Tracking Tiny Drones against Clutter: Large-Scale Infrared Benchmark with Motion-Centric Adaptive Algorithm

Supplementary Material

1.1. Datasets

1.1.1. TDTIV Dataset(ours)

We have established a benchmark for tiny drone tracking in infrared videos (TDTIV). For anonymity, the TDTIV dataset will be released after the paper is accepted. In the data folder, we provide two sample videos to give you a more complete understanding of the contents of the TDTIV dataset. Based on the definitions in [1], we conducted a statistical analysis of the videos in the TDTIV test set according to the following attributes: (1) Out-of-View (OV) indicates that the target has moved outside the camera's field of view. (2) Occlusion (OC) refers to the target being partially or fully obstructed. (3) Thermal Crossover (TC) occurs when the target exhibits a temperature similar to other objects or the background. (4) Fast Motion (FM) signifies that the target's movement exceeds 60 pixels between consecutive frames. (5) Scale Variation (SV) occurs when the bounding box ratio between the initial frame and the current frame falls outside the range of [0.66, 1.5]. (6) Dynamic Background Clutter (DBC) refers to the presence of dynamic changes in the background surrounding the target. Table 1 represents the number of videos for different attributes in the TDTIV test set.

OV	OC	TC	FM	SV	DBC
49	13	82	53	57	61

Table 1. Number of videos per attribute in the TDTIV test set.

1.1.2. Anti-UAV410 Dataset

To validate the effectiveness of our proposed method, we conducted tests on the publicly available Anti-UAV410 counter-drone dataset. You can access detailed information about this dataset via the following link: https://github.com/HwangBo94/Anti-UAV410.

1.2. Robustness Analysis

To better assess the performance of our method in different scenarios, we manually classified the proposed TDTIV dataset into four categories: sky, buildings, hills, and forests. For each category and each attribute, we conducted experiments and compared the results with state-of-the-art (SOTA) models to assess the robustness of our approach.

As shown in Table 2, our method, MCATrack, outperforms the state-of-the-art drone tracker, SiamDT, in terms of SA values across all four scenarios. It demonstrates exceptional robustness in all categories, performing particularly well in sky and hill backgrounds, significantly surpassing the other two methods. Furthermore, as observed in Table 3, our method achieves the best performance across all six attributes. This indicates that our approach is highly adaptable to complex backgrounds and conditions, enabling more stable target tracking.

Methods	sky	buildings	hills	forests
GlobalTrack	32.34	28.62	62.33	59.84
SiamDT	33.36	32.69	64.18	62.46
MCATrack	44.94	39.74	68.14	65.04

Table 2. SA values for different backgrounds in TDTIV dataset.

Methods	OV	OC	TC	FM	SV	DBC
GlobalTrack	44.19	47.70	48.59	44.81	49.65	55.73
SiamDT	46.47	49.19	51.67	47.11	52.92	58.48
MCATrack	51.40	57.20	56.73	52.19	56.67	60.99

Table 3. SA values for various attributes in TDTIV dataset.

1.3. Long-term Tracking Performance Analysis

To verify the stability of our method over long-term tracking, we conducted an analysis of the time-averaged error for two tracking algorithms: Stark and MCATrack. Figure 1 shows the average error of these algorithms over time. The x-axis represents the frame index, ranging from 0 to 1400, indicating the progression of video frames. The y-axis represents the error in pixels, ranging from 0 to 200, with higher values indicating greater deviation from the true position. For the Stark algorithm (depicted on the left), the error starts at a relatively low value and exhibits a gradual increase over time. This upward trend suggests that the Stark algorithm experiences cumulative errors or tracking drift over extended periods, making it less reliable for long-term tracking tasks. In contrast, our MCATrack algorithm (shown on the right) demonstrates a more stable performance. The error remains relatively consistent throughout the sequence, fluctuating only slightly, indicating that MCATrack maintains its accuracy better over time. This stability suggests that MCATrack is more robust for long-term tracking applications.

