region proposal-based techniques [33, 34], YOLO uses a single convolutional network that simultaneously predicts multiple bounding boxes and class probabilities for those boxes, then trains the network on full images, and directly optimizes detection. The structure has many advantages, such as high speed and robustness. YOLOv3 is the latest version of YOLO, which is

one of the most balanced target detection networks for speed and accuracy. Through the integration of various advanced methods, the short board of YOLO series was modified.

Following YOLO 9000, the YOLOv3 predicts bounding box using the anchor boxes. Four coordinates were predicted by network for each boning box (b_x, b_y, b_w, b_h). Once the cell is offset from the top left corner of the image by (c_x, c_y), then the bounding box point has width and height, and then the predictions correspond to the following.

$$b_x = \sigma(t_x) + c_x \tag{8a}$$

$$b_y = \sigma(t_y) + c_y \tag{8b}$$

$$b_w = p_w e^{t_w} \tag{8c}$$

$$b_h = p_h e^{t_h} \tag{8d}$$

The equations above are convenient for the computation of the ground truth value. They predict an objectness score for each boning box using logistic regression. If the bounding box prior overlaps a ground truth object by more than any other bounding box prior, it should be 1. If the bounding box is not the best but does overlap a ground truth object by more than a threshold of 0.5, the prediction will be ignored. Different from the YOLO9000, YOLOv3 will only assign one bounding box prior for each ground truth object. If previous bounding box is not assigned to grounding box objects, the coordinates or category forecast will not cause damage.

Different from the past YOLO, which always struggled with small target, YOLOv3 uses multiscale feature fusion, so the number of bounding boxes is much higher than before. The YOLOv3 is more sensitive for small targets. It predicts boxes at three different scales.

A new network called Darknet53 used for performing feature extraction was proposed, which has 53 convolutional layers. It uses successive 3×3 and 1×1 convolutional kernel and much more layers than before. The Dark-53 is more powerful than the ResNet-101, as well as the Dark-19. It achieves the highest measured floating-point operations per second and will better utilize the GPU.

In summary, YOLOv3 is a superior detector, which is fast and accurate. In the CTAD, the superior performance will benefit the tracking accuracy of the tracker.

4. Experimental Results

4.1. Implementation Details and Dataset. Our tracker is implemented in MATLAB and the detector D uses TensorFlow on a single NVIDIA GTX 1080 GPU with 4GB memory. The two networks communicated with each other via a TCP-IP socket. The detector D is used to verify the tracking module when LCT works nearly all over the tracking process. The detector D is based on the YOLOv3, which was trained with a training dataset composed of a sequence of 2200 IR images. And, it contains a new IR image sequence with 1783 images and 154 images from the sequence 1, 120 images from the sequence 2, and 143 images form the sequence 3. When we trained the network of the YOLOv3, we made an iteration of 5000 times.

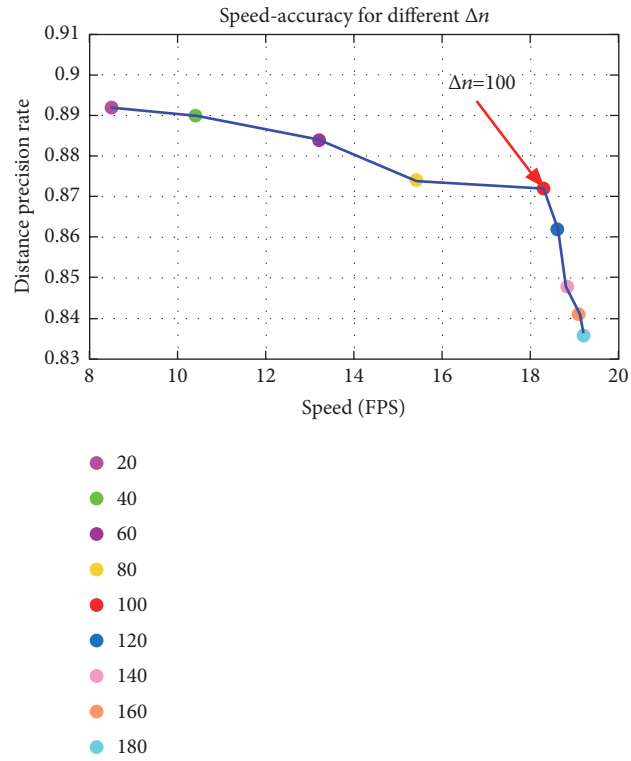


Figure 6: Speed-accuracy for different Δn.

In order to evaluate performance related to Δn, we change the value of Δn form 20 to 180 during the experiment on twelve IR sequences. The result of the experiment was shown in Figure 6. In order to find a balance between accuracy and speed, we chose Δn=100 in our algorithm.

We valuated our algorithm CTAD in three representative IR sequences supplied by simulation software used for infrared image processing and two sequences form VOT-TIR2016 dataset. The VOT-TIR2016 dataset are public and available online (http://www.votchallenge.net/vot2016/dataset.html). The simulated sequence lengths varied from 600 to 1100 figures. The individual sequence typically consists of different background (modicum cloud, middling cloud, vast cloud, clear sky), number of decoys, sorts of maneuvers, which pose challenging of interference and scale variations, fast motion.Each IR image sequence contains a complete tracking process. The details of the five image sequences is shown in Table 1.

4.2. Evaluation. As performance measure, we chosen six kinds of classic tracking algorithm. All the tracking methods are evaluated by two metrics: (i) Precision plot, which shows the percentage of image frames whose tracked location is within the given threshold distance of ground truth. (ii) Success plot, the metric of the success plot evaluates the tracker with bounding box overlap [32]. which shows the given the

Table 1: Details of the three sequences.

	Sequence 1	Sequence 2	Sequence 3	VOT-TIR 1	VOT-TIR 2
Size	512×512	512×512	512×512	640×480	640×480
Figure number	972	772	740	178	1035

Table 2: Comparisons with six state-of-the-art tracking methods on three infrared image sequences and two sequences from VOT-TIR 2016.

