J. Vis. Commun. Image R. 110 (2025) 104458

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Full Length Article

Learning disruptor-suppressed response variation-aware multi-regularized correlation filter for visual tracking[☆]

Sathishkumar Moorthy [a,b], Sachin Sakthi K.S. [a], Sathiyamoorthi Arthanari [a], Jae Hoon Jeong [a], Young Hoon Joo [a,*]

[a] School of IT Information and Control Engineering, Kunsan National University, 558 Daehak-ro, Gunsan-si, Jeonbuk 54150, Republic of Korea
[b] Department of Artificial Intelligence and Data Science, Sejong University, Seoul 05006, Republic of Korea

ARTICLE INFO

Keywords:
Visual tracking
Correlation filters
Disruptor suppression
Temporal regularization
Channel reliability

ABSTRACT

Discriminative correlation filters (DCF) are widely used in object tracking for their high accuracy and computational efficiency. However, conventional DCF methods, which rely only on consecutive frames, often lack robustness due to limited temporal information and can suffer from noise introduced by historical frames. To address these limitations, we propose a novel disruptor-suppressed response variation-aware multi-regularized tracking (DSRVMRT) method. This approach improves tracking stability by incorporating historical interval information in filter training, thus leveraging a broader temporal context. Our method includes response deviation regularization to maintain consistent response quality and introduces a receptive channel weight distribution to enhance channel reliability. Additionally, we implement a disruptor-aware scheme using response bucketing, which detects and penalizes areas affected by similar objects or partial occlusions, reducing tracking disruptions. Extensive evaluations on public tracking benchmarks demonstrate that DSRVMRT achieves superior accuracy, robustness, and effectiveness compared to existing methods.

1. Introduction

Visual object tracking (VOT) is a fundamental task in computer vision that involves the automatic detection and continuous monitoring of objects in video sequences. With the proliferation of digital cameras and the increasing availability of high-resolution video data, object tracking has become essential in various real-world applications. From surveillance and security systems to autonomous vehicles and augmented reality, accurate and robust object tracking plays a pivotal role in enabling intelligent systems to perceive and interact with their environment effectively [1,2]. This has inspired both academia and industry to dedicate substantial attention and resources to the comprehensive research and development of VOT. The goal is to consistently and accurately estimate the state and trajectory of a target of interest in subsequent frames, relying solely on its initial state (i.e., center position, extent, and motion model). Nevertheless, there is plenty of opportunity for further development, particularly concerning these attributes. Recently, many algorithms have been presented to perform feature extraction, target localization, model update, and so on. Wherein, deep learning-based trackers exploit convolutional features from different layers and show encouraging results. However, the challenge remains significant: ensuring precise and reliable target tracking in constrained scenarios, particularly when dealing with external disruptions (e.g., partial or complete occlusion, background clutter) and internal disturbances (e.g., object deformation, rotation, or scale variation).

Researchers have shown an enormous amount of interest in DCF-based trackers thanks to their remarkable efficacy in laying the foundation for many subsequent methodologies [3–5]. Technically, in correlation filter (CF) trackers, the ridge regression model is trained using the circulant structure of a base sample. On the other hand, the DCF-based trackers employ a fast Fourier transform (FFT) to improve computational performance. The circular shift operation can produce unrealistic training samples at the extension boundaries, leading to what is known as boundary effects. These corrupted samples, resulting from unwanted boundary effects, can degrade the performance of the tracker during training. Despite continuous advancements, the tracking accuracy of DCF trackers has seen significant improvements; however, certain challenges persist: (1) Dynamic changes in the background, such as rapid object or fast motion and varying camera viewpoints, can introduce disruptive information near the tracked object. This issue manifests as secondary peaks emerging around the primary peak in the response map, which may eventually dominate, causing tracking failure. (2) A

[☆] This paper has been recommended for acceptance by Zicheng Liu.
[*] Corresponding author.
E-mail address: yhjoo@kunsan.ac.kr (Y.H. Joo).

https://doi.org/10.1016/j.jvcir.2025.104458
Received 3 November 2024; Received in revised form 20 January 2025; Accepted 8 April 2025
Available online 17 April 2025

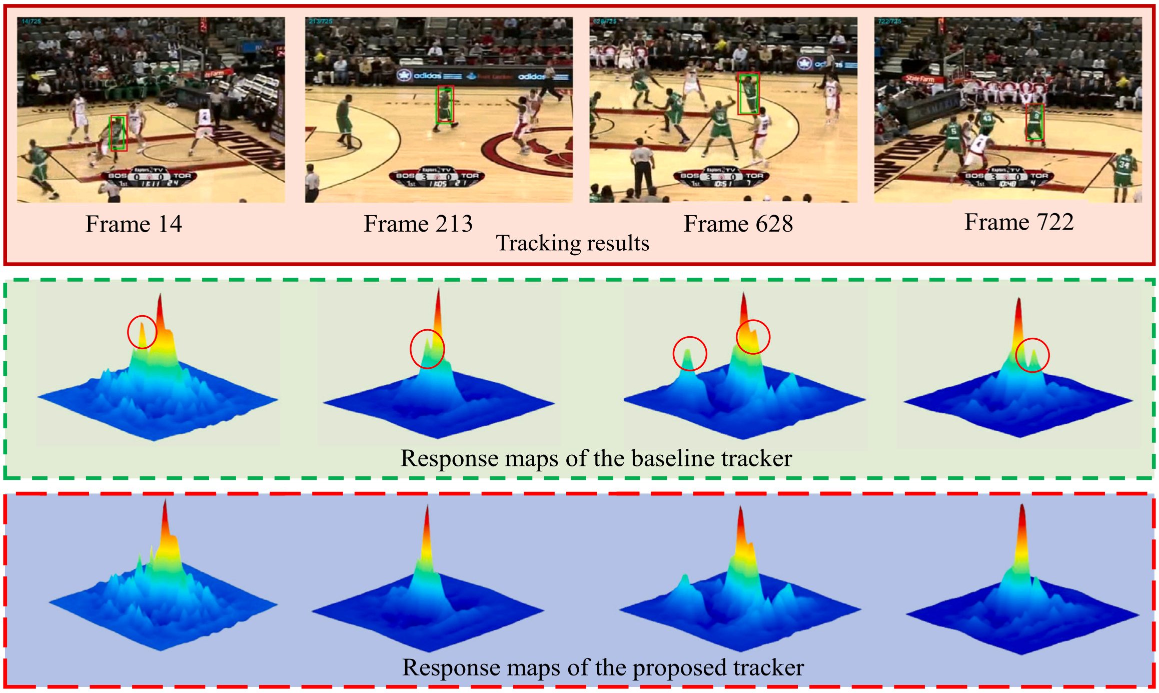


Fig. 1. The response maps and tracking results of the Basketball sequence from the OTB benchmark were compared between the proposed DSRVMRT (highlighted in red) and its baseline BACF (highlighted in green). This sequence presents a major challenge due to its significant occlusion and deformation. Our method successfully detects and suppresses the disruptors that cause occlusion-related responses, allowing for precise object relocation when it reappears.

lack of adequate focus on the tracked object frequently leads to object loss during the tracking process.

In recent years, several research directions have been explored to address these issues. Specifically, the CF trackers presented in [6,7] not only mitigate the futile boundary effects but also enhance the performance capabilities. As a development, the authors of [6] have presented an SRDCF method, which tackles the boundary effects issues and penalizes filter coefficients based on their location, reducing the impact of surrounding areas. When the object being tracked moves quickly or undergoes deformation, SRDCF may not perform as well. On the other hand, the Gauss–Seidel solving method used in SRDCF may not have the necessary computational speed for typical tracking applications. To mitigate the above-mentioned problem, Galoogahi et al. have used the binary mask in BACF tracker [7] on training the filter to generate real negative samples. Specifically, the BACF tracker incorporates negative samples from the background surrounding the target, which helps the filter distinguish the object from similar background features. By learning the difference between the object and its background variations, the BACF filter becomes more robust to changes in shape caused by deformation. Besides, the BACF tracker may update corrupted samples inaccurately, particularly during occlusion or when the target object moves outside the bounding box, which leads to tracking drift and fails to detect the target object until the end of the video sequence. To suppress the disturbance information in the background, ARCF [8] employs a binary cropping matrix by enlarging the search region, which enables the CF to be trained with actual negative samples. To be more specific, a temporal regularization component in ARCF restricts the disparity between response maps from two consecutive frames to prevent model drift. However, this regularization only incorporates information from two frames, which limits its robustness. While the inclusion of additional regularization terms helps to mitigate background noise, these approaches fail to effectively suppress disturbance region samples involved in filter training. Regarding this issue, we have proposed a disruptor-suppressed method for CF learning to handle object appearance variations more efficiently.

