Aircraft tracking in infrared imagery with adaptive learning and interference suppression

Sijie Wu, Kai Zhang, Shaoyi Li, and Jie Yan
School of Astronautics, Northwestern Polytechnical University, Xi’an, People’s Republic of China

Correspondence
Sijie Wu, School of Astronautics, Northwestern Polytechnical University, Xi’an 710072, People’s Republic of China.

E-mail: wusijie@mail.nwpu.edu.cn

Airborne target tracking is a crucial part of infrared imaging guidance. In contrast to visual tracking tasks, the target in infrared imagery shows different visual patterns. Moreover, severe background clutter and frequent occlusion caused by infrared interference make it a challenging task. Recently, discriminative correlation filter (DCF)-based trackers have shown impressive performance. However, the features adopted in DCF-based trackers are either handcrafted or pre-trained from a different task, which do not closely intertwine with the domain-specific video. To settle this problem, it is proposed to full use of online training to learn domain-specific features. By integrating the correlation filter layer into the convolutional neural networks, the feature domain and the response maps of the DCF can be optimized iteratively in the initial frame. Meanwhile, utilizing the measurement of the response maps’ peak strength, further adjustments to the feature domain can be made to achieve a sharper peak and suppress the interference region during the tracking process. Evaluations are conducted to prove the validity of proposed aircraft-tracking algorithm.

Introduction: Airborne target tracking based on infrared technology is capable of working in various weather conditions. However, infrared images of the airborne target are often presented with low signal-to-noise ratios. In the meantime, the tracking process is accompanied by frequent interference caused by cloud or infrared decoys, as shown in Figure 1, making it a challenging task to maintain a robust aircraft tracking.

Recently, DCF-based trackers have achieved great success due to their robustness and high-speed performance. Feature representations play a critical role in DCF-based trackers. Bohm et al. [1] firstly apply correlation filter to visual tracking and adopt single-channel grayscale as feature representations. Henriques et al. [2] extend the work of [1] by employing the multi-dimensional HOG features and achieve significant improvement. With the integration of convolutional neural networks (CNN) features, DCF based trackers have achieved a remarkable increase. Danelljan et al. propose a continuous formulation to fuse multi-resolution CNN feature maps [3], achieving outstanding performance in the Visual Object Tracking Challenge.

For thermal infrared object tracking, Liu et al. [4] combine multi-layer CNN features extracted from the pre-trained VGG-Net with kernelized correlation filters (KCF) trackers, and the response maps from multiple trackers are fused to consolidate the performance of the ensemble tracker. To obtain richer feature representations, Li et al. [5] design a Siamese CNN that integrates multi-level features, and propose a spatial-aware network to further improve the performance. Further, they propose to learn two complementary feature models composed of infrared-specific discriminative features and fine-grained correlation features from a larger infrared dataset, providing a performance boost [6].

Despite the integration of CNN features provides an additional performance boost, the CNN features designed for them are usually pre-trained on large datasets, which might not be optimal for aircraft tracking in infrared imagery. Motivated by the current progress of formulating the correlation filters as a differentiable layer and the training mechanism of correlation filters, we propose to combine the learning ability of CNN and high detection efficiency of kernelized correlation filters (KCF) in a unified framework to learn domain-specific features online which fit correlation filters. Compared with DCFNet [7], the proposed method does not require pre-training on large datasets. The shifted versions of the target in the initial frame are constructed as the training data. Thus the training of the network is consistent with the training of correlation filters. Therefore, the features obtained from the network are tightly coupled with both the current video domain and the correlation filters-based tracker. The optimization of the network is based on the response maps of the correlation filters, which enables tailoring feature space for the current video domain in discriminative correlation filters framework. Furthermore, we can adaptively tune the network after encountering large fluctuation in the response maps caused by background clutters or infrared decoys to suppress the interference region. Experiments are conducted to prove the tracking performance improvement in contrast to the baseline method.