		CTAD (Ours)	TLD	LSST	KCF	C-COT	ECO	LCT
Sequence 1	DPR(%)	86.6	29.4	29.3	59.2	90.6	75.4	69.7
	OSR(%)	78.5	6.9	3.4	27.3	61.4	41.7	35.3
	Speed (fps)	17.6	37.9	7.0	178.0	0.2	2.2	18.5
Sequence 2	DPR(%)	75.1	51.3	6.7	69.3	100	100	81.3
	OSR(%)	53.6	18	0.1	28.4	46.5	45.7	52.8
	Speed (fps)	18.4	25.0	6.4	185.0	0.2	2.2	20.9
Sequence 3	DPR(%)	100	13.6	70.9	40.8	100	100	100
	OSR(%)	93.4	0.4	0.1	25.4	61.6	53.5	75.8
	Speed (fps)	18.9	31.2	6.9	174.0	0.2	2.2	19.2
VOT-TIR 1	DPR(%)	79.8	0.6	0.6	51.0	61.2	48.9	8.4
	OSR(%)	77.5	0.6	0.6	0.6	16.9	54.5	9.0
	Speed (fps)	17.6	25.4	7.5	192.0	0.3	1.68	18.5
VOT-TIR 2	DPR(%)	75.7	4.8	12.1	12.1	77.4	12.1	80.1
	OSR(%)	74.5	2.0	12.1	12.1	66.6	12.1	59.4
	Speed (fps)	18.6	28.0	9.4	197.0	0.2	2.3	19.9
Average	DPR(%)	81.1	19.9	23.9	46.5	85.9	67.3	67.1
	OSR(%)	75.7	5.6	3.3	18.7	50.6	41.5	46.4
	Speed (fps)	18.1	29.5	7.4	185.2	0.2	2.1	19.4

tracked bounding box and the ground truth bounding box, the overlap score is defined as in the following equation:

$$S = \frac{\left|B_t \cap B_{gt}\right|}{\left|B_t \cup B_{gt}\right|} \tag{9}$$

where $\cup$ and $\cap$ represent the intersection and union operators, $|\cdot|$ denotes the number of pixels in a region.

The speed of tracking methods is affected by different factors despite the very of the platform. For aircraft tracking, the images sequences change fast as the high speed. The speed of the algorithm is the basis of application requirement. As the length of the image sequence is different, we make an operation of average to make the evaluation more exactly. The fps (figure per second) was calculated with the following equation:

$$fps = \frac{\sum_{i=1}^{n} N_i}{\sum_{i=1}^{n} t_i} \tag{10}$$

where N_i represent the length of the image sequence, t_i denotes the cost time, i denotes the index of the image sequence.

4.3. Experiment on the IR Dataset. The state-of-the-art methods, including TLD, LSST, KCF, C-COT, ECO, LCT were estimated on the dataset above. Those method contain deep trackers, Discriminative Correlation Filter method and machine learning method. All the trackers above were employed in the three image sequences and two sequences form VOT-TIR2016 dataset. We employ one-pass evaluation (OPE) to compare those trackers.

Figure 7 shows the success plot and precision plot of our algorithm and six kinds of state-of-the-art trackers in simulated IR sequences and two sequences form VOT-TIR2016 dataset. Our approach shows comparable performance with the deep trackers like C-COT and ECO, while run at a fast speed of about 18.1 fps. Moreover, higher accuracy than traditional correlation filtering method and machine learning methods.

Following the protocol in [39], we compare the results of all the trackers in this paper in Table 2, which based on OPE using distance precision rate (DPR) at a threshold of 20 pixels and and overlap success rate (OSR) at an overlap threshold. This table shows that CTAD outperforms other state-of-the-art trackers in most cases, which achieves a DPR 81.1%of and OSR of 75.7%. Although the deep trackers show higher accuracy in some cases, the speed of our approach is about 8.3 times than the deep trackers. Compared with the base-line LCT, our CTAD achieves a higher improvement in many cases, especially when the scale of the target changes severely.

In the experiment, we made a qualitative evaluation of our algorithm and other trackers. Figure 8. summarizes qualitative comparisons of CTAD with six state-of-the-art

