Channel Attentional Correlation Filters Learning With Second-Order Difference for UAV Tracking

Yang Zhang, Yu-Feng Yu, Ke-Kun Huang, *Member, IEEE*, and Yingxu Wang

***Abstract*— Unmanned aerial vehicle (UAV) visual tracking has been a hot research topic in the field of remote sensing. Many filter-based UAV trackers have achieved excellent performance. However, existing methods do not distinguish the importance of different feature channels with semantic information and background information, which may hinder the tracker's ability to adapt to changing environments. To deal with this problem, we propose a channel attentional correlation filters (CACFs) learning model. Specifically, we introduce the fuzzy C-means (FCMs) algorithm to preclassify the extracted features and then perform weight penalty to feature channels with different membership degrees. In addition, the filter can adapt more effectively to the background's rapid changes during the UAV tracking process by learning the second-order difference between adjacent three frame features. Finally, the comparative experiments are conducted on three mainstream UAV datasets, including DTB70, UAV123@10fps, and UAVDT. The experimental results demonstrate the effectiveness of the proposed method. The tracking performance of CACF surpasses that of other state-of-the-art trackers.**

***Index Terms*— Fuzzy C-means (FCMs), second-order difference, weight penalty.**

I. Introduction

THE discriminative filter model is a crucial element in target tracking, and its research has significantly advanced in the past decade. Many scholars have contributed to the maturity of this field, and its practical applications have expanded into society and technology. For example, the algorithms based on filtering models for tracking satellite videos in the field of remote sensing [1], [2] have achieved satisfactory results.

The discriminative tracker works by analyzing the position of the target in the initial frame and learning the relevant features from the initial frame. Then the features of the target in the following frames are extracted, and the filter is convolved with these features to generate a response map. The predicted location of the target is determined by finding the region with the highest value in the response map.

When tracking an object, there are two common difficulties that arise: boundary effects and temporal filter degradation. Several object tracking algorithms have been developed to address these difficulties. One such algorithm is the SRDCF [3] that uses a spatial regularization technique. This technique incorporates a penalty matrix to limit the filter's constraint, prevents overfitting of the training data, and enhances the tracker's generalization ability. BACF [4] enlarges the learning sample and reduces background noise interference by clipping the matrix and allowing the tracker to learn the central region of the target. LADCF [5] adopts adaptive spatial feature selection and temporal consistency constraints to achieve joint spatial–temporal filter learning. In addition, STRCF [6] incorporates temporal regularization to restrict the updating of the tracker, which increases the accuracy of the tracker and minimizes the possibility of drift. However, in the context of unmanned aerial vehicle (UAV) visual tracking, where targets undergo rapid changes, the existing trackers may struggle to perform effectively. Therefore, it is necessary to make specific improvements to enhance the tracking performance in these challenging scenarios.

In this letter, we propose a joint discriminant correlation filter (DCF) framework that incorporates channel attentional correlation filter (CACF) learning and second-order difference measurement. In particular, we use the BACF model as the benchmark approach and develop a novel tracker (named as CACF) that effectively addresses the concerns by combining accuracy, robustness, and tracking speed. As illustrated in Fig. 1, our CACF tracker is different from the tracking procedure used in BACF. We use the fuzzy c-means (FCMs) clustering algorithm [7] for preclassifying the extracted multichannel features and perform weight punishment for feature channels, which is able to distinguish the importance of different feature channels with semantic information and background information. Meanwhile, we further explore the differences in interframe features to address the challenges of drone tracking, such as target drift, occlusion, and deformation (DEF), and calculate the second-order difference between the features of adjacent frames. This allows the tracker to better adapt to the changes in the environment and target.

It should be noted that there are some trackers to assign channels with different weights for distinguishing the

Manuscript received 18 April 2023; revised 18 July 2023; accepted 30 August 2023. Date of publication 4 September 2023; date of current version 15 September 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62006056 and Grant 61976104 and in part by the Natural Science Foundation of Guangdong Province under Grant 2022A1515011595. *(Corresponding author: Yu-Feng Yu.)*

Yang Zhang and Yu-Feng Yu are with the Department of Statistics, Guangzhou University, Guangzhou 510006, China (e-mail: yuyufeng220@163.com).