Earlier investigations have put forth various techniques to evaluate the excellence of response maps for example peak-to-sidelobe ratio (PSR) in GCF [1]. In the context of visual odometry, authors of [9] devised a bucketing mechanism to enhance the accuracy of estimating the vehicle's overall ego-motion and decrease the occurrence of drift rates. The existing methods do not possess the capability to adjust constraint coefficients flexibly with respect to changes in the target and background within the spatial domain. This implies that, despite being aware of the background information, these methods are prone to generating distractors which can potentially lead to model drifts and cause the tracking of the target to fail. Inspired by this, the authors of [10,11] utilized an innovative temporal consistency for UAV tracking to denoise the response maps. In order to solve the problems mentioned above, this paper proposes a disruptor-suppressed response variation-aware multi-regularized correlation filter framework for reliable and accurate tracking. With multiple regularization strategies, this proposed tracker can effectively alleviate boundary discontinuities and promote tracking performance. The suitability of the DSRVMRT tracker for robust object tracking is confirmed by its competitive performance as shown in Fig. 1. It is worth noting that the baseline method contains disruptors due to various challenges whereas the presented methodology can locate objects efficiently.

This work's key contributions are:

1. The proposed mechanism utilizes response bucketing to identify disruption areas and creates a customized penalty mask that effectively suppresses background noise, enhancing tracking performance in the presence of disruptors.
2. An innovative regularization method is introduced that smooths response fluctuations by regularizing the second-order difference, improving the tracker's sensitivity to changes in the object's appearance.
3. A channel-reliability-aware regularization is introduced to optimize channel weights alongside filters during training, allowing the tracker to focus on the most reliable channels and ultimately enhancing its performance.
4. Comprehensive experimental results on seven challenging benchmarks demonstrate that the proposed tracker outperforms state-of-the-art algorithms.

2. Related work

Despite extensive research conducted in recent decades, VOT continues to be one of the most popular subjects in the field of computer vision. This section examines closely related methods that serve as the baseline and motivation for the research presented in this paper. The review covers various aspects of tracking studies, such as DCF-based trackers, spatial and temporal information, channel reliability information, and deep learning strategies.

2.1. Tracking with DCF

DCF-based methods have achieved an elegant balance between tracking accuracy, robustness, and computational efficiency by leveraging a supervised approach to solve a linear regression problem. As a pioneering work, the authors in [3] have introduced the MOSSE tracker, which can adapt to changes in the target object's appearance over time. Besides, the MOSSE filter may struggle with distinguishing between the target object and similar-looking background elements, leading to tracking drift or failures in cluttered scenes. Following this, the author in [4] has introduced the KCF tracker, which uses rectangular correlation windows to mitigate boundary effects and improve tracking accuracy along the edges of the target object. Furthermore, the authors of [12] have introduced a novel method that utilizes a DCF tracker to achieve accurate scale estimation, leading to improved robustness against object size variations in videos. Despite the higher accuracy these trackers have, there persistently exist some issues to be resolved, such as target deterioration, boundary effects, and scale distinctions. Moreover, authors in [13] have leveraged surrounding information on the target object to mitigate the boundary effects. In spite of this, typical trackers heavily depend on training filters that only use the training patches retrieved from the present frame, which prevents them from incorporating temporal clues. The performance of these trackers tends to degrade in the presence of substantial appearance variations, scale discrepancies, and movements of the object or platform, conditions often encountered during prolonged tracking. In response to these issues, we have implemented a tracking system that incorporates disruptor-suppressed regularization, which efficiently generates a penalty mask to address problematic regions.

2.2. Tracking with spatial and temporal information

A fundamental challenge in visual object tracking lies in uncovering spatial and temporal information. Current trackers can generally be classified into two main types: those focusing solely on spatial information and those integrating both spatial and temporal dynamics. This occurrence can cause contamination of the samples and subsequently result in a degradation of performance. To address the issue mentioned earlier, the authors of [6] have introduced SRDCF, which penalizes the filter coefficients based on their spatial distance from the center of the sample. Likewise, BACF (Kiani et al. 2017) addresses the boundary effects problem by applying a binary mask to the image patches. Further, an adaptive spatial regularizer was introduced by the authors of ASRCF [14] to incorporate temporal consistency into filter training, thereby mitigating abrupt appearance variation. Moreover, the authors in [15] have presented a tracker using adaptive spatial and temporal information in the ensemble method to raise the performance. Technically, the authors of [16] incorporate multiple constraint terms into the model to enhance the resilience of tracking. Nevertheless, existing spatial regularization methods only enforce limitations on filters through basic predefined measures like binary masks, overlooking the rich diversity and redundancy inherent in the input features. Technically, a novel DIoU network-based bounding-box regression model is formulated for target tracking [17]. While preserving the advantages of the IoU network in tracking tasks, the DIoU network directly minimizes the distance between the ground-truth bounding box and the predicted bounding box, enabling the tracker to obtain more accurate tracking results. A Conjugate-Gradient-based strategy is adopted to efficiently address the optimization problem in the target classification component, allowing for efficient online processing. On the other hand, the authors in [18] propose an improved BACF tracker (TRBACF), which incorporates temporal regularization to account for the potential relationship of the moving target object in a time series, thereby enhancing tracking performance. To mitigate computational complexity, the improved alternating direction method of multipliers (ADMM) is utilized to solve the objective function in the Fourier domain. Different from the aforementioned approaches, the proposed method incorporate response variation information to mitigate the boundary effect. With the response variation calculated by the second-order difference of the filters among adjacent three frames, the proposed algorithm can robustly locate the target in complicated tracking scenarios.

2.3. Tracking with channel reliability information

The goal of channel reliability is to utilize channels that provide useful information, enhancing tracking accuracy while minimizing the impact of less valuable channels. The authors in [19] presented spatial and channel reliability approaches to boost tracking efficiency. In [20], authors proposed a group feature selection algorithm for multichannel image representations. This algorithm not only decreases dimensionality but also improves the discriminability and interpretability of the learned filters. To achieve favorable tracking accuracy, the CACF [21] technique dynamically chooses feature channels from convolutional neural network features that are representative and discriminative. On the one hand, the authors of CGRCF [22] have introduced channel regularization CF to acquire channel weights, while graph regularization is utilized to maintain the similarity of significance among distinct feature channels. On the other hand, the authors of CFRP (correlation filter with region proposal) [23] have utilized a channel regularization in CF which multiplies the weight vector with the feature channels. The tracking performance is effectively enhanced by incorporating channel-wise information in these studies. In contrast, this research introduces a novel regularization technique that takes into account the reliability of each channel and optimizes it in conjunction with the filter. By employing this reliable channel information, the target localization for filter learning is significantly refined and the channels with redundant information are effectively suppressed.

2.4. Tracking with deep convolutional neural networks

The convolutional neural network (CNN) has attracted significant attention in the field of deep learning due to its impressive feature representation capabilities. This has inspired researchers in the VOT community to abandon conventional hand-crafted features in favor of convolutional layer features, leading to promising results. Deep CNN has been integrated into VOT to enhance the tracker's efficacy, leveraging their robust feature representation capabilities [24,25]. For the first time, Ma et al. [26] leverages the hierarchical convolutional layer into the DCF methods to enhance tracking performance. In the process of filter learning, LADCF integrates spatial and temporal information to sustain a robust appearance model for effectively tracking the target. Technically, the authors in [27] employed a multi-expert approach, which implements the convolutional layer and HOG features. Recently, the SiamRAKPN [28] tracker introduced keypoint prediction-enhanced Siamese networks with self-attention for robust visual object tracking. The method integrates keypoint prediction for spatial context and a progressive heatmap refinement technique to enhance target localization accuracy in complex and challenging scenarios. By drawing inspiration from these techniques, a highly effective combination of handcrafted and deep convolutional features is utilized for the tracking task, resulting in a significant improvement in overall tracking performance.