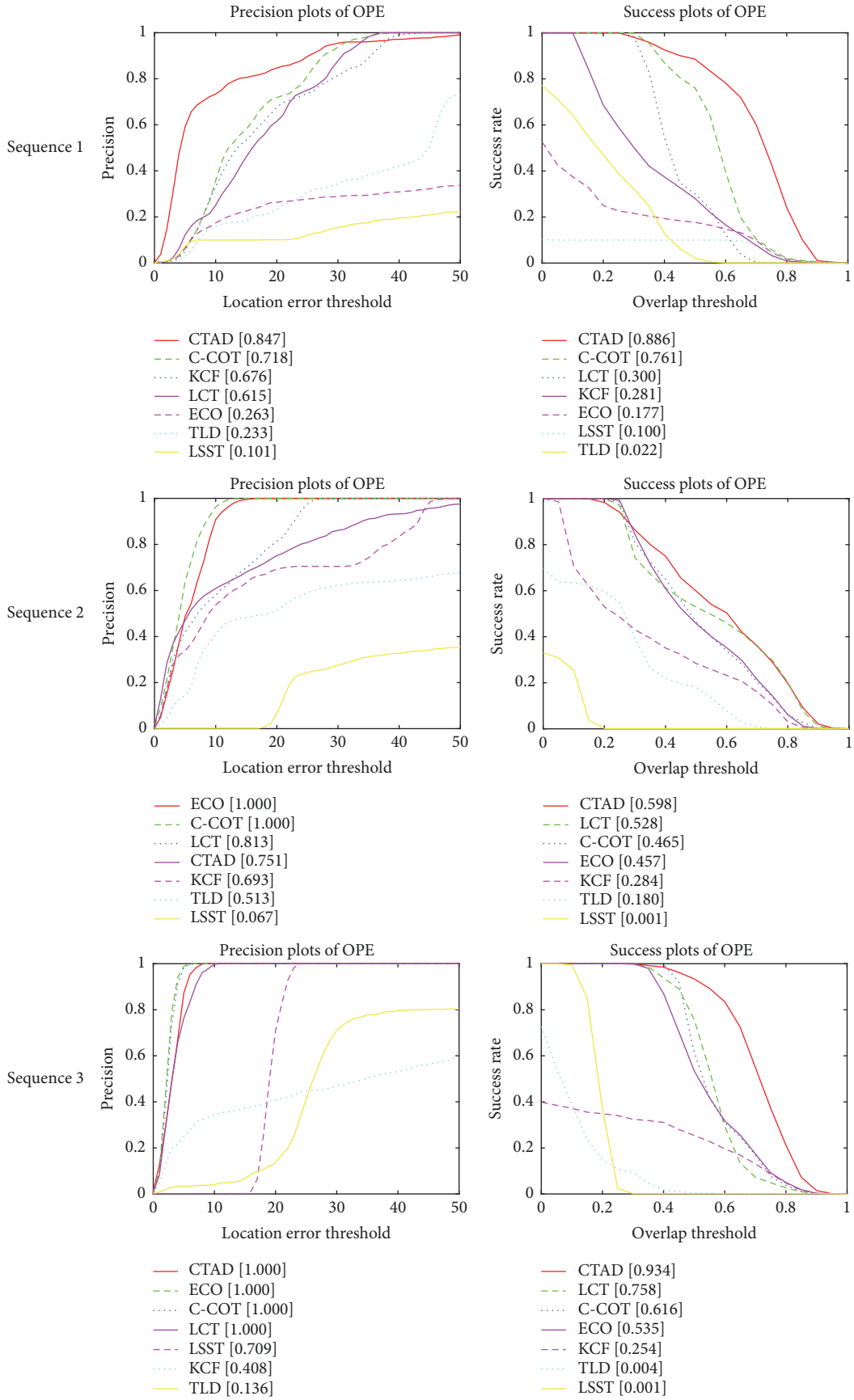


Figure 7: Continued.