Ke-Kun Huang is with the School of Mathematics, Jiaying University, Meizhou 514015, China.

Yingxu Wang is with the Shandong Provincial Key Laboratory of Network-Based Intelligent Computing, University of Jinan, Jinan 250024, China.

This article has supplementary downloadable material available at https://doi.org/10.1109/LGRS.2023.3311441, provided by the authors.

Digital Object Identifier 10.1109/LGRS.2023.3311441

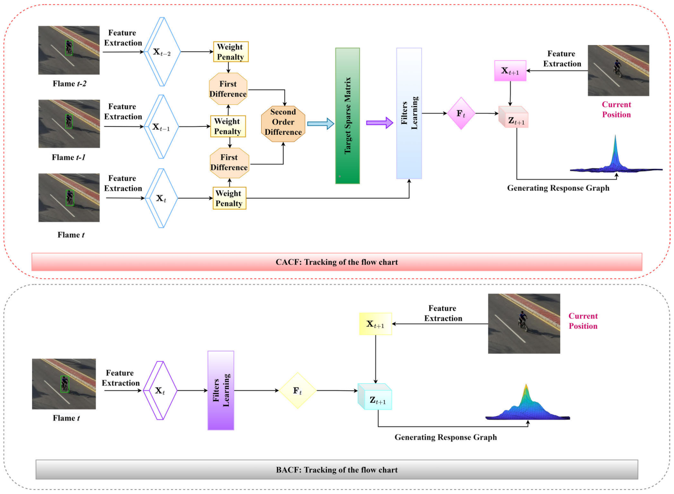


Fig. 1. Comparison of tracking processes between CACF and BACF trackers.

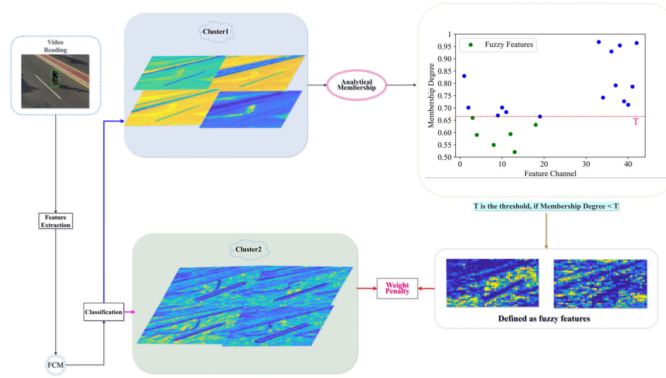


Fig. 2. Using FCM to preclassify feature channels and applying weight penalty to feature channels with membership degree below a certain threshold.

importance of different feature channels, such as JCRCF [8] and CGRCF [9]. These trackers assign different weights to each channel and learn channel weights in the DCF framework, which result in complex models and high training costs. Different from these trackers, our CACF treats the feature channels with high semantic information equally and assigns the weight penalty to the feature channels with high background information. As depicted in Fig. 2, feature channels belonging to cluster 1 contain many appearance, location, and semantic information of the target, while those belonging to cluster 2 have more background information. Assigning feature channels equal weight coefficients can adversely affect tracking accuracy. Thus, we apply weight penalty to reduce the degree to which the tracker learns from them. For the feature channels belonging to cluster 1, we only perform weigh penalty when the membership degrees are small. For cluster 2, we apply weight penalty to all the feature channels because they have more background information to reduce their impact on the tracker's learning process.

In addition, we explore the differences in interframe features to better adapt to the changes in the environment. By calculating the second-order difference between adjacent frames' features, our tracker can effectively handle challenges such as target drift and DEF.

The main contributions of our letter are summarized as follows.

1) The FCMs clustering algorithm is introduced to perform preclassification of feature channels. Using the degree of membership as a criterion, background features and fuzzy features are identified.
2) Background features and fuzzy features are subjected to weight penalty before performing the ADMM algorithm. In contrast, the weights of clear target features are higher, resulting in an enhancement of learning efficiency and tracking speed.
3) By learning the second-order difference between adjacent three-frame features, the filter is able to more effectively adapt to rapid changes in the background during the tracking process.