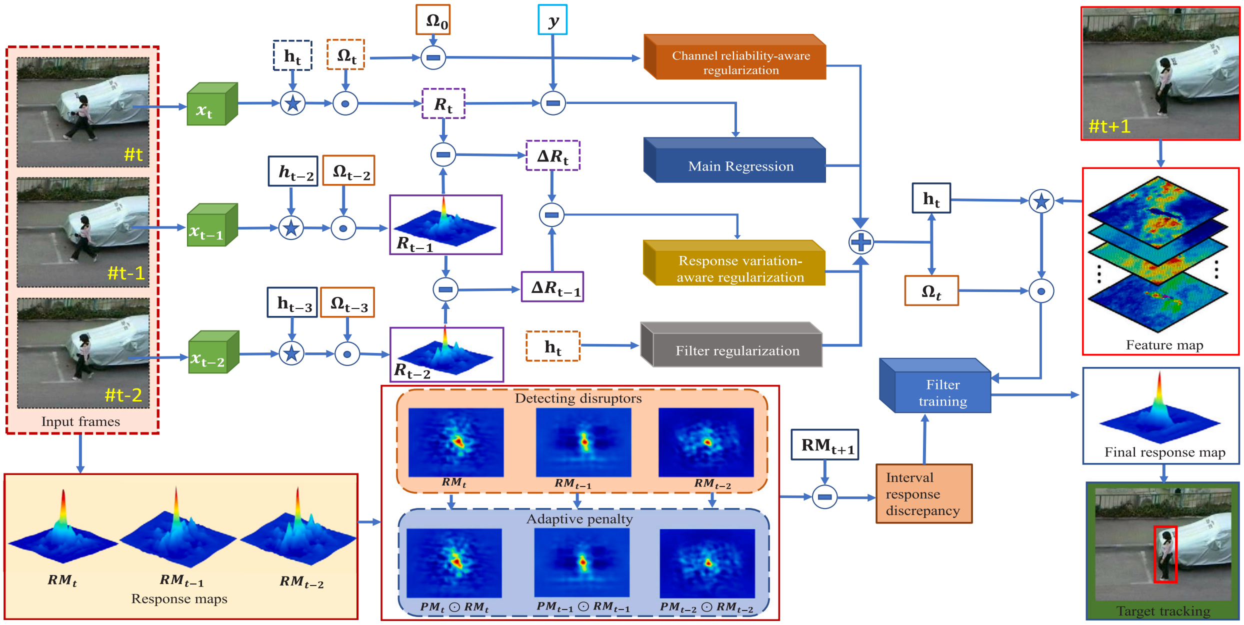


Fig. 2. The tracking procedure of the proposed DSRVMRT method encompasses both training and detection stages. During training, the framework computes the 2-norm difference between the target output and response as the primary regression, while employing three regularizations: filter regularization (current filter norm), response variation-aware regularization (2-order response difference), and channel reliability-aware regularization (tracking weight changes). For detection, an optimized filter and weight coefficient generate response maps using the search patch's feature map. Historical response maps enhance the filter's adaptability to object variations. Finally, the filter is iteratively refined using the ADMM algorithm, ensuring precise object localization in subsequent frames.

3. The proposed method

The proposed DSRVMRT tracking framework consists of three progressive stages, as illustrated in Fig. 2: response variation regularization scheme, channel reliability information, and disruptor-suppression mechanism. In the following sections, a detailed description of each component's structure and role will be provided. The general framework of our tracker is summarized in Algorithm 1.

3.1. Preview of BACF

The BACF tracker integrates the concepts of multi-feature channels and background information, which is depicted as a ridge regression to enhance tracking performance. The objective function is illustrated as follows:

$$E(h_t) = \frac{1}{2}\left\| y - \sum_{d=1}^{D} P x_t^d \odot h_t^d \right\|_2^2 + \frac{\upsilon}{2} \sum_{d=1}^{D} \left\| h_t^d \right\|_2^2, \tag{1}$$

where $x = [x_1, \dots, x_d] \in R^{N \times D}$ denotes the vectorized image. $y \in R^N$ indicates the correlation response. $P \in R^{M \times N}$ represents the binary matrix. $\odot$ and υ denote the correlation operator and regularization parameter.

3.2. Objective function of DSRVMRT

We introduce a disruptor-suppressed response variation-aware multiple regularized CF framework for visual tracking, which aims to mitigate the limitations of the baseline tracker. The presented method can dynamically adjust to evolving tracking conditions, yielding tracking results that are more accurate and dependable, which is achieved by incorporating an updating mechanism for the appearance model. Hence, the objective function is determined as follows:

$$E(h_t) = \frac{1}{2}\left\| y - \sum_{d=1}^{D} P x_t^d \star h_t^d \right\|_2^2 + R_{RVAR} + R_{CRAR} + R_{DSR} \tag{2}$$

where $R_{RVAR}, R_{CRAR}, R_{DSR}$ denotes the response variation-aware term, channel reliability term, and disruptor-suppressed correlation filter term, respectively.

Algorithm 1 DSRVMRT tracking algorithm

```
Input: Video with $t$ frames.
Initial target location $P_1$ and the scale $S_1$ of the first frame $F_1$.
Output: The target position $P_t$ and scale $S_t$ in each frame.
  Construct the target regression $y$.
  Compute the sequential interval responses $RM_{t-f}, \cdots, RM_{t-1}$.
  for frame $f = 1$ to end do
    if $t = 1$ then
      Crop and extract features of training sample $x_1$ from $F_1$ with $P_1$ and $S_1$.
      Format the advent model $X_1^{model} = x_1^1$.
      Compute the correlation filter $h_1$.
      Compute the response map $RM_1 = x_1 \odot h_1$.
    else
      Crop and extract features of training sample $x_t$ from $F_t$ with $P_{t-1}$ and $S_{t-1}$.
      Compute the response variation-aware regularization $S(\Delta R_t^d) - S(\Delta R_{t-1}^d)$ (Section 3.3).
      Generate channel reliability regularization $\Omega_t$ for the frame $t$ (Section 3.4).
      Compute the response map $RM_t = \Omega_t x_t \odot h_{t-1}$.
      Discover the disruptors in the $RM_t$ and compute the penalty mask $PM_t$ (Section 3.5).
      Estimate the target position and scale based on $RM_t$.
      Relocate $RM_{t-f}, \cdots, RM_{t-1}$ to construct their maximum be compatible with $RM_t$.
      Crop and extract features of the training sample $x_t$ from $F_t$ with $P_t$ and $S_t$.
      Update the filter model $h_t$ based on Eq. (24).
    end if
  end for
```

3.3. Response variation-aware regularization

Updating the DCF filter in every frame using the standard approach may result in over-fitting of the filter for the present frame. Furthermore, the trained filter is simply added to the previous filter model, which may cause issues when there are significant changes in the target. When the target changes continuously, the update method works well, but when the changes are substantial, the trained filter and the previous filter may differ significantly. When the two filters are combined, the filter might not be optimal for either the present or preceding target location. Additionally, occlusion and blurring often occur during the target tracking process, which can cause the filter to contain background information and noise if it learns from those training samples. This can lead to tracking drift and failure. To boost the efficacy, response variation can be incorporated into the DCF learning process with the purpose to take advantage of information from the target trusted region. Therefore, we introduce a regularization term for response variation, which can further improve the tracker's performance by adaptively considering the spatial difference of the previous frame. This is achieved by introducing a constraint on the second-order difference of responses. Hence, the resulting term, denoted as R_{RVAR}, can be expressed as follows:

$$R_{RVAR} = \frac{\lambda}{2}\sum_{d=1}^{d}\left\|\Delta R_t^d - \Delta R_{t-1}^d\right\|_2^2, \tag{3}$$

where λ denotes the regularization parameters. Here, $\Delta R_t^d \in R^N$ obtained by calculating the difference between two distinct response maps R_t^d and R_{t-1}^d as follows

$$\Delta R_t^d = S(\Delta R_t^d) - S(\Delta R_{t-1}^d), \tag{4}$$

where the peaks of two response maps R_t^d and R_{t-1}^d are shifted to the center by shift operator $S(.)$.