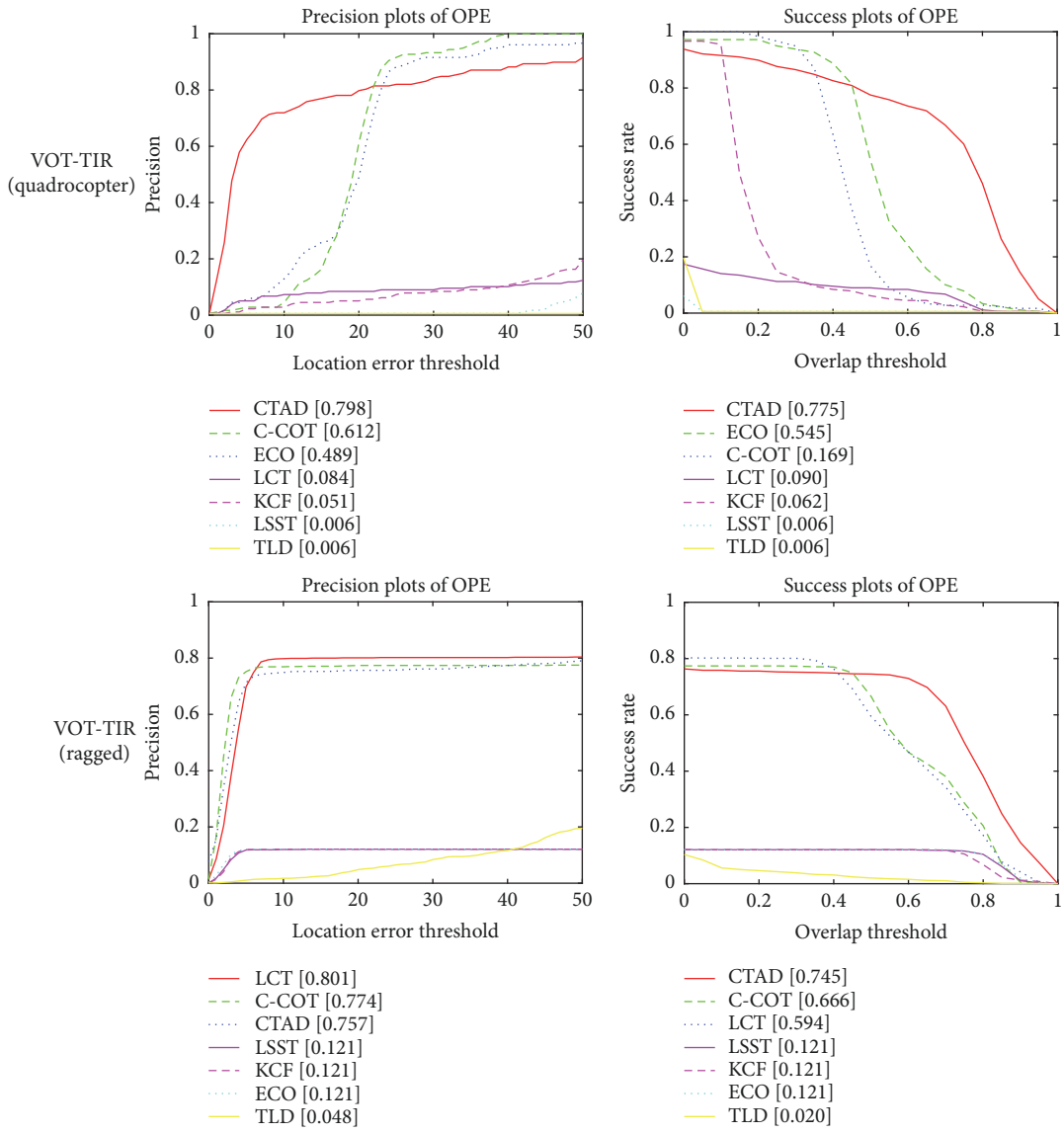


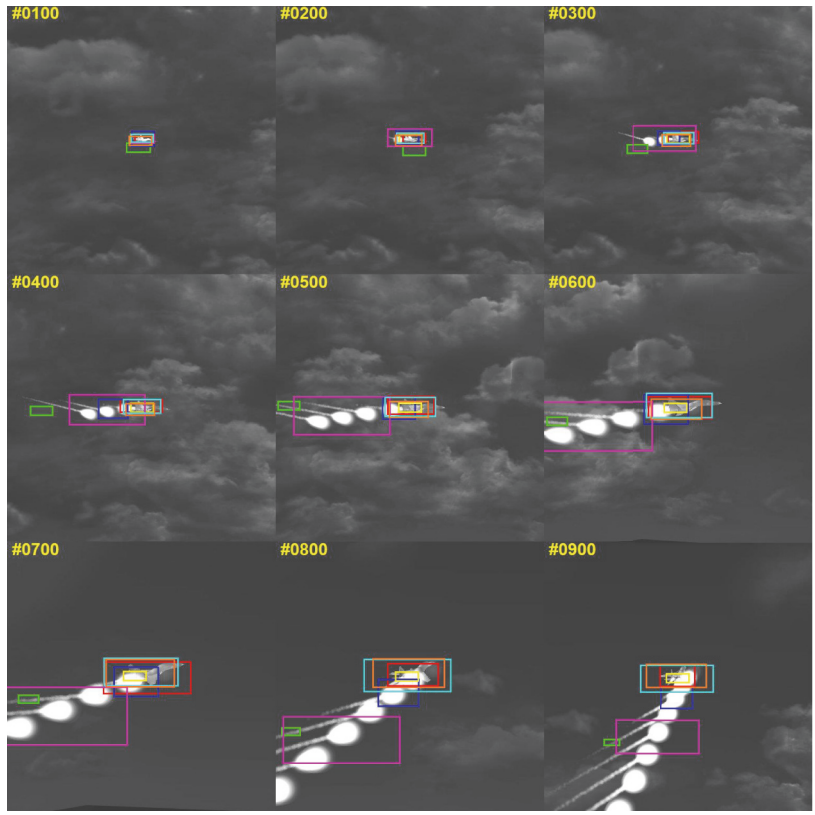
Figure 7: Comparison of DPR and OSR on 3 IR sequences and 2 sequences of VOT-TIR2016.

trackers (TLD, LSST, KCF, C-COT, ECO, LCT) on three kinds of IR image sequences from our IR image dataset.

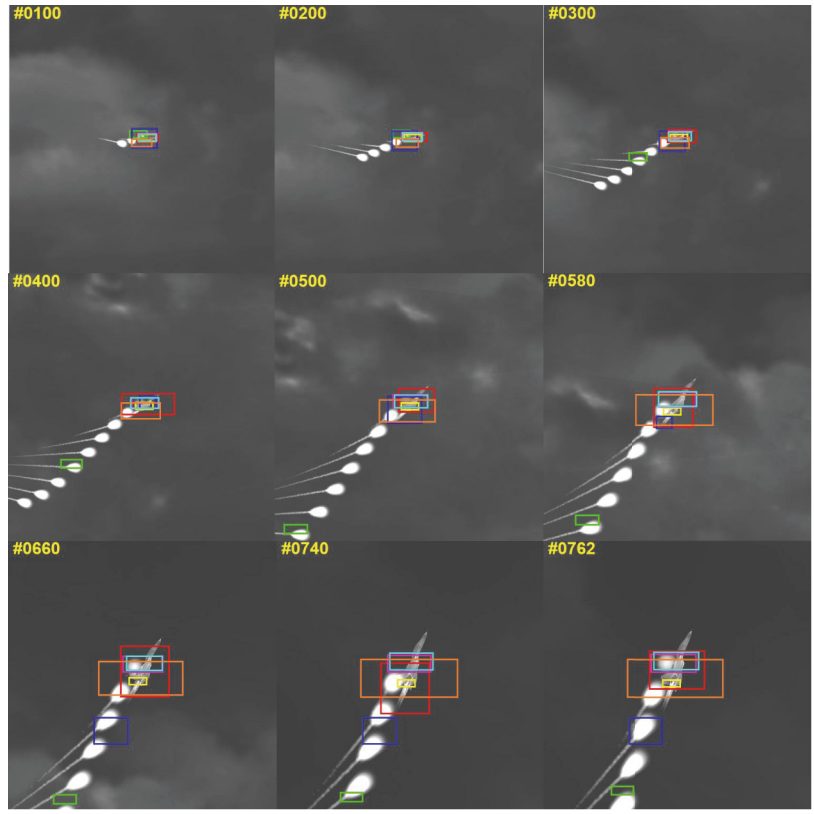
Compared with other trackers, our approach could track the aircraft more reliably. Even in the cases of the interference, complex background, fast motion. As showed in Figure 8, the precision of LCT decrease in the case of heavy scale changing. When we use YOLOv3 to verity the tracking result, the tracking precision could be improved obviously. The precision of ECO is misadventure in the case of the scale of the target changes severely, especially in the sequence 1. The bonding box of C-COT and ECO is usually much smaller than the aircraft, which will threat the tracking performance of the IR imaging missile. KCF could tracking the aircraft reliable at first, however, when the target made moves fast in the figure, it could not track the aircraft effectively.

In order to evaluate our tracker move scientifically, As illustrated in Figure 9, we compared it with the top 5 trackers of VOT-TIR2016 [40], SRDCFir [41], EBT [42], TCNN [43], Staple_TIR [44], SHCT. The tracking result of VOT-TIR2016 are public and available online (http://www.votchallenge.net/vot2016/results.html). Our algorithm has a comparable performance with the top-5 trackers from VOT-TIR2016 on DPR and OSR. As shown in Figure 8, the sequence VOT-TIR 1 which contain 178 frames, our algorithm ranking third. For the sequence VOT-TIR 2 which contain 1035 frames, it ranking first. From the data from the Table 3, we could clearly see that, our algorithm has a comparable performance with the top 5 trackers of the VOT-TIR in both accuracy and speed. Especially, when it is a long-term tracking task, our algorithm has a state-of-the-art performance compared with the best tracker of the VOT-TIR2016.

In addition, we also evaluated our algorithm in two sequences from the famous forward-looking infrared (FLIR) dataset named AMCOM. The datasets were available to us



Sequence 1



Sequence 2

Figure 8: Continued.