II. Related Work

The field of object tracking has seen significant advancements; in this section, we review some of the notable approaches in this domain.

Looking back at excellent deep learning algorithms, Siamese network-based methods [10], [11] focus on improving the generation of high-quality anchor proposals and have achieved excellent tracking performance. SE-SiamFC [12] proposes a novel approach to equip the Siamese network with an additional built-in scale to capture the natural variations in the target. ALT [13] proposes an active learning method to select and annotate unlabeled samples for training a deep CNN model. In addition, ATL effectively improves the tracking efficiency using a nearest-neighbor discrimination method to filter out isolated and low-quality samples. Yuan et al. [14] introduce the AMFT model, which integrates hand-crafted features and CNN features to tackle complex thermal infrared tracking problems. With the recent surge in popularity of Transformers, HiFT [15] and STARK [16] leverage the powerful capability of Transformers to capture global relationships, thereby enhancing and optimizing the prediction ability of trackers. This has led to significant performance improvements in the field of tracking. While deep-learning-based tracking models, supported by powerful computational resources, can achieve astonishing tracking performance through extensive training, the challenges of expensive experimental equipment and significant time costs still need to be addressed.

Overall, these algorithms represent important advancements in the field of object tracking. However, in UAV visual tracking, the tracker encounters challenging scenarios where the target undergoes rapid changes, such as fast movement, DEF, and occlusion. To achieve excellent tracking performance, it is essential to make specific improvements to the tracker. ARTracker [17] constructs a Gaussian-like function label for correlation filter training and proposes an accurate incremental update to mitigate model degradation by combining target samples with adaptive aspect ratios. IBRI [18] introduces disruptor-aware interval-based response inconsistency and a response bucketing scheme to improve the accuracy. MRCF [19] proposes the multiregularization for deviation of responses and reliability of channels to enable adaptive channel weight distributions and enhance tracking discriminability for UAV tracking and self-localization. ASTCA [20] effectively handles UAV tracking tasks by learning spatial context information and spatial–temporal context weights to reduce the impact of boundary effects.

While numerous scholars have put forward innovative ideas for UAV trackers, the majority of models fail to consider discrimination between the target and the background in filter learning and treat semantic information and background information equally. Furthermore, the significance of inter frame information in the tracking process is also overlooked. To address aforementioned problems, we propose a joint DCF framework of CACFs' learning and second-order difference measurement. CACF combines accuracy, robustness, and tracking speed by addressing the limitations of previous models. Specifically, we focus on the discrimination between the target and the background in filter learning and consider the importance of interframe information in the tracking process.

III. Proposed Method

A. Review Baseline Model

In this section, we briefly review the objective function of the baseline model BACF

$$E(\mathbf{F}) = \frac{1}{2}\left\|\sum_{d=1}^{D}\mathbf{F}^d * \mathbf{P}\mathbf{X}^d - \mathbf{Y}\right\|_2^2 + \frac{\lambda}{2}\sum_{d=1}^{D}\left\|\mathbf{F}^d\right\|_2^2 \tag{1}$$

where $\mathbf{X}^d \in \mathbb{R}^N$ and $\mathbf{F}^d \in \mathbb{R}^N$ represent the feature of the dth channel and the corresponding filter. D represents the total number of feature channels. $\mathbf{P} \in \mathbb{R}^{M\times N}(M \ll N)$ is a binary matrix to crop the central region of the feature $\mathbf{X}^{\mathbf{d}}$. $*$ denotes the correlation operator. λ is the regularization hyperparameter.