3.4. Channel reliability-aware regularization

To enhance the accuracy of target localization, the contribution of each feature channel is emphasized by using the channel reliability metric to weigh its response toward the target location. Based on this idea, the proposed tracking method suggests the effectiveness of incorporating the learning of a channel weight distribution Ω_t into the model training process and suppressing misleading ones. As a result, the regularization term R_{CRAR} is created which accounts for channel reliability.

$$R_{CRAR} = \frac{\xi}{2}\|\Omega_t - \Omega_0\|_2^2, \tag{5}$$

where $\Omega_t = diag(\Omega_t^1, \Omega_t^2, \dots, \Omega_t^D)$ represents the diagonal matrix composed of all D channel weights. ξ represents the preset constant. Similarly, the Ω_0 represents the initial weight distribution. To move from frame to frame seamlessly and prioritize the robust channels simultaneously, the current frame's channel weights are indispensable.

3.5. Disruptor-suppressed regularization

The procedure involves developing a response map by comparing the DCF to the characteristics of the area of interest region. The peak with the greatest height in the response pattern is used to calculate the target's estimated localization in the new frame. The resulting output response maps accurately reflect the target's state. As a result, any variations in the object's appearance are initially and immediately reflected in the response maps. Drawing inspiration from this, we leverage these responses when designing a temporal regularization. The responses may contain misleading background noises. To identify and eliminate these disruptive elements, a new disruptor-aware scheme based on response bucketing has been introduced.

Fig. 3 illustrates the process of dividing a response map RM_{t-f} into $\alpha \times \alpha$ non-overlapping rectangles using the bucketing scheme for the $(t-f)^{th}$ frame. Then, the background noises (disruptors) in the individual bucket are detected based on the peak of the response map. Based on this, the penalty factors are identified and the adaptive penalty mask PM_{t-f} is developed. Accordingly, the disruptor detection scheme calculate $\theta(i,j)$ as follows:

$$\theta_{i,j} = \frac{m_{(i,j)}}{m_{max}}, \tag{6}$$

where, $m(i,j)$ is the local maximum value in the $i \times j$ response bucket, and m_{max} is the global maximum value on the entire response map. Then, the penalty factor τ is calculated as

$$\begin{cases} \tau_{(i,j)} = \frac{1}{\rho\theta_{(i,j)}}, & \varkappa \leq \theta_{(i,j)} \leq 1 \\ \tau_{(i,j)} = 1, & otherwise, \end{cases} \tag{7}$$

where ρ represents the predefined constant. A threshold named $varkappa$ is used to determine the possibility that the response bucket consists of disruptions.

Similarly, the penalty mask PM_{t-f} is expressed as follows

$$PM_{t-f} = \begin{bmatrix} \tau_{(1,1)} & \tau_{(1,2)} & \cdots & \tau_{(1,\alpha)} \\ \tau_{(2,1)} & \tau_{(2,2)} & \cdots & \tau_{(2,\alpha)} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{(\alpha,1)} & \tau_{(\alpha,2)} & \cdots & \tau_{(\alpha,\alpha)} \end{bmatrix}. \tag{8}$$

As a result, the response map RM_{t-f} of the $(t-f)^{th}$ frame is then denoised as $PM_{t-f} \odot RM_{t-f}$. The disruptor-aware scheme put forward is capable of effectively and accurately identifying and suppressing disruptors in responses. Fig. 3 indicates that the response map of the DSRVMRT tracker remains stable despite disturbances caused by similar objects, enabling it to precisely locate the target. The prediction results of the DSRVMRT tracker align closely with the ground-truth labels of the tracking target, providing compelling evidence of the disruptor-aware model's effectiveness.

3.5.1. Interval response discrepancy

VOT poses a multitude of challenges, including fast-moving objects or aerial platforms that make it difficult to maintain a consistent tracking performance using only two consecutive frames. On the other hand, relying on hundreds of historical training samples can be impractical due to the large memory footprint. Therefore, we present an innovative approach that attempts to reduce interval response inconsistency has been suggested as an alternative to these problems. The learnt filter can better adapt to object modifications and produce more precise predictions in a subsequent frame by using response maps from the previous F frames throughout the filter training process for the tth frame. By incorporating historical interval responses, which require minimal memory but effectively suppress aberrations, this approach can substantially boost the robustness of object tracking, as illustrated in Fig. 2.

3.6. Overall objective of DSRVMRT

The overall objective of DSRVMRT is described as follows:

$$\begin{aligned} E(h_t, \Omega_t) = & \frac{1}{2}\left\|y - \sum_{d=1}^{D}\Omega_t^d x_t^d \star P^T h_t^d\right\|_2^2 + \frac{\upsilon}{2}\sum_{d=1}^{D}\left\|h_t^d\right\|_2^2 \\ & + \frac{\lambda}{2}\sum_{d=1}^{D}\left\|\Delta R_t^d - \Delta R_{t-1}^d\right\|_2^2 + \frac{\xi}{2}\|\Omega_t - \Omega_0\|_2^2 \\ & + \sum_{f=1}^{f}\Gamma_f\|PM_{t-f} \odot RM_{t-f}[\Psi_{uf,vf}] - RM_t\|_2^2, \end{aligned} \tag{9}$$

where Γ_f is the penalty coefficient.

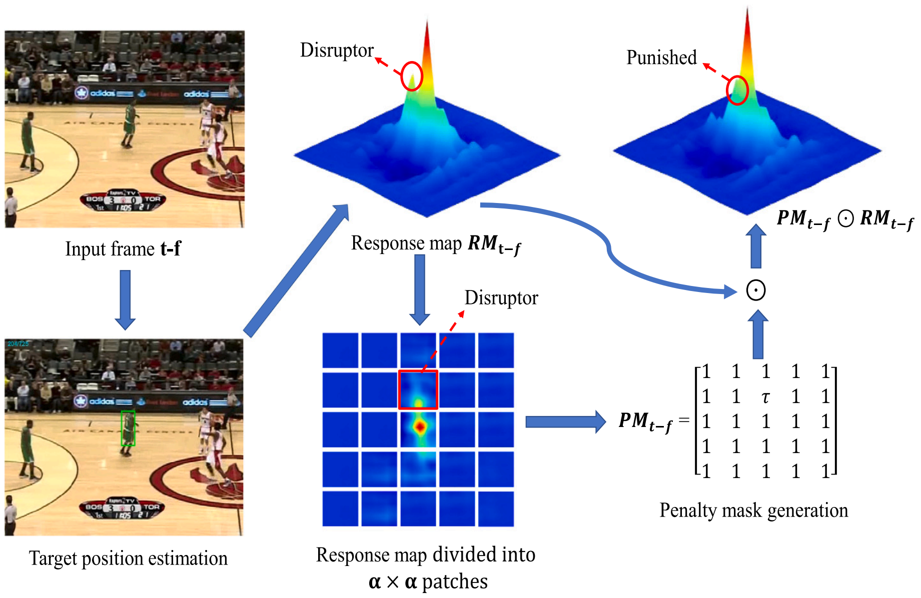

Disruptor
Punished
Input frame t-f
Response map RM_{t-f}
$PM_{t-f} \odot RM_{t-f}$
Disruptor
$\odot$
$PM_{t-f} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & \tau & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$
Penalty mask generation
Target position estimation
Response map divided into
$\alpha \times \alpha$ patches

Fig. 3. The process flow of the disruptor-aware approach that utilizes response bucketing.