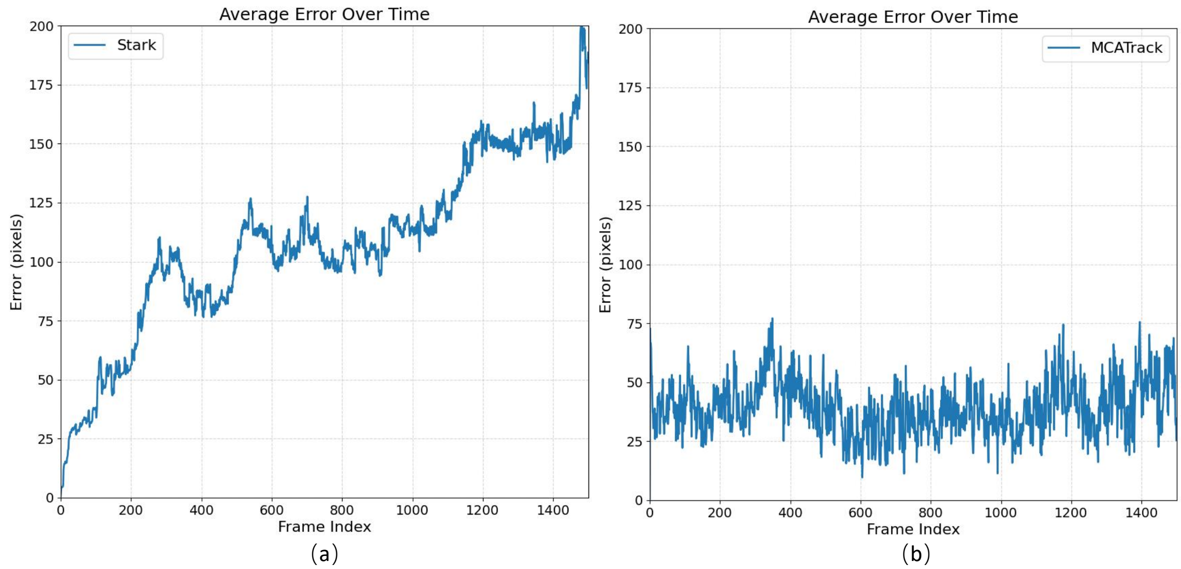

Figure 1. Average error over time comparison of Stark and MCATrack on the TDTIV dataset.

1.4. Attribution Evaluation on Anti-UAV410

To validate the effectiveness of our method in various scenarios, we evaluated it based on the different attributes defined by Anti-UAV410, along with the precision plot and the success plot for each attribute, as shown in Figure 2. Our method demonstrates outstanding performance across several attributes, particularly in handling occlusions. We observe a significant increase in precision by 16.6% and a notable improvement in the success rate by 8.3%. By using motion information to compensate for occluded parts, we prevent target loss, an issue that traditional methods struggle to address during occlusion. Furthermore, our method exhibits strong robustness in other challenging conditions. Whether dealing with dynamic background changes, rapid target movement, or scale variation, our method accurately extracts target motion features while suppressing background noise. As a result, it maintains stable performance in complex tracking scenarios. Evaluations of these attributes confirm that our method achieves high precision and success rates, validating its robustness and adaptability under diverse conditions.

1.5. Size Evaluation on Anti-UAV410

To evaluate the performance of our method with tiny targets, we tested it on various scales within the Anti-UAV410 dataset, with results shown in Figure 3. The target size is defined as the diagonal length of the bounding box. To distribute the different scaling properties as evenly as possible, we divided the size range into four intervals: tiny [2, 10), small [10, 30), medium [30, 50), and normal [50, inf]. For tiny targets, we observed a significant 7.1% increase in precision and a notable 2.9% improvement in success scores compared to SiamDT, further demonstrating our method's effectiveness in handling tiny targets.

1.6. Visualizing Motion Information

In order to better showcase the effectiveness of our proposed motion information module in enhancing the contrast between targets and the background and effectively suppressing background noise, we provide two processed examples of the TDTIV dataset in the magno_data folder, which correspond to two original examples in the data folder with the same naming convention. For a more intuitive view of the processed results, we display the original image, extracted motion information, and the combined image as Figure 4. It is evident that our method adeptly extracts motion information of small targets in complex backgrounds and effectively suppresses background noise.

References

[1] Bo Huang, Jianan Li, Junjie Chen, Gang Wang, Jian Zhao, and Tingfa Xu. Anti-uav410: A thermal infrared benchmark and customized scheme for tracking drones in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 1