B. CACFs' Learning With Second-Order Difference

Suppose $\mathbb{X} \in R^{B\times W\times D}$ contains D feature channels with size of $B \times W$ in the tth frame. We turn the feature channels of a 3-D array $\mathbb{X} \in R^{B\times W\times D}$ into a 2-D array $\mathbf{X} \in R^{L\times D}$. Here, $L = BW$ is the dimension. Then we perform FCM algorithm to calculate the membership degree of features

$$J(\mathbf{U},\mathbf{H}) = \sum_{d=1}^{D}\sum_{j=1}^{C}u_{dj}^m\left\|\mathbf{x}_d - \mathbf{h}_j\right\|^2 \tag{2}$$

where C is the number of clusters (The value is set to 2.), and the variable m represents an exponential weighting that regulates the degree of fuzziness in the membership function. $\mathbf{h}_j$ represents the jth cluster center. $\mathbf{x}_d$ represents the dth feature channel. The element u_{dj} of $\mathbf{U}$ represents the membership degree of channel $\mathbf{x}_d$ to the clustering center $\mathbf{h}_j$. $\mathbf{h}_j$ and u_{dj} can be updated alternatively using the following equations:

$$u_{dj} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\|\mathbf{x}_d - \mathbf{h}_j\|^2}{\|\mathbf{x}_d - \mathbf{h}_k\|^2}\right)^{\frac{2}{m-1}}} \tag{3}$$

$$\mathbf{h}_j = \frac{\sum_{d=1}^{D}u_{dj}^m \cdot \mathbf{x}_d}{\sum_{d=1}^{D}u_{dj}^m}. \tag{4}$$

As defined above, $\mathbf{U}$ is a $D \times 2$ membership matrix, representing the degree of membership of D feature channels to the first and second clusters, respectively. We embed the weight penalty by comparing the elements u_{d1} and u_{d2} in the membership matrix $\mathbf{U}$. If $u_{d2} \le u_{d1}$, the dth channel belongs to cluster 1 and it has more semantic information. We adopt the following weight penalty strategy:

$$\alpha_d = \begin{cases} e^{-\frac{1}{\sigma_1}\cdot\frac{\mathrm{dis}_{d1} - \mathrm{Min}\,\mathbf{Dis}_1}{\mathrm{Max}\,\mathbf{Dis}_1 - \mathrm{Min}\,\mathbf{Dis}_1}}, & \text{if } u_{1d} \le T \\ 1, & \text{otherwise.} \end{cases} \tag{5}$$

On the other hand, if $u_{d1} \le u_{d2}$, the dth channel belongs to cluster 2 and it has more background information. We set the weight as follows:

$$\alpha_d = e^{-\frac{1}{\sigma_2}\cdot\frac{\mathrm{dis}_{d2} - \mathrm{Min}\,\mathbf{Dis}_2}{\mathrm{Max}\,\mathbf{Dis}_2 - \mathrm{Min}\,\mathbf{Dis}_2}} \tag{6}$$

where dis_{di} indicates the distance between the dth channel and the center of the ith cluster. $\mathbf{Dis}_1$ and $\mathbf{Dis}_2$ represent the distances from the feature channels to the centroids of the first and second clusters, respectively. σ_i denotes the standard deviation of $\mathbf{Dis}_i$. The threshold T is obtained by sorting the membership degree of the feature channels belonging to cluster 1 in descending order and selecting the value that ranks at the 70th percentile in the ranked sequence of membership degree. This allows the model to focus on learning essential features, thereby enhancing the model's learning efficiency and improving the accuracy of target recognition, without compromising the tracking speed.

We further combine the weight penalization mechanism with the second-order difference of features and formulate the CACF model as follows:

$$\begin{aligned} E(\mathbf{F}) = {} & \frac{1}{2}\left\|\sum_{d=1}^{D}\mathbf{F}^d * \alpha_d\mathbf{P}\mathbf{X}_t^d - \mathbf{Y}\right\|_2^2 + \frac{\lambda}{2}\sum_{d=1}^{D}\left\|\mathbf{F}^d\right\|_2^2 \\ & + \frac{\mu}{2}\left\|\sum_{d=1}^{D}\mathbf{F}^d * \mathbf{P}\left[\left(\mathbf{X}_t^d - \mathbf{X}_{t-1}^d\right) - \left(\mathbf{X}_{t-1}^d - \mathbf{X}_{t-2}^d\right)\right]\right\|_2^2 \end{aligned} \tag{7}$$

where $\mathbf{X}_{t-1}^d$ and $\mathbf{X}_{t-2}^d$ represent the dth features of the $t-1$th frame and the $t-2$th frame, respectively. During the model's updating process, we also carefully consider the details by applying a target-sparse binary matrix to highlight the background regions in the second-order difference term. By learning the second-order difference regularization term of target sparsity features, the model becomes more responsive to rapid background changes, thus improving the tracking stability and adaptability.