3.7. Frequency domain transformation

We introduce an auxiliary variable $\hat{g}_t = \sqrt{T} F P^T h_t^d (\hat{G} = [\hat{g}_t^1, \hat{g}_t^2, \dots \hat{g}_t^D])$ in Eq. (9) as:

$$
\begin{aligned}
E(h_t, \hat{g}_t, \Omega_t) =& \frac{1}{2}\left\| y - \sum_{d=1}^{D} \Omega_t^d x_t^d \odot \hat{g}_t^d \right\|_2^2 + \frac{\upsilon}{2} \sum_{d=1}^{D} \left\| h_t^d \right\|_2^2 \\
&+ \frac{\lambda}{2} \sum_{d=1}^{D} \left\| \Delta R_t^d - \Delta R_{t-1}^d \right\|_2^2 + \frac{\xi}{2} \| \Omega_t - \Omega_0 \|_2^2 \\
&+ \sum_{f=1}^{F} \Gamma_f \| PM_{t-f} \odot RM_{t-f}[\Psi_{uf,vf}] - RM_t \|_2^2 .
\end{aligned} \tag{10}
$$

For the computational purpose, the Eq. (10) is transformed into a frequency domain. Subsequently, the dth channel response of the tth frame could be expressed as $R_t^d = \Omega_t^d x_t^d \odot \hat{g}_t^d$ and substitute in Eq. (10) then in Eq. (11) becomes

$$
\begin{aligned}
E(h_t, \hat{g}_t, \Omega_t) =& \frac{1}{2}\left\| y - \sum_{d=1}^{D} \Omega_t^d x_t^d \odot \hat{g}_t^d \right\|_2^2 + \frac{\upsilon}{2} \sum_{d=1}^{D} \left\| h_t^d \right\|_2^2 \\
&+ \frac{\lambda}{2} \sum_{d=1}^{D} \left\| \Omega_t^d x_t^d - 2\widehat{S(R_{t-1}^d)} + \widehat{S(R_{t-2}^d)} \right\|_2^2 + \frac{\xi}{2} \| \Omega_t - \Omega_0 \|_2^2 \\
&+ \sum_{f=1}^{F} \Gamma_f \| PM_{t-f} \odot RM_{t-f}[\Psi_{uf,vf}] - \sum_{d=1}^{D} x_t^d \odot g_t^d \|_2^2 .
\end{aligned} \tag{11}
$$

3.8. Optimization through ADMM

ADMM, which stands for alternative direction method of multipliers, is utilized to accelerate computations in a manner similar to the BACF tracker. Similarly, the Augmented Lagrangian function is exploited to solve the Eq. (11) as below:

$$
\begin{aligned}
L(h_t, \hat{g}_t, \Omega_t, \hat{M}_t) =& \frac{1}{2}\left\| y - \sum_{d=1}^{D} \Omega_t^d \hat{x}_t^d \odot \hat{g}_t^d \right\|_2^2 + \frac{\upsilon}{2} \sum_{d=1}^{D} \left\| h_t^d \right\|_2^2 \\
&+ \frac{\lambda}{2} \sum_{d=1}^{D} \left\| \Omega_t^d x_t^d \odot g_t^d - 2\widehat{S(R_{t-1}^d)} + \widehat{S(R_{t-2}^d)} \right\|_2^2 \\
&+ \frac{\xi}{2} \| \Omega_t - \Omega_0 \|_2^2 \\
&+ \sum_{f=1}^{F} \Gamma_f \| PM_{t-f} \odot RM_{t-f}[\Psi_{uf,vf}] - \sum_{d=1}^{D} \hat{x}_t^d \odot \hat{g}_t^d \|_2^2 \\
&+ \sum_{d-1}^{D} ({\hat{g}_t}^d - \sqrt{T} F P^T h_t^d) {\hat{m}_t}^d \\
&+ \frac{\mu}{2} \sum_{d=1}^{D} \left\| \hat{g}_t^d - \sqrt{T} F P^T h_t^d \right\|_2^2 ,
\end{aligned} \tag{12}
$$

where m denotes the Lagrange multiplier. μ is the penalty factor. For simplification we define $\hat{g}_t = [\hat{g}_t^1, \hat{g}_t^2, \hat{g}_t^3, \dots, \hat{g}_t^d]$ and $\hat{M}_t = [\hat{m}_t^1, \hat{m}_t^2, \hat{m}_t^3, \dots, \hat{m}_t^d]$. Let we consider $\hat{\varsigma}_t^d = \frac{1}{\mu} \hat{m}_t$ in Eq. (12), the Eq. (13) becomes,

$$
\begin{aligned}
L(h_t, \hat{g}_t, \Omega_t, \hat{\varsigma}_t) =& \frac{1}{2}\left\| y - \sum_{d=1}^{D} \Omega_t^d \hat{x}_t^d \odot \hat{g}_t^d \right\|_2^2 + \frac{\upsilon}{2} \sum_{d=1}^{D} \left\| h_t^d \right\|_2^2 \\
&+ \frac{\lambda}{2} \sum_{d=1}^{D} \left\| \Omega_t^d x_t^d \odot g_t^d - 2\widehat{S(R_{t-1}^d)} + \widehat{S(R_{t-2}^d)} \right\|_2^2 \\
&+ \frac{\xi}{2} \| \Omega_t - \Omega_0 \|_2^2 \\
&+ \sum_{f=1}^{F} \Gamma_f \| PM_{t-f} \odot RM_{t-f}[\Psi_{uf,vf}] - \sum_{d=1}^{D} \hat{x}_t^d \odot \hat{g}_t^d \|_2^2 \\
&+ \frac{\mu}{2} \sum_{d=1}^{D} \left\| \hat{g}_t^d - \sqrt{T} F P^T h_t^d + \hat{\varsigma}_t^d \right\|_2^2 .
\end{aligned} \tag{13}
$$

3.8.1. Subproblem $\hat{g}_t^*$

Given $h_t, \hat{g}_t, \Omega_t, \theta_t, \hat{M}$ can be acquired by solving:

$$
\begin{aligned}
\hat{g}_t^* =& \underset{G_t}{\operatorname{argmin}} \left\{ \frac{1}{2}\left\| \hat{y} - \sum_{d=1}^{D} \Omega_t^d \hat{x}_t^d \odot \hat{g}_t^d \right\|_2^2 \right. \\
&+ \frac{\lambda}{2} \sum_{d=1}^{D} \left\| \Omega_t^d x_t^d \odot g_t^d - 2\widehat{S(R_{t-1}^d)} + \widehat{S(R_{t-2}^d)} \right\|_2^2 \\
&+ \sum_{f=1}^{F} \Gamma_f \| PM_{t-f} \odot RM_{t-f}[\Psi_{uf,vf}] - \sum_{d=1}^{D} \hat{x}_t^d \odot \hat{g}_t^d \|_2^2 \\
&+ \frac{\mu}{2} \sum_{d=1}^{D} \left\| \hat{g}_t^d - \sqrt{T} F P^T h_t^d + \hat{\varsigma}_t^d \right\|_2^2 .
\end{aligned} \tag{14}
$$