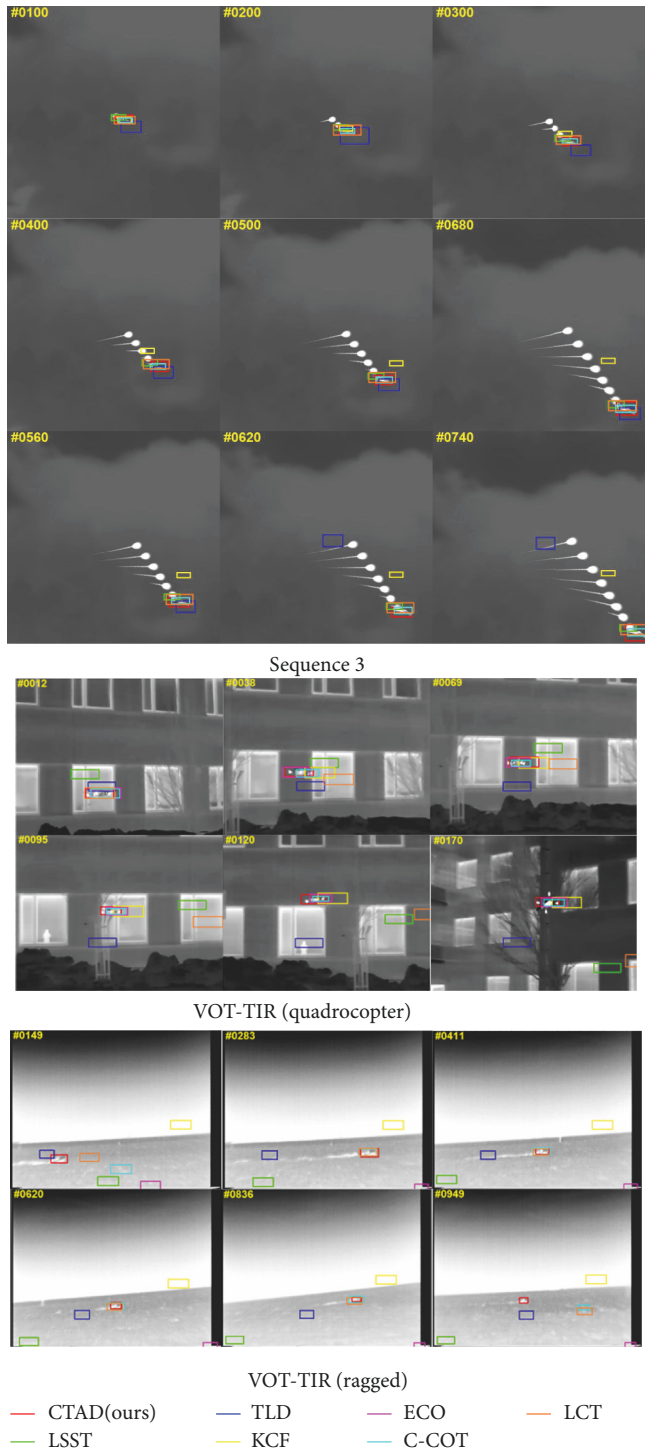


Figure 8: A visualization of tracking of our CTAD algorithm and the state-of-the-art visual trackers TLD, LSST, KCF, C-COT, and ECO on three IR image sequences and two sequences form VOT-TIR2016 dataset. The image sequences pose challenge of interference and scale variations, fast motion.

in grayscale format and each frame is 128×128 pixels. As illustrated in Figure 10, we compared the DPR and OSR in AMCOM sequences lwir_1608 and lwir_1913. Our algorithm has excellent performance on location precision and an acceptable OSR. Table 4 shows that the DPR of our algorithm ranks first in sequence lwir_1608. When it comes to sequence lwir_1913, the DPR of our algorithm ranks second. It is worth noting that the OSR of all the trackers using deep features only has ordinary performance in the assess of OSR. This could have been caused by the poor image quality as it is only 128×128. With the development of the IR technology in air-to-air missile, the image quality has been improved vastly.

Table 3: Comparisons with top 5 trackers on two sequences from VOT-TIR 2016.

		CTAD (Ours)	SRDCFir	EBT	TCNN	Staple_TIR	SHCT
VOT-TIR 1	DPR(%)	79.8	53.4	66.3	98.9	29.2	93.8
	OSR(%)	77.5	16.9	75.8	80.9	24.2	89.3
	EFO	2.17	2.48	1.99	0.76	14.25	0.91
VOT-TIR 2	DPR(%)	75.7	33.0	24.4	30.0	16.4	20.9
	OSR(%)	74.5	26.0	18.0	25.3	12.1	16.7
	EFO	2.31	2.48	1.99	0.76	14.25	0.91