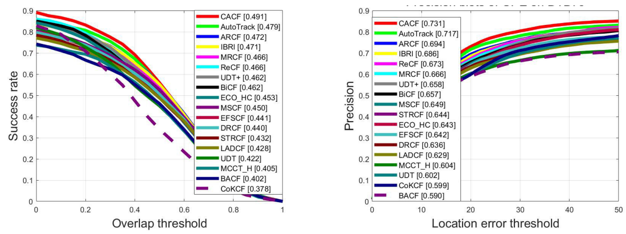
IV. Experiments

In this section, we conduct comparative experiments with the state-of-the-art trackers on multiple challenging datasets to demonstrate the efficacy and superiority of our proposed CACF. The proposed CACF tracker is evaluated on several authoritative drone benchmarks: UAV123@10fps [21], UAVDT [22], and DTB70 [23]. UAV123@10fps is a series of 123 challenging high-definition video sequences captured from a low-altitude aerial view. UAVDT consists of 100 video

TABLE I

COMPARISON OF PERFORMANCE ON THE UAVDT BENCHMARK BETWEEN CACF AND OTHER EXCELLENT DEEP TRACKERS. THE TOP THREE METHODS IN PREC, SUCC, AND FPS ARE DENOTED BY DIFFERENT COLORS: RED, GREEN, AND BLUE. ⇑ INDICATES OUTPERFORMANCE AGAINST BASELINE

Tracker	MDNet	GOTURN	C-COT	ECO	CFNet	MCPF	Staple-CA	CoKCF	MCCT	D-STRCF	UDT	ASRCF	TADT	MN_ECO	LUDT	LUDT+	Ocean	SiamAPN	SE-SiamFC	HiFT	STARK	BACF	CACF
Prec	0.725	0.702	0.659	0.702	0.680	0.660	0.697	0.605	0.671	0.667	0.674	0.700	0.677	0.691	0.631	0.701	0.725	0.710	0.626	0.734	0.704	0.686	0.737⇑
Succ	0.492	0.488	0.437	0.488	0.447	0.432	0.403	0.389	0.478	0.479	0.489	0.485	0.478	0.482	0.464	0.450	0.523	0.516	0.405	0.522	0.523	0.478	0.499⇑
FPS	1.0	16.5	1.1	16.4	41.0	0.7	56.21	21.2	8.6	6.6	76.4	14.1	32.5	30.6	78.8	59.4	14.3	36.2	5.6	–	37.9	52.04	50.5



(a) (b)

Fig. 3. Success and precision plots on the DTB70 database. (a) Success plots of OPE on DTB70. (b) Precision plots of OPE on DTB70.

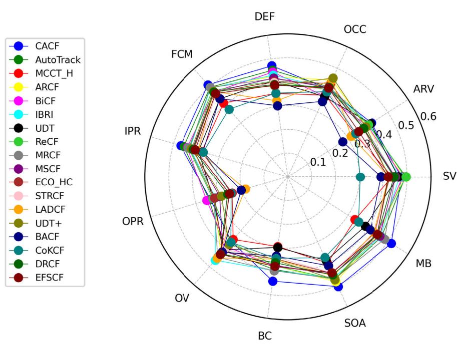
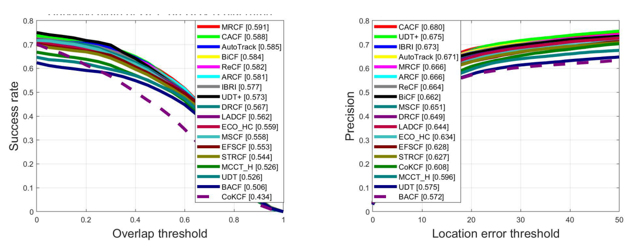


Fig. 4. Attribute-based success evaluation on the DTB70 dataset.