Moreover, we employ the ADMM technique to obtain better results. Here, $\hat{y}$ dependent on $\hat{x}_t(n) = [\hat{x}_t^1(n), \hat{x}_t^2(n), \dots, \hat{x}_t^C(n)]^T$. Introducing, $\hat{g}_t(n) = [conj(\hat{g}_t^1(n)), conj(\hat{g}_t^2(n)), \dots, conj(\hat{g}_t^D(n))]^T$ where $conj(.)$ denotes the conjugate function. The Eq. (14) transferred as sub-expressions:

$$
\hat{g}_{t+1}^*(n) = \frac{1}{2}\left\| \hat{y}(n) - \sum_{d=1}^{D} \Omega_t^d \hat{x}_t^d(n) \odot \hat{g}_t^d(n) \right\|_2^2
$$

$$\begin{aligned}
&+ \frac{\lambda}{2}\sum_{d=1}^{D}\left\|\Omega_t^d x_t^d \odot g_t^d - 2\widehat{S(R_{t-1}^d)} + \widehat{S(R_{t-2}^d)}\right\|_2^2 \\
&+ \sum_{f=1}^{F}\Gamma_f \| PM_{t-f} \odot RM_{t-f}[\Psi_{uf,vf}] - \sum_{d=1}^{D}\hat{x}_t^d \odot \hat{g}_t^d \|_2^2 \\
&+ \frac{\mu}{2}\sum_{d=1}^{D}\left\|\hat{g}_t^d - \sqrt{T}FP^T h_t^d + \hat{\varsigma}_t^d\right\|_2^2,
\end{aligned} \tag{15}$$

where $\hat{h}_t(n) = [\hat{h}_t^1(n), \hat{h}_t^2(n), \ldots, \hat{h}_t^C(n)]^T$ and $\hat{h}_t^c$ denotes the DFT of h_t^c, i.e., $\hat{h}_t^c = \sqrt{T}FB^T h_t^c$. The solution of $\hat{g}_t^*$ is obtained as follows:

$$\begin{aligned}
\hat{g}_{t+1}^*(n) =& b\Bigg[\ \Omega_t^T \hat{x}_t(n)\hat{x}_t(n)^T \Omega_t + \lambda \Omega_t^T \hat{x}_t(n)\hat{x}_t(n)^T \Omega_t \\
&+ \sum_{f=1}^{F}\Gamma_f \Omega_t^T \hat{x}_t(n)\hat{x}_t(n)^T \Omega_t + \mu I_D\ \Bigg]^{-1} \\
&\Bigg[\ \Omega_t^T \hat{x}_t(n)\hat{y}_t(n) + 2\lambda \Omega_t^T \hat{x} S\widehat{(R_{t-1})}(n) - \lambda \Omega_t^T \hat{x} S\widehat{(R_{t-2})}(n) \\
&+ \sum_{f=1}^{F}\Gamma_f PM_{t-f} \odot RM_{t-f}[\Psi_{uf,vf}] + \mu\hat{h} - \mu\hat{r}\ \Bigg].
\end{aligned} \tag{16}$$

To avoid the inverse problem, we employ the Sherman–Morrison formula in the Eq. (16), i.e., $(S+uv^T)^{-1} = S^{-1}-S^{-1}u(I_D+v^T S^{-1}u)^{-1}v^T S^{-1}$. In this case, $S = \frac{\mu}{1+\lambda+\sum_{f=1}^{F}\Gamma_f}$, and $u = v = \Omega_t^T \hat{x}_t(n)$. As a result, the Eq. (16) is equivalent to the following expression:

$$\hat{g}_{t+1}^* = \frac{1}{\mu}\left[I_D - \frac{\Omega_t^T \hat{x}_t(n)\hat{x}_t(n)^T \Omega_t}{\mu b + \hat{x}_t(n)^T \Omega_t^T \Omega_t \hat{x}_t(n)}\right]\rho, \tag{17}$$

where, $\rho = \Big[\Omega_t^T \hat{x}_t(n)\hat{y}_t(n) + 2\lambda \Omega_t^T \hat{x} S\widehat{(R_{t-1})}(n) - \lambda \Omega_t^T \hat{x} S\widehat{(R_{t-2})}(n) + \sum_{f=1}^{F}\Gamma_f PM_{t-f} \odot RM_{t-f}[\Psi_{uf,vf}] + \mu\hat{h} - \mu\hat{r}\Big]$.

3.8.2. Subproblem h_t^*

The subproblem h_t^* is determined as follows:

$$h_{t+1}^* = \underset{h_t}{\operatorname{argmin}}\left\{\frac{v}{2}\|h_t^d\|_2^2 + \frac{\mu}{2}\|\hat{g}_t^d - \sqrt{T}FP^T h_t^d + \hat{\varsigma}_t^d\|_2^2\right\}, \tag{18}$$

the Eq. (18) is expressed as follows:

$$h_{t+1}^* = \frac{\mu T \odot (\hat{\varsigma}_t^d + \hat{g}_t^d)}{v + \mu T}. \tag{19}$$

3.8.3. Subproblem Ω_t^*

The subproblem Ω_t^* is formulated as follows:

$$\begin{aligned}
\Omega_{t+1}^* =& \frac{1}{2}\left\|y - \sum_{d=1}^{D}\Omega_t^d \hat{x}_t^d \odot \hat{g}_t^d\right\|_2^2 \\
&+ \frac{\lambda}{2}\left\|\Omega_t^d x_t^d \odot \hat{g}_t^d - 2\widehat{S(R_{t-1}^d)} + \widehat{S(R_{t-2}^d)}\right\|_2^2 + \frac{\xi}{2}\|\Omega_t - \Omega_0\|_2^2 \\
&+ \sum_{f=1}^{F}\Gamma_f \| PM_{t-f} \odot RM_{t-f}[\Psi_{uf,vf}] - \sum_{d=1}^{D}\hat{x}_t^d \Omega_t \odot \hat{g}_t^d \|_2^2,
\end{aligned} \tag{20}$$

the simplified form of Eq. (20) is,

$$\Omega_{t+1}^* = \frac{(\hat{x}_t^d \odot \hat{g}_t^d)(\hat{y} + 2\lambda \widehat{S(R_{t-1}^d)} - \lambda \widehat{S(R_{t-2}^d)} + \sum_{d=1}^{D}\Gamma_f(\vartheta)) + \xi N \Omega_0^d}{(1+\lambda)(\sum_{f=1}^{F}\Gamma_f(\hat{x}_t^d \odot \hat{g}_t^d)^T(\hat{x}_t^d \odot \hat{g}_t^d)) + \xi N}, \tag{21}$$

where $\vartheta = PM_{t-f} \odot RM_{t-f}[\Psi_{uf,vf}]$.

3.8.4. Lagrangian update

The Lagrangians can be updated as follows:

$$\hat{\varsigma}_t^{i+1} = \hat{\varsigma}_t^i + \mu^i\left(\hat{g}_t^{i+1} - \hat{h}_t^{i+1}\right), \tag{22}$$

$$\mu^{i+1} = min(\mu_{max}, \beta\mu^i), \tag{23}$$

where superscript i and $i+1$ denote the iteration and β indicates the scale factor.

3.9. Model update

The model $\hat{x}^{model}$ is updated as follows:

$$\hat{x}_{t+1}^{model} = (1-\eta)\hat{x}_t^{model} + \eta\hat{x}_{t+1}, \tag{24}$$

where η indicates the learning rate of the model.

4. Experimental results

To evaluate the tracker performance, we provide a comparison of DSRVMRT with state-of-the-art methods across seven challenging benchmark datasets: OTB2013 [29], OTB2015 [29], TempleColor128 [30], GOT-10k [31], UAV123 [32], UAVDT [33], LaSOT [34] and DTB70 [35]. The attributes include illumination variation (IV), scale variation (SV), occlusion (OCC), deformation (DEF), motion blur (MB), fast motion (FM), in-plane rotation (IPR), out-of-plane rotation (OPR), out-of-view (OV), background clutter (BC), low resolution (LR), aspect ratio change (ARC), fast camera motion (FCM), camera motion (CM), full occlusion (FOC), partial occlusion (POC), similar object (SIB), small objects (SOB), viewpoint change (VC), large occlusion (LOC), occlusion with background clutter (O-B), object motion (OM), long-term tracking (LT), rotation (ROT). Table 1 reports the applications, scenarios, characteristics, and the overlap of single object tracking datasets. By different evaluation protocols, existing visual tracking benchmarks assess the accuracy and robustness of trackers in realistic scenarios.

To ensure an unbiased comparison, the tracking results of various trackers are evaluated based on their precision and success rates, as noted in [29]. The error between the centers of the tracked and ground truth bounding boxes is quantified using a precision plot. Trackers are ranked based on their performance at a common threshold of 20 pixels (P(20px)). On the other hand, a success plot measures the intersection over union (IoU) between the tracked and ground truth bounding boxes. This plot shows the percentage of correctly predicted bounding boxes on the y-axis for varying overlaps on the x-axis. The compared trackers are ranked based on the area under the curve (AUC) of the success plot.