Aircraft tracking algorithm: In the implementation of DCF, the correlation filters \( w \) are obtained based on the following function:

\[
\min_w ||w \cdot f(x) - y||^2 + \lambda ||w||^2, \tag{1}
\]

where \( f(x) \) denotes the feature of the image patch \( x \), \( \lambda \) refers to a regularization factor, and \( y \) stands for a Gaussian label. The solution can be derived as,

\[
F(w) = \frac{\mathcal{F}(y) \circ \mathcal{F}(f(x))^\top}{\mathcal{F}(f(x)) \circ \mathcal{F}(f(x))^\top + \lambda}, \tag{2}
\]

where \( \mathcal{F}(\cdot) \) represents the FFT, \( \circ \) denotes the element-wise product. After obtaining the correlation filters \( F(w) \) and the feature \( f(z) \) of the new frame, the aircraft’s position is estimated based on the response map \( r \), which is given by,

\[
r = \mathcal{F}^{-1}(F(w) \circ F(f(z))), \tag{3}
\]

where \( \mathcal{F}^{-1}(\cdot) \) refers to the inverse FFT. The features space \( f(\cdot) \) adopted in Equation (1) is crucial to the generation of response map. For an ideal feature space, the response maps should approximate the Gaussian labels with high response values to the target and vice versa. By integrating the correlation filter layer into the convolutional neural networks, we can take advantage of the learning ability of CNN to project \( x \) into various feature spaces, and choose a feature space that best fits the current video domain. In our formulation of the training process, the inputs of the network consist of a fixed original sample \( x_0 \) and a shifted sample \( z \). The network is implemented via Siamese architecture with tied parameters.

The training is achieved by minimizing:

\[
L(w) = \frac{1}{M} \sum_{i=1}^{M} ||r_i - y_i||^2 \tag{4}
\]

\[
= \frac{1}{M} \sum_{i=1}^{M} \left| \left| \mathcal{F}^{-1}\left( \frac{\mathcal{F}(y_i) \circ \mathcal{F}(f(x_i))^\top}{\mathcal{F}(f(x_i)) \circ \mathcal{F}(f(x_i))^\top + \lambda} \circ \mathcal{F}(f(z_i)) \right) - y_i \right| \right|^2,
\]

where \( y_i \) and \( M \) stand for the Gaussian label and the count of shifted samples. According to [7], the backpropagation gradients with respect to \( f(x_0) \) and \( f(z) \) are formulated as follows,

\[
\frac{\partial r_i}{\partial f(x_0)} = \mathcal{F}^{-1}\left( \frac{\mathcal{F}(y_i) \circ \mathcal{F}(f(z_i)) - \mathcal{F}(y_i) \circ \mathcal{F}(f(z_i)) \mathcal{F}(f(x_0))^\top}{\mathcal{F}(f(x_0)) \mathcal{F}(f(x_0))^\top + \lambda} \right), \tag{5}
\]

\[
\frac{\partial r_i}{\partial f(z_i)} = \mathcal{F}^{-1}(\mathcal{F}(w_0)), \tag{6}
\]

\[
F(w_0) = \frac{\mathcal{F}(y_i) \circ \mathcal{F}(f(z_i))^\top}{\mathcal{F}(f(z_i)) \mathcal{F}(f(z_i))^\top + \lambda}. \tag{7}
\]

Fig. 1 The aircraft-tracking process faces the challenges of poor signal-to-noise ratios and frequent interferences
After training the Siamese network in the initial frame, we adopt one branch to extract the features of subsequent frames. During the tracking process, the aircraft may encounter clouds or infrared decoys, resulting in large fluctuations in the response map. To settle this problem, we propose to finetune the network according to the response maps’ peak strength and employ peak to sidelobe ratio \([1]\) as the measurement, thereby improves the tracking performance.

The operation of interference suppression is mainly to suppress the interference region. The overall procedure of the aircraft-tracking algorithm is shown in Figure 2.

The operation of interference suppression guides the adjustment of the network. The two losses refer to the same region, and thereby improves the tracking performance. The network would be further adjusted to suppress the interference region. The two losses refer to the same region and thereby improves the tracking performance. The overall procedure of the aircraft-tracking algorithm is shown in Figure 2.

The variance of the historical frames’ PSR is defined as:

\[
\sigma_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (PSR_{i,n} - \mu_{n-1})^2.
\]

If \(\sigma_n^2\) exceeds \(\sigma_{n-1}^2\), and the PSR of the current frame is lower than the historical average, the network would be further adjusted to suppress the interference region. The overall procedure of the aircraft-tracking algorithm is shown in Figure 2.

The operation of interference suppression mainly to suppress the interference region. The overall procedure of the aircraft-tracking algorithm is shown in Figure 2.

The operation of interference suppression is mainly to suppress the interference region. The overall procedure of the aircraft-tracking algorithm is shown in Figure 2.

The operation of interference suppression is mainly to suppress the interference region. The overall procedure of the aircraft-tracking algorithm is shown in Figure 2.