(a) (b)

Fig. 5. Success and precision plots on the UAV123@10fps. (a) Success plots of OPE on UAV123@10fps. (b) Precision plots of OPE on UAV123@10fps.

sequences with about 20 320 video frames. DTB70 is a highly diverse benchmark video dataset consisting of more than 70 drone videos.

A. Experimental Results and Analysis

On the DTB70 dataset, we compare our CACF tracker with other 17 advanced trackers, namely, MCCT [24], AutoTrack [25], ARCF [26], BiCF [27], IBRI [18], UDT [28], UDT+ [28], EFSCF [29], DRCF [30], MRCF [19],ReCF [31] MSCF [32], ECO [33], STRCF [6], LADCF [5], CoKCF [34], and BACF [4]. The tracking results of each algorithm are shown in Fig. 3. Among the evaluated tracking algorithms, our CACF tracker demonstrates the superior performance in both precision and success plots. Our tracker achieves an 8.9% improvement in success rate over the baseline algorithm BACF, while outperforming the baseline by 14.1% in the precision plot. Moreover, the CACF tracker is evaluated on eleven attributes, and the results in Fig. 4 show that our method outperforms other trackers. CACF obtains the highest scores for seven attributes, specifically FCM, DEF, in-plane rotation (IPR), aspect ratio variation (ARV), BC, similar objects around (SOA), and MB, while maintaining a top-three ranking across eight attributes.

In the experiments conducted on the UAV123@10fps dataset. It can be clearly understood from Fig. 5. Our tracker slightly trails MRCF in success rate. However, in terms of accuracy, CACF outperforms both MRCF and the second-place UDT+ by 1.4% and 0.5% points, respectively. Furthermore, our CACF tracker achieves remarkable improvements over the baseline approach. It outperforms BACF by 8.2% and 10.8% in terms of success and accuracy rate, respectively. These results validate the superiority and robustness of our tracking method.

B. Comparison With Deep-Based Trackers

In this section, we compare our CACF with deep tracking algorithms on the UAVDT dataset. These deep algorithms include MN_ECO [35], LUDT [36], LUDT+ [36], MDNet [37], ECO [33], GOTURN [38], Staple-CA [39], CFNet [40], TADT [41], UDT [28], DeepSTRCF [6], MCPF [42], C-COT [43], CoKCF [34], MCCT [24], ASRCF [44], Ocean [10], SiamAPN [11], SE-SiamFC [12], HiFT [15], and STARK-ST101 [16].

The detailed results are shown in Table I. The CACF tracker outperforms the second-ranked HiFT by 0.3% in tracking accuracy and improves by 5.1% compared with the baseline method. Regarding tracking success rate, CACF also exhibits remarkable performance, obtaining 2.1% increase over the baseline method. Despite the CACF tracker does not obtain a place among the top three in the FPS ranking, the real-time tracking speed of 50.5 is all-right.

V. CONCLUSION

In the article, we introduce the FCM algorithm to preclassify features into two clusters. We not only apply the weight penalty to background features in cluster 2 but also perform the weight penalty to fuzzy features in cluster 1 based on their membership degrees. In addition, filters are learned by combining the second-order differences of features between adjacent three frames. These improvements allow the tracker to adapt to rapid changes in the tracked object. Ultimately, the proposed CACF tracker demonstrates outstanding tracking performance and robustness on the DTB70, UAV123@10fps, and UAVDT datasets.

It should be noted that there are still limitations to our tracker. The tracking performance of CACF tends to decrease when dealing with scenes that have occlusion attributes. This is because when the target is occluded, the features extracted by the model are prone to noise interference, resulting in inaccurate binary classification results in FCM and further impacting the effectiveness of weight penalization. Therefore, in future work, we will focus on how to preprocess the noise-disturbed features. In addition, we will conduct in-depth research on the emerging field of multimodal tracking, which has gained prominence in recent years.

REFERENCES

[1] Y. Li and C. Bian, "Object tracking in satellite videos: A spatial–temporal regularized correlation filter tracking method with interacting multiple model," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6511105.

[2] Y. Li, C. Bian, and H. Chen, "Object tracking in satellite videos: Correlation particle filter tracking method with motion estimation by Kalman filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5630112.