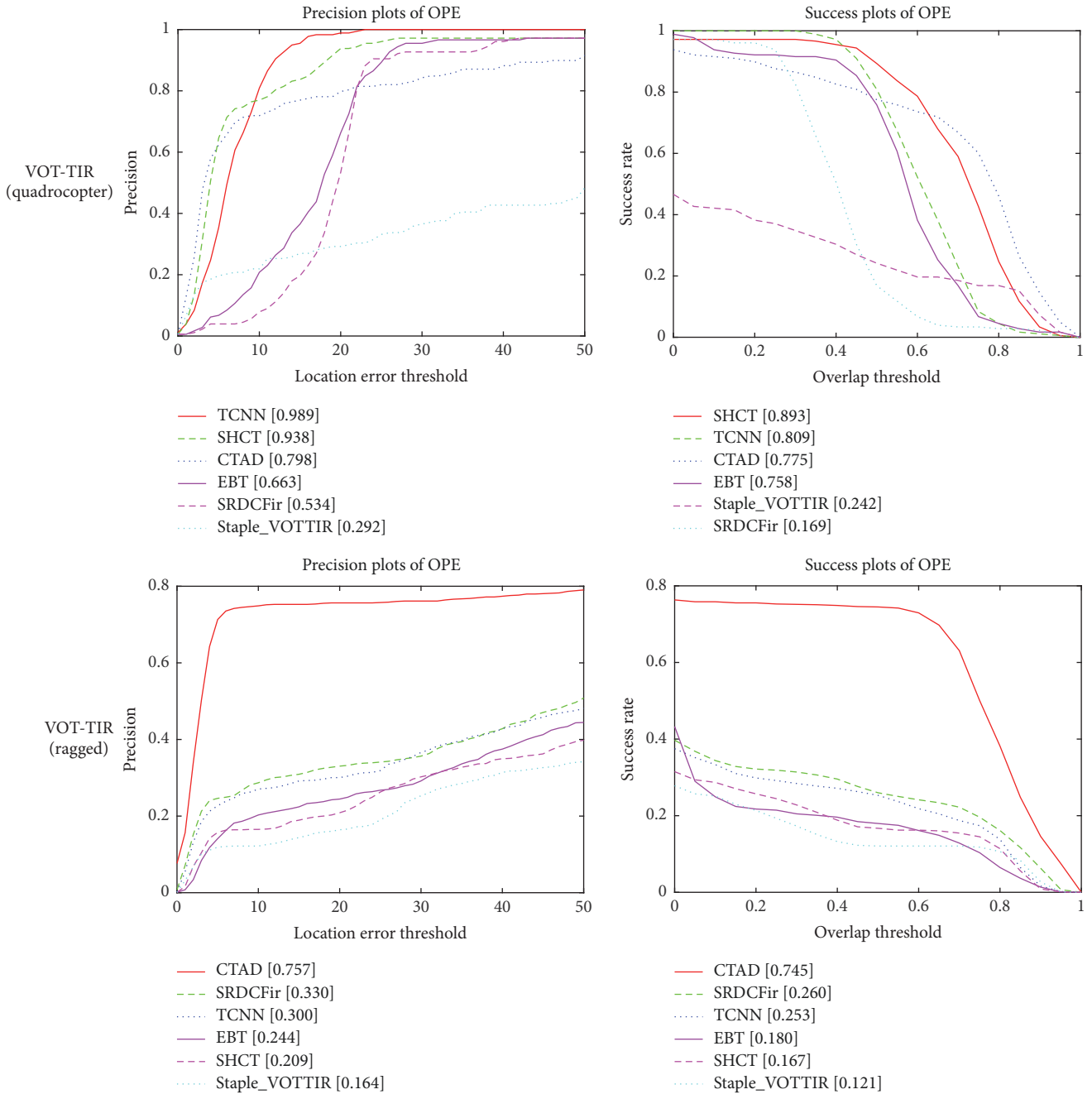


Figure 9: Comparison of DPR and OSR with the top 5 trackers of VOT-TIR2016.

Table 4: Comparisons with six excellent trackers on two sequences from AMCOM.

		CTAD (Ours)	TLD	LSST	KCF	C-COT	ECO	LCT
lwir_1608	DPR(%)	100	85.2	84.1	68.3	100	100	94.5
	OSR(%)	29.3	0.3	73.8	1.7	89.3	97.6	13.8
	Speed (fps)	15.3	168	11.7	2448	0.3	2.3	16.4
lwir_1913	DPR(%)	97.0	91.3	99.6	60.0	83.0	82.3	63.0
	OSR(%)	38.5	0.4	48.7	51.7	41.9	48.7	22.6
	Speed (fps)	17.9	172	13.1	2432	0.3	2.3	17.5

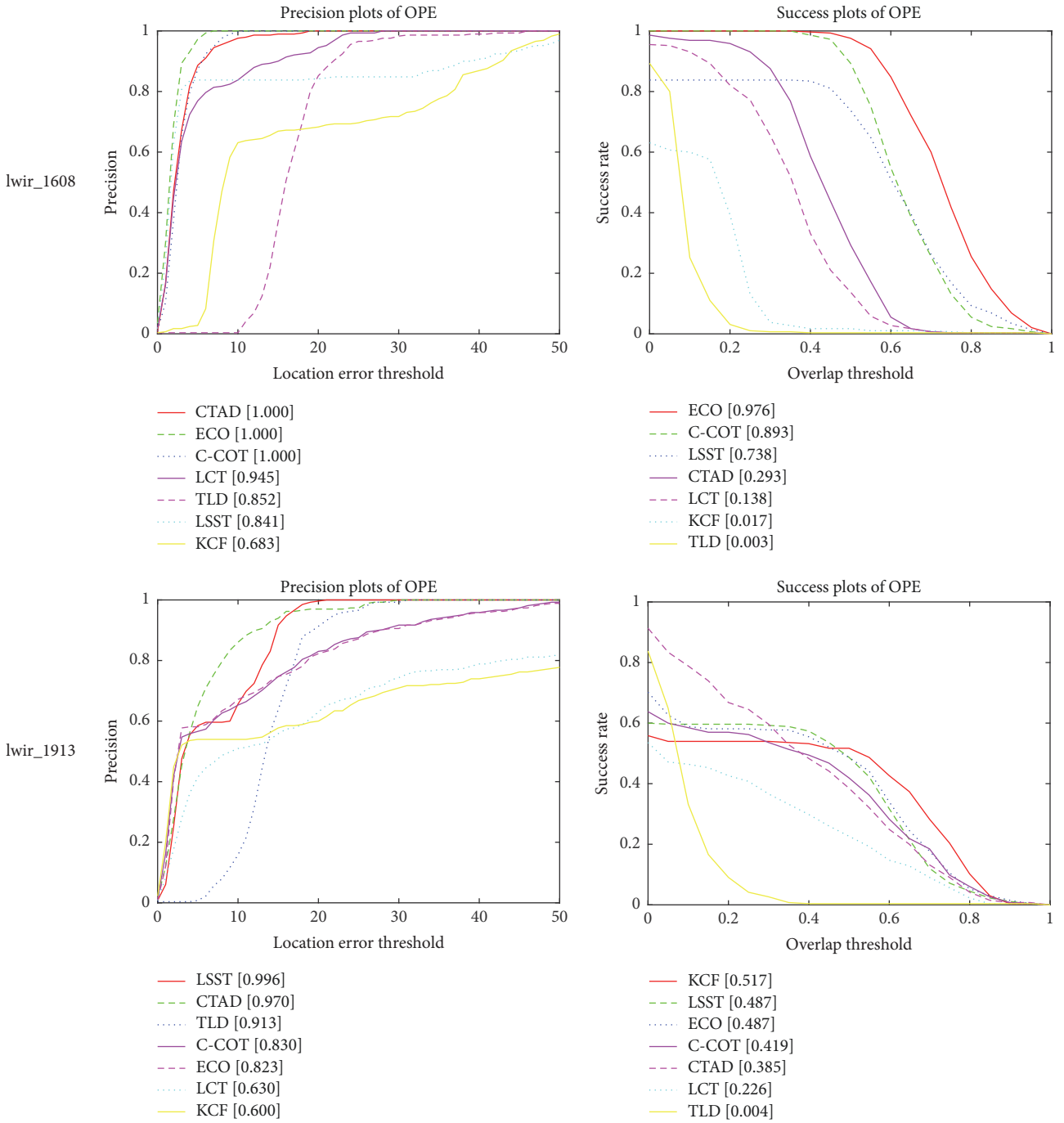


Figure 10: Comparison of DPR and OSR with six excellent trackers on two sequencs form AMCOM.