4.1. Implementation details

In the domain of VOT, it has been established that a reliable and distinctive feature representation is crucial. In our proposed DSRVMRT approach, we incorporate manually-crafted features such as HOG, CN and Saliency, as well as deep features extraction from the pre-trained ResNet-50 model. The parameters in our implementation are presented in Table 2.

4.2. Ablation analysis

The effect of various combinations of main components in DSRVMRT, which includes base tracker (BT), response variation-aware regularization ($RDAR$), channel reliability-aware regularization ($CRAR$), disruptor-suppressed regularization (DSR), and deep features (DF), is presented in Table 3. The contributions of each component were evaluated by individually removing them from the entire framework and measuring the results. The detailed results showcase the performance achieved by different combinations of components. From the Table 3, we notice that by including $RVAR$ to BT, $BT + RVAR$ achieves the DP/AUC of $(82.6\%, 62.3\%)$, the performs better than BT $(80.4\%, 61.1\%)$. Additionally, we observe that the $CRAR$ on the $BT + RVAR + CRAR$ gives significant amendment of 1.5% in precision and 0.8% in success, respectively. Next, we examine the significance of disruptor suppression (DSR) for the proposed tracker. By integrating the DSR into *BT+RVAR+CRAR*, the *BT+RVAR+CRAR+DSR* increases DP by 2.1% and AUC by 2.3% respectively. To further upgrade the tracking performance, we integrate the deep features into the proposed tracker *DSRVMRT* resulting in best performance improvement and achieving $(91.0\%, 68.1\%)$ in DP/AUC scores.

Table 1
Characteristics of benchmark datasets. The abbreviations are denoted as Nov : Number of Videos, NoF : Number of Frames, NoA : Number of Attributes.

Year	Application	Dataset	NoV	NoF	NoA	Attributes
2013	Generic	OTB 2013	51	29K	11	IV, SV, OCC, DEF, MB, FM, IPR, OPR, OV, BC, LR
2015	Generic	OTB 2015	100	59K	11	IV, SV, OCC, DEF, MB, FM, IPR, OPR, OV, BC, LR
2015	Generic	TempleColor128	129	55K	11	IV, SV, OCC, DEF, MB, FM, IPR, OPR, OV, BC, LR
2016	UAV	UAV123	123	113K	12	IV, SV, FM ,OV, BC, LR, ARC, CM, FCM, FOC, POC, SIB, VC
2017	UAV	DTB	70	15K	11	SV, OCC, DEF, MB, IPR, OPR, OV, BC, ARC, FCM, SIB
2018	Generic	GOT10K	10 000	1.5M	6	IV, SV, OCC, FM, ARC, LOC
2018	UAV	UAVDT	50	80K	9	BC, CM, OM, SOB, IV, OB, SV, LOC, LT
2019	Generic	LaSOT	1400	3.5M	14	IV, SV, DEF, MB, FM, OV, BC, LR, ARC, CM, FOC, POC, VC, ROT

Table 2
Parameters of the proposed tracker.

Parameter	Value	Parameter	Value
Number of ADMM iterations	3	Ω_0^d	1
Regularization Parameter ξ	26	Regularization Parameter υ	0.01
Regularization Parameter λ	0.004	Frame weight base Γ_1	0.443
Bucketing number α	12 × 12	Disruptor Threshold $\varkappa$	0.47
Penalty factor μ	0.844	Learning rate η	0.0193
Step Length β	10	Interval Length F	3
Disruptor factor ρ	3	μ_{max}	10 000

Table 3
Ablation analysis.

Method	BT	RVAR	CRAR	DSR	DF	DP %	AUC %
1	✓					80.4	61.1
2	✓	✓				82.6	62.3
3	✓	✓	✓			84.1	63.1
4	✓	✓	✓	✓		86.2	65.4
5	✓	✓	✓	✓	✓	91.0	68.1

4.3. Parameter analysis

This section investigates the effect of core parameters on the overall performance of DSRVMRT. Analysis is conducted on key parameters using the UAVDT benchmark. Different numerical values are assigned to the bucketing number $\alpha \times \alpha$ and the interval length F for further verification.

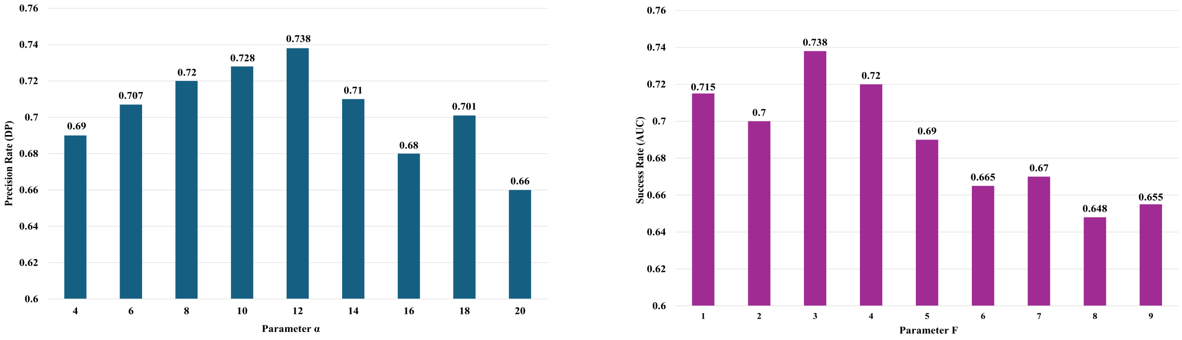
1. **Bucketing Number** $\alpha \times \alpha$ **:**
The bucketing number $\alpha \times \alpha$ is set to range from 4 to 20 for the trials. In most instances, the peak of the response map is located at the center. To avoid damaging the object region, α is adjusted to specific values. Results for precision is shown in Fig. 4(a). As increases, the success rate reaches its peak of 0.738 when $\alpha = 12$. Beyond this point, the success rate fluctuates slightly until $\alpha = 18$. Similarly, precision follows a comparable trend, achieving its best score of 0.738 at $\alpha = 12$. These results highlight the effectiveness of the bucketing-based disruptor-aware scheme in detecting and suppressing disruptors when α is set within an optimal range. Consequently, the discriminative power of the system is enhanced, leading to overall performance improvements. Based on these findings, α is set to 12 in this work.
2. **Interval Length F:** F is set to range from 1 to 9 for the trial, with a step size of 1. Results for precision and success rate are presented in Fig. 4(b). As F increases, both precision and success rate reach their highest values at $F = 3$. Beyond this point, both metrics gradually decline and stabilize. From this, it can be concluded that setting F within an appropriate range allows the interval-based response inconsistency to enhance tracking performance. Consequently, F is set to 3 in this work.
3. **Parameter Analysis:** λ, ξ and η
The presented DSRVMRT approach has three key parameters: λ, ξ and η. First, we investigate the response variation-aware regularization parameter λ in Eq. (21). Fig. 5 displays the outcomes obtained with different values of λ and ξ on the OTB2015 dataset. The λ changes from 0.001 to 0.01 and obtains the highest value when $\lambda = 0.004$. When the value rises above 0.004, the precision score declines abruptly. The DSRVMRT approach acquires the maximum precision at $\lambda = 0.004$. Then, we examine the channel reliability-aware regularize parameter ξ in Eq. (21). We fixed $\lambda = 0.004$ and ξ changes from 5 to 50 the performance tends to degrade when ξ is excessively large. Specifically, the DP score achieved the maximum value at 91.0% with a value of $\lambda = 0.004$ and ξ of 25. Additionally, the parameter η is employed to update the filters that are used for VOT. Among other things, we investigate the impact of η in Eq. (24) in the proposed algorithm. For this, we conducted a series of experiments with different values. The DSRVMRT approach obtains the preeminent accuracy score when η is set to 0.019. Furthermore, we observed a moderate decline in tracking performance when η is set to other values.