[3] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proc. ICCV*, Dec. 2015, pp. 4310–4318.

[4] H. K. Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in *Proc. ICCV*, Oct. 2017, pp. 1144–1152.

[5] T. Xu, Z.-H. Feng, X.-J. Wu, and J. Kittler, "Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5596–5609, Nov. 2019.

[6] F. Li, C. Tian, W. Zuo, L. Zhang, and M.-H. Yang, "Learning spatial–temporal regularized correlation filters for visual tracking," in *Proc. CVPR*, Jun. 2018, pp. 4904–4913.

[7] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, and T.-J. Chen, "Fuzzy C-means clustering with spatial information for image segmentation," *Comput. Med. Imag. Graph.*, vol. 30, no. 1, pp. 9–15, 2006.

[8] F. Du, P. Liu, W. Zhao, and X. Tang, "Joint channel reliability and correlation filters learning for visual tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 6, pp. 1625–1638, Jun. 2020.

[9] M. Jain, A. Tyagi, A. V. Subramanyam, S. Denman, S. Sridharan, and C. Fookes, "Channel graph regularized correlation filters for visual object tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 2, pp. 715–729, Feb. 2022.

[10] Z. Zhang, H. Peng, J. Fu, B. Li, and W. Hu, "Ocean: Object-aware anchor-free tracking," in *Proc. 16th Eur. Conf. Comput. Vis. (ECCV)*. Berlin, Germany: Springer, 2020, pp. 771–787.

[11] C. Fu, Z. Cao, Y. Li, J. Ye, and C. Feng, "Siamese anchor proposal network for high-speed aerial tracking," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May/Jun. 2021, pp. 510–516.

[12] I. Sosnovik, A. Moskalev, and A. Smeulders, "Scale equivariance improves Siamese tracking," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 2765–2774.

[13] D. Yuan et al., "Active learning for deep visual tracking," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, May 10, 2023, doi: 10.1109/TNNLS.2023.3266837.

[14] D. Yuan, X. Shu, Q. Liu, X. Zhang, and Z. He, "Robust thermal infrared tracking via an adaptively multi-feature fusion model," *Neural Comput. Appl.*, vol. 35, no. 4, pp. 3423–3434, 2023.

[15] Z. Cao, C. Fu, J. Ye, B. Li, and Y. Li, "HiFT: Hierarchical feature transformer for aerial tracking," in *Proc. ICCV*, Feb. 2021, pp. 15437–15446.

[16] B. Yan, H. Peng, J. Fu, D. Wang, and H. Lu, "Learning spatial–temporal transformer for visual tracking," in *Proc. ICCV*, Oct. 2021, pp. 10448–10457.

[17] J. Chen, T. Xu, B. Huang, Y. Wang, and J. Li, "ARTracker: Compute a more accurate and robust correlation filter for UAV tracking," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6514605.

[18] C. Fu, J. Ye, J. Xu, Y. He, and F. Lin, "Disruptor-aware interval-based response inconsistency for correlation filters in real-time aerial tracking," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6301–6313, Aug. 2021.

[19] J. Ye, C. Fu, F. Lin, F. Ding, S. An, and G. Lu, "Multi-regularized correlation filter for UAV tracking and self-localization," *IEEE Trans. Ind. Electron.*, vol. 69, no. 6, pp. 6004–6014, Jun. 2022.

[20] D. Yuan, X. Chang, Z. Li, and Z. He, "Learning adaptive spatial–temporal context-aware correlation filters for UAV tracking," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 18, no. 3, pp. 1–18, 2022.

[21] M. Mueller, N. Smith, and B. Ghanem, "A benchmark and simulator for UAV tracking," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*. Berlin, Germany: Springer, 2016, pp. 445–461.

[22] D. Du et al., "The unmanned aerial vehicle benchmark: Object detection and tracking," in *Proc. ECCV*, 2018, pp. 370–386.

[23] S. Li and D.-Y. Yeung, "Visual object tracking for unmanned aerial vehicles: A benchmark and new motion models," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, no. 1, pp. 1–7.

[24] N. Wang, W. Zhou, Q. Tian, R. Hong, M. Wang, and H. Li, "Multi-cue correlation filters for robust visual tracking," in *Proc. CVPR*, Jun. 2018, pp. 4844–4853.