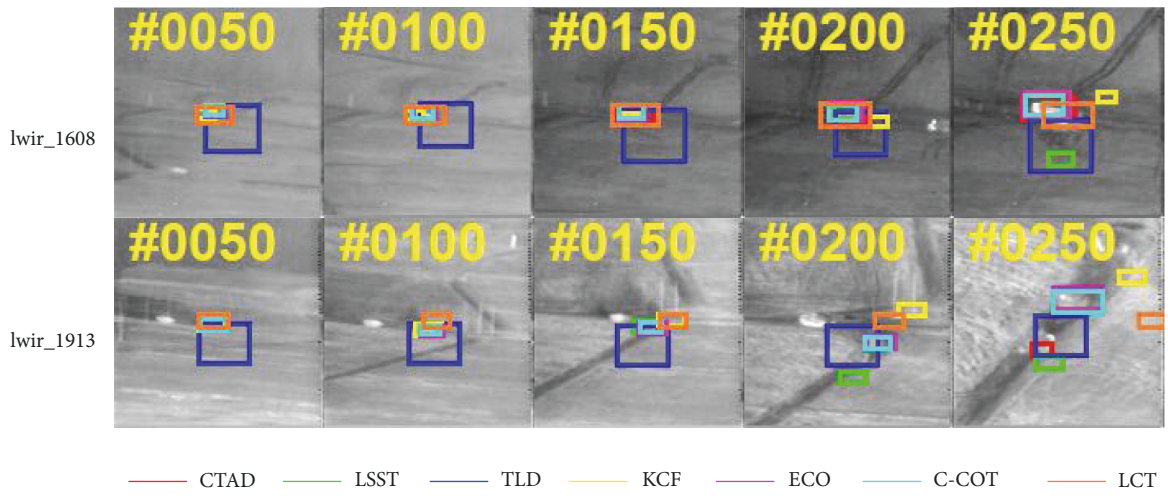


Figure 11: A visualization of tracking result of our algorithm and 6 kinds of excellent trackers on two sequences from the AMCOM.

In the experiment, we also made a qualitative evaluation of our algorithm and other trackers. Figure 11 summarizes qualitative comparisons of CTAD with six state-of-the-art trackers on two sequences from AMCOM dataset. In sequence lwir_1608, our algorithm could track the target reliably all the time. When it comes to the sequence lwir_1913, the SNR of the target decreased severely starting from frame 139 as the target made a turn. The low SNR resulted in the tracking failure of the tracker LCT, but our algorithm could still track the target because of the detecting mechanism. The detecting mechanism plays an important role in the case of tracking failure.

4.4. Detailed Analysis of CTAD. CTAD is composed of the tracking T and the detector D. T is required for CTAD to be efficient and accurate most of the time, which has a crucial influence on the performance of CTAD. In our approach, we choose LCT as the tracker based on a large amount of experiment in the IR image sequence. The YOLOv3 is a state-of-the-art detecting method based on the deep learning. Through the study of performance related to different value of the threshold, the Δn, with the development of the deep learning and the visual tracking, it is possible for us to change both the tracking T and the detector D. As a result, the structure of our approach is flexible, which is suitable for updating in the following research.

As an air-to-air weapon system, the computational capability was severely restricted by the limited space of the weapon. However, the real-time performance is quite crucial for the missile. As a result, we need to balance the relationship between the accuracy and the real-time performance. In our approach, we use a structure combining both deep learning and correlation filters, which could employ both the accuracy of the deep learning method and the high efficiency of the correlation filters. In the research filed of aerial target tracking, the combination framework is a high-efficiency solution. Our research is of great significance for the development of the IR imaging missile.

5. Conclusions

In this paper, we focus on the aerial target tracking for the IR imaging missile. we propose an effective algorithm for infrared target tracking under complex environment. Our algorithm CTAD decomposes target tracking task into two parts, the tracker T and the detector D. We have evaluated our algorithm with many classical methods in our IR sequences and two sequences from the VOT-TIR dataset. In addition, we also present the experiments performed on an AMCOM dataset of the proposed tracking algorithm. The result shows a comparable performance with the deep tracker, while running at a fast speed of about 18.1 fps. The superior result is due to the advantage of both the deep learning and correlation. Extensive experimental results show that the proposed algorithm performs favorably against the state-of-the-art methods in terms of efficiency, accuracy, and robustness.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (61703337) and the Aerospace Science and Technology Innovation Fund of China (SAST2017-082).

References

[1] A. Liaghat and M. A. Masnadi-Shirazi, "Aerial target detection and tracking in infrared image sequences by using morphological operations Kalman filtering and elliptical representation,"

in *Proceedings of the 24th Iranian Conference on Electrical Engineering, ICEE '16*, pp. 1841–1847, Iran, 2016.

[2] J. Leitloff, D. Rosenbaum, F. Kurz, O. Meynberg, and P. Reinartz, "An operational system for estimating road traffic information from aerial images," *Remote Sensing*, vol. 6, no. 11, pp. 11315–11341, 2014.

[3] Y. Li, S. Liang, B. Bai, and D. Feng, "Detecting and tracking dim small targets in infrared image sequences under complex backgrounds," *Multimedia Tools and Applications*, vol. 71, no. 3, pp. 1179–1199, 2014.

[4] H. Xin and S. Tang, "Target detection and tracking in forward-looking infrared image sequences using multiscale morphological filters," in *Proceedings of the International Symposium on Image & Signal Processing & Analysis*, 2007.

[5] Y. Cao, G. Wang, D. Yan, and Z. Zhao, "Two algorithms for the detection and tracking of moving vehicle targets in aerial infrared image sequences," *Remote Sensing*, vol. 8, no. 28, 2015.