The operation of interference suppression is mainly to suppress the interference region. The overall procedure of the aircraft-tracking algorithm is shown in Figure 2.

The operation of interference suppression is mainly to suppress the interference region. The overall procedure of the aircraft-tracking algorithm is shown in Figure 2.

The operation of interference suppression is mainly to suppress the interference region. The overall procedure of the aircraft-tracking algorithm is shown in Figure 2.

The operation of interference suppression is mainly to suppress the interference region. The overall procedure of the aircraft-tracking algorithm is shown in Figure 2.
suppression (KCF_ALIS), as shown in Figure 6. In frame 71, the peak strength is weakened due to the interference, showing a multimodal distribution. In subsequent frame 72, the response value of the interference region surpasses the region of the target without interference suppression. As a result, the tracking results drift to the region of the decoy. In contrast, the response value of the interference region decreases after suppressing the interference region, alleviating the model drift problem. The visualization of the tracking results is shown in Figure 7.

**Conclusion:** In this article, an aircraft-tracking algorithm based on kernelized correlation filters is presented. In order to learn features fitting the current video domain, we integrate a correlation filter layer into the KCF tracking framework. Thus we can optimize the feature domain and the response map in a unified framework. The combination of adaptive learning in the initial frame and interference suppression during the tracking process improves the tracking performance of the baseline method.

**Acknowledgements:** This research was funded by the National Natural Science Foundation of China under Grant Number 61703337 and the Aerospace Science and Technology Innovation Fund of China under Grant Number SAST2017-082.

---

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Received: 29 December 2020 Accepted: 19 April 2021
doi: 10.1049/ell2.12209

**References**

1. Bolme, D.S., et al.: Visual object tracking using adaptive correlation filters. In: *Computer Vision and Pattern Recognition*, 2010, pp. 2544–2550.
2. Henriques, J.F., et al.: High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* 37(3), 583–596 (2015)
3. Danelljan, M., et al. ECO: Efficient convolution operators for tracking. In: *Computer Vision and Pattern Recognition*, vol. 1, 2017, p. 3.
4. Liu, Q., et al.: Deep convolutional neural networks for thermal infrared object tracking. *Knowledge-Based Systems* 134, 189–198 (2017)
5. Li, X., et al.: Hierarchical spatial-aware Siamese network for thermal infrared object tracking. *Knowledge-Based Systems* 166, 71–81 (2019)
6. Liu, Q., et al.: Multi-task driven feature models for thermal infrared tracking. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 11604–11611. AAAI Press, Palo Alto (2020)
7. Wang, Q., et al.: DeNet: Discriminant correlation filters network for visual tracking. arXiv preprint, arXiv:1704.04057, 2017
8. Bertinetto, L., et al.: Fully-convolutional siamese networks for object tracking. In: European Conference on Computer Vision, pp. 850–865. Springer, Cham (2016)
9. Li, B., et al.: High performance visual tracking with siamese region proposal network. In: *Computer Vision and Pattern Recognition*, pp. 8971–8980.(2018)
10. Zhu, Z., et al.: Distractor-aware siamese networks for visual object tracking. In: European Conference on Computer Vision, pp. 101–117. Springer, Cham (2018)
11. Ma, C., et al.: Robust visual tracking via hierarchical convolutional features. *IEEE Trans. Pattern Anal. Mach. Intell.* 41(11), 2709–2723 (2018)
12. Kiani Galoogahi, H., et al.: Learning background-aware correlation filters for visual tracking. In: International Conference on Computer Vision (ICCV), pp. 1135–1143. IEEE, Piscataway (2017)
13. Kalal, Z., et al.: PN learning: Bootstrapping binary classifiers by structural constraints. In: *Computer Vision and Pattern Recognition*, pp. 49–56.(2010)
14. Hare, S., et al.: Struck: Structured output tracking with kernels. *IEEE Trans. Pattern Anal. Mach. Intell.* 38(10), 2096–2109 (2015)
15. Grabner, H., et al.: Real-time tracking via on-line boosting. In: British Machine Vision Conference, vol. 1, p. 6. BMVA Press, London (2006)
16. Wu, Y., Lim, J., Yang, M.-H.: Object tracking benchmark. *IEEE Trans. Pattern Anal. Mach. Intell.* 37(9), 1834–1848 (2015)
17. Grabner, H., Leistner, C., Bischof, H.: Semi-supervised on-line boosting for robust tracking. In: European Conference on Computer Vision, pp. 234–247. Springer, Cham (2008)