[25] Y. Li, C. Fu, F. Ding, Z. Huang, and G. Lu, "Autotrack: Towards high-performance visual tracking for UAV with automatic spatio-temporal regularization," in *Proc. CVPR*, Jun. 2020, pp. 11920–11929.

[26] Z. Huang, C. Fu, Y. Li, F. Lin, and P. Lu, "Learning aberrance repressed correlation filters for real-time UAV tracking," in *Proc. ICCV*, Oct. 2019, pp. 2891–2900.

[27] F. Lin, C. Fu, Y. He, F. Guo, and Q. Tang, "BiCF: Learning bidirectional incongruity-aware correlation filter for efficient UAV object tracking," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 2365–2371.

[28] N. Wang, Y. Song, C. Ma, W. Zhou, W. Liu, and H. Li, "Unsupervised deep tracking," in *Proc. CVPR*, Jun. 2020, pp. 1308–1317.

[29] J. Wen, H. Chu, Z. Lai, T. Xu, and L. Shen, "Enhanced robust spatial feature selection and correlation filter learning for UAV tracking," *Neural Netw.*, vol. 161, pp. 39–54, Apr. 2023.

[30] C. Fu, J. Xu, F. Lin, F. Guo, T. Liu, and Z. Zhang, "Object saliency-aware dual regularized correlation filter for real-time aerial tracking," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8940–8951, Dec. 2020.

[31] F. Lin, C. Fu, Y. He, W. Xiong, and F. Li, "ReCF: Exploiting response reasoning for correlation filters in real-time UAV tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 10469–10480, Aug. 2022.

[32] G. Zheng, C. Fu, J. Ye, F. Lin, and F. Ding, "Mutation sensitive correlation filter for real-time UAV tracking with adaptive hybrid label," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May/Jun.2021, pp. 503–509.

[33] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, "ECO: Efficient convolution operators for tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6931–6939.

[34] L. Zhang and P. N. Suganthan, "Robust visual tracking via co-trained kernelized correlation filters," *Pattern Recognit.*, vol. 69, pp. 82–93, Sep. 2017.

[35] J. Zhao, K. Dai, D. Wang, H. Lu, and X. Yang, "Online filtering training samples for robust visual tracking," in *Proc. 28th ACM Int. Conf. Multimedia*. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1488–1496.

[36] N. Wang, W. Zhou, Y. Song, C. Ma, W. Liu, and H. Li, "Unsupervised deep representation learning for real-time tracking," *Int. J. Comput. Vis.*, vol. 129, no. 2, pp. 400–418, 2021.

[37] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in *Proc. CVPR*, Jun. 2016, pp. 4293–4302.

[38] D. Held, S. Thrun, and S. Savarese, "Learning to track at 100 FPS with deep regression networks," in *Proc. ECCV*, 2016, pp. 749–765.

[39] M. Mueller, N. Smith, and B. Ghanem, "Context-aware correlation filter tracking," in *Proc. CVPR*, Jul. 2017, pp. 1387–1395.

[40] J. Valmadre, L. Bertinetto, J. Henriques, A. Vedaldi, and P. H. S. Torr, "End-to-end representation learning for correlation filter based tracking," in *Proc. CVPR*, Jul. 2017, pp. 5000–5008.

[41] X. Li, C. Ma, B. Wu, Z. He, and M.-H. Yang, "Target-aware deep tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1369–1378.

[42] T. Zhang, C. Xu, and M.-H. Yang, "Multi-task correlation particle filter for robust object tracking," in *Proc. CVPR*, Jul. 2017, pp. 4819–4827.

[43] M. Danelljan, A. Robinson, F. S. Khan, and M. Felsberg, "Beyond correlation filters: Learning continuous convolution operators for visual tracking," in *Proc. ECCV*, 2016, pp. 472–488.

[44] K. Dai, D. Wang, H. Lu, C. Sun, and J. Li, "Visual tracking via adaptive spatially-regularized correlation filters," in *Proc. CVPR*, Jun. 2019, pp. 4665–4674.