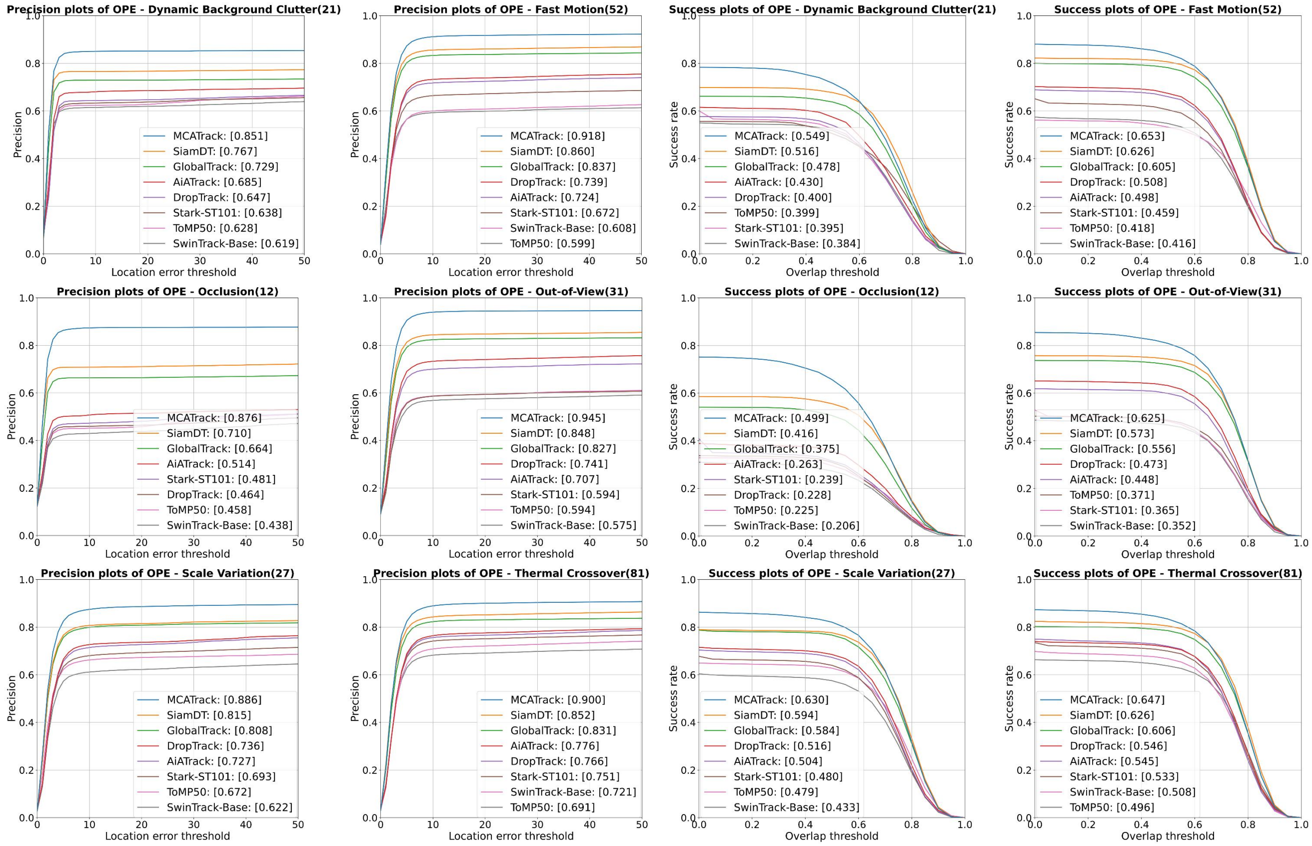

Figure 2. Attribution Evaluation on Anti-UAV410: the precision plot (left) and success plot (right) are presented. For the best viewing experience, it is recommended to view the plots in color and at an enlarged scale.