[6] U. Braga-Neto, M. Choudhary, and J. Goutsias, "Automatic target detection and tracking in forward-looking infrared image sequences using morphological connected operators," *Journal of Electronic Imaging*, vol. 13, no. 4, pp. 802–813, 2004.

[7] F. S. Marvasti, M. R. Mosavi, and M. Nasiri, "Flying small target detection in IR images based on adaptive toggle operator," *IET Computer Vision*, vol. 12, no. 4, pp. 527–534, 2018.

[8] M. Liu, Z. Huang, Z. Fan, S. Zhang, and Y. He, "Infrared dim target detection and tracking based on particle filter," in *Proceedings of the 36th Chinese Control Conference, CCC '17*, pp. 5372–5378, China, 2017.

[9] M. Kristan, A. Eldesokey, and Y. Xing, "The visual object tracking vot2017 challenge results," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 1949–1972, 2017.

[10] J. Choi, H. J. Chang, T. Fischer et al., "Context-aware deep feature compression for high-speed visual tracking," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '18)*, pp. 479–488, Salt Lake City, UT, USA, 2018.

[11] D. Wang, H. Lu, and M.-H. Yang, "Least soft-threshold squares tracking," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 2371–2378, June 2013.

[12] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409–1422, 2012.

[13] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015.

[14] C. Ma, X. Yang, C. Zhang, and M.-H. Yang, "Long-term correlation tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR '15*, pp. 5388–5396, 2015.

[15] M. Danelljan, A. Robinson, F. S. Khan, and M. Felsberg, "Beyond correlation filters: learning continuous convolution operators for visual tracking," in *Proceedings of the European Conference on Computer Vision*, pp. 472–488, 2016.

[16] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, "ECO: efficient convolution operators for tracking," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR '17*, pp. 6931–6939, 2017.

[17] A. Sanna, B. Pralio, F. Lamberti, and G. Paravati, "A novel ego-motion compensation strategy for automatic target tracking in FLIR video sequences taken from UAVs," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 2, pp. 723–734, 2009.

[18] H. Xin, "A novel infrared target detection and tracking algorithm based on morphological filters," in *Proceedings of the International Workshop on Information Security and Application*, pp. 260–263, 2009.

[19] K. Shafique, N. Lobo, X. Li, and T. Olson, "Target-tracking in flir imagery using mean-shift and global motion compensation," in *Proceedings of the Workshop on Computer Vision Beyond the Visible Spectrum*, pp. 54–58, 2001.

[20] F. Lamberti, A. Sanna, and G. Paravati, "Improving robustness of infrared target tracking algorithms based on template matching," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 2, pp. 1467–1480, 2011.

[21] N. A. Mould, C. T. Nguyen, and J. P. Havlicek, "Infrared target tracking with AM-FM consistency checks," in *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation, SSIAI '08*, pp. 5–8, 2008.

[22] W. Yang, J. Li, D. Shi, and S. Hu, "Mean shift based target tracking in FLIR imagery via adaptive prediction of initial searching points," in *Proceedings of the 2nd International Symposium on Intelligent Information Technology Application, IITA '08*, pp. 852–855, China, 2008.

[23] V. Venkataraman, G. Fan, X. Fan, and J. P. Havlicek, "Appearance learning by adaptive kalman filters for FLIR tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR '09*, pp. 46–53, 2009.

[24] A. A. Buznikov and E. Y. Énbert, "Correlation gas-filter analyzer with microprocessor signal processing," *Journal of Optical Technology C/c of Opticheskii Zhurnal*, vol. 71, no. 3, pp. 140–142, 2004.

[25] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 2544–2550, San Francisco, Calif, USA, 2010.

[26] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Computer Vision—ECCV 2012*, vol. 7575 of *Lecture Notes in Computer Science*, pp. 702–715, Springer, Berlin, Germany, 2012.

[27] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Discriminative scale space tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 8, pp. 1561–1575, 2017.

[28] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[29] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Convolutional features for correlation filter based visual tracking," in *Proceedings of the 15th IEEE International Conference on Computer Vision Workshops, ICCVW '15*, pp. 621–629, Chile, 2015.

[30] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Hierarchical convolutional features for visual tracking," in *Proceedings of the 15th IEEE International Conference on Computer Vision, ICCV '15*, pp. 3074–3082, Chile, 2015.

[31] Y. Qi, S. Zhang, L. Qin et al., "Hedged deep tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4303–4311, 2016.

[32] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '15)*, 2015.

[33] R. Girshick, "Fast R-CNN," in *Proceedings of the 15th IEEE International Conference on Computer Vision (ICCV '15)*, Santiago, Chile, 2015.

[34] S. Ren, K. He, R. Girshick, and J. Sun, "Faster rcnn: towards real-time object detection with region proposal networks," in *Proceedings of the International Conference on Neural Information Processing Systems*, pp. 91–99, 2017.

[35] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR '15*, pp. 779–788, 2015.

[36] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, (CVPR '17)*, pp. 6517–6525, 2017.

[37] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018, https://arxiv.org/abs/1804.02767v1.

[38] C. Robert, *Machine Learning, a Probabilistic Perspective*, MIT Press, 2012.

[39] Y. Wu, J. Lim, and M.-H. Yang, "Object tracking benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1834–1848, 2015.

[40] M. Felsberg and M. Kristan, "The thermal infrared visual object tracking VOT-TIR2015 challenge results," in *Proceedings of the IEEE International Conference on Computer Vision: Workshop (ICCVW '15)*, pp. 639–651, Santiago, Chile, 2015.

[41] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proceedings of the 15th IEEE International Conference on Computer Vision (ICCV '15)*, Santiago, Chile, 2015.

[42] G. Zhu, F. Porikli, and H. Li, "Beyond local search: tracking objects everywhere with instance-specific proposals," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR '16*, 2016.

[43] H. Nam, M. Baek, and B. Han, "Modeling and propagating cnns in a tree structure for visual tracking," 2016, https://arxiv.org/abs/1608.07242v1.

[44] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. S. Torr, "Staple: complementary learners for real-time tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR '16*, 2016.