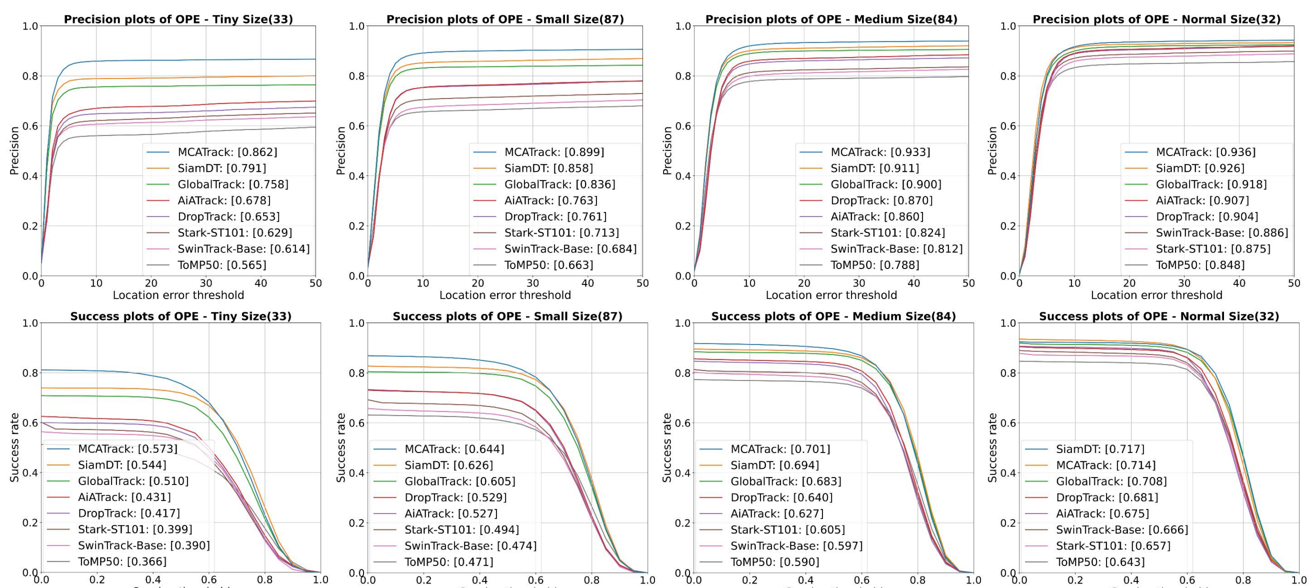

Figure 3. Size Evaluation on Anti-UAV410. the precision plot (top) and success plot (bottom) are presented. For the best viewing experience, it is recommended to observe the plots in color and at an enlarged scale.

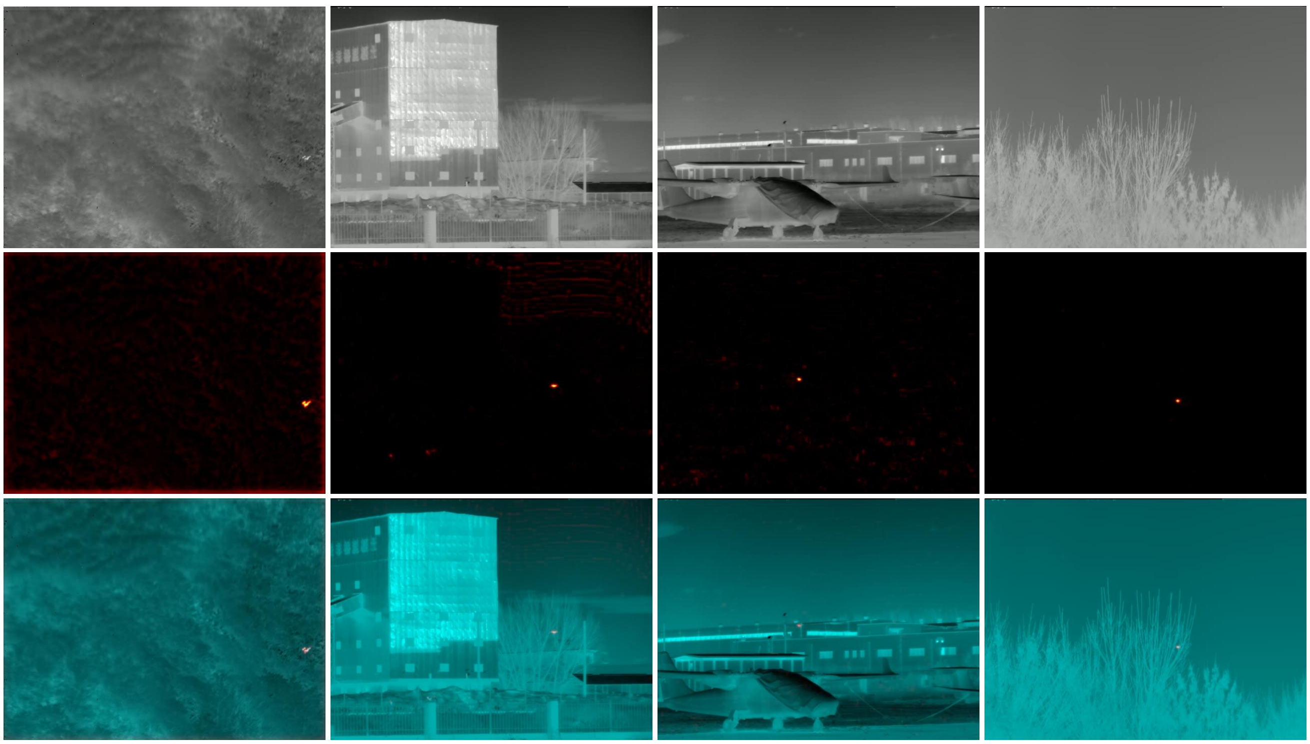

Figure 4. The first row represents the original images, the second row represents the extracted motion information, and the third row represents the final image resulting from combining the original image with the motion information. Each column corresponds to a processed image related to an original image.