4.4. Quantitative results

To evaluate the tracker efficiency, we compare the proposed DSRVMRT method with 50 different trackers such as KCF [4], SRDCF [6], LADCF [36], DSAR-CF [37], HDT [38], MEEM [39], MEVT [27], DSST [12], SiamRPN [40], MCCT [41], STRCF [42], SiamFC [43], AutoTrack [44], DaSiamRPN [45], ATOM [46], MDNet [47], STAR [48], Staple [49], SAMF [50], ARCF [8], MEEM [39], SiamRPN [40], SiamFC [43], DaSiamRPN [45], ATOM [46], MDNet [47], BACF [7], A3DCF [51], HCFM [52], BSTCF [53], STAR [54], ARCF [8], AutoTrack [44], SASR [55], SAMF [50], AMCF [56], EFSCF [57], SiamRKPN [28], DUSTNET [58], CAERDCF [59], HSIC-SRCF [60], FWRDCF [61], SiamMask [62], TADT-PAF [63], HiFT [64], TCTrack++ [65], DSiam [66], CFNet [67], CSRDCF [19], DSST [12].

4.4.1. Experiments on OTB2013 dataset

We conducted a comparison between the DSRVMRT tracker and 19 other trackers, including ATOM, DaSiamRPN, CF2, and various other approaches, as shown in Figs. 6(a) and 6(b). The outcomes demonstrate that the DSRVMRT tracker obtains best efficiency in DP and AUC scores of 91.8% and 69.7%, respectively. The proposed method outperforms the trackers implemented with multiple feature fusion such as CF2 by (2.7%/9.2%) and HDT by (2.9%/9.4%) in DP/AUC metrics. In particular, our method demonstrates a better enhancement in DP/AUC, outperforming the STRCF tracker with deep features by (2.8%/1.8%). Furthermore, the tracking outcomes of our presented approach on the OTB2013 dataset are depicted in Table 4. In particular, the DSRVMRT outperforms DaSiamRPN by (3.2%, 4.2%), SiamRPN by (3.4%, 3.9%), and ATOM by (4.3%, 3.8%). When compared with ARCF, BACF, and SRDCF methods, our approach enhances precision by (9%, 7.5%, 8%) and overlap by (7.1%, 5.1%, 7.1%). Overall, our proposed method demonstrates superior performance compared to the other tracked methods. We also assess the tracking speed of our method in comparison to popular deep-learning based tracking methods. Table 4 presents the FPS (frames per second) for each tracking method. Notably, our DSRVMRT method achieves 45 FPS, outperforming MCCT, which reaches only 7.8 FPS (see Table 5).



(a) Precision rate of the DSRVMRT under different values of bucketing number $\alpha \times \alpha$ on the UAVDT benchmark. At $\alpha = 12$, the precision plots reach the highest points.

(b) Precision of the DSRVMRT under different values of interval length F on the UAVDT benchmark. At $F = 3$, the precision plots reach the highest points.

Fig. 4. Parameter analysis of bucketing number and interval length.

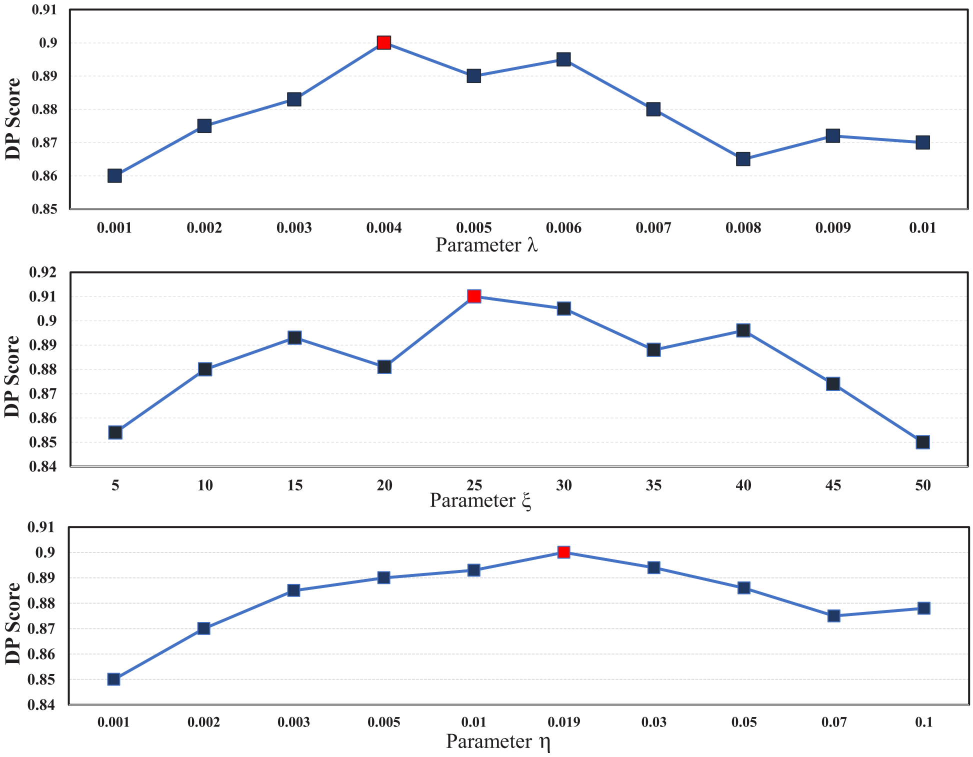


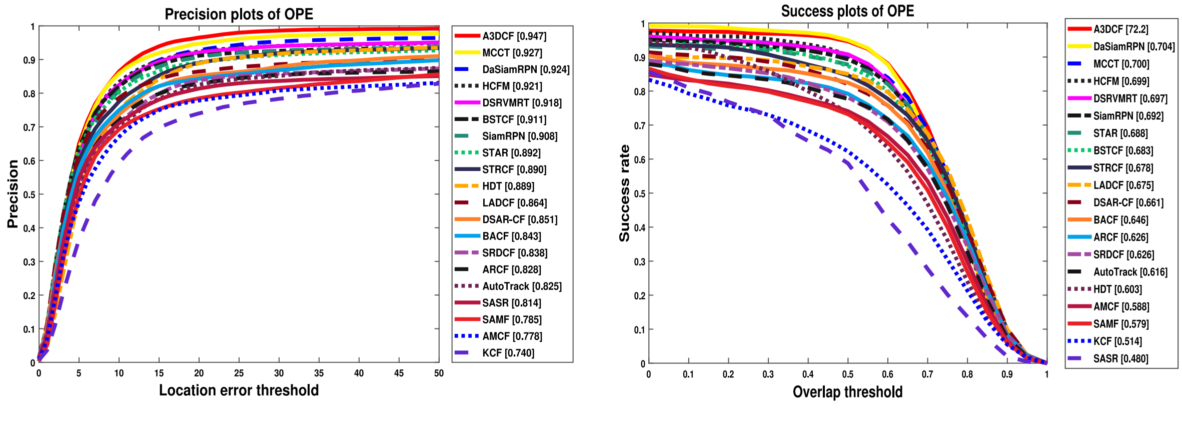
Fig. 5. Analysis of parameters λ, ξ, and η on OTB2015 dataset.

Table 4
Comparative results of conventional trackers with the proposed tracker on the OTB-2013 dataset with 50 sequences. The results of the top 15 trackers are presented as follows:

Methods	Metrics	DSRVMRT	A3DCF	MCCT	DaSiamRPN	HCFM	BSTCF	SiamRPN	STAR	STRCF	HDT	LADCF	DSAR-CF	BACF	SRDCF	ARCF
OTB-2013	DP	91.8	94.7	92.7	92.4	92.1	91.1	90.8	89.2	89.0	88.9	86.4	85.1	84.3	83.8	82.8
	AUC	69.7	72.2	70.0	70.4	69.8	68.3	69.2	68.8	67.8	60.3	67.5	66.1	64.6	62.6	62.6
FPS		**45**	–	**7.8**	**160.0**	**2.7**	**19.0**	**200.0**	**5.5**	**5.8**	**10.0**	**1.5**	**16.0**	**25.4**	–	**26.2**

Table 5
Comparative results of conventional trackers with the proposed tracker on the OTB-2015 dataset with 100 sequences. The results of the top 15 trackers are presented as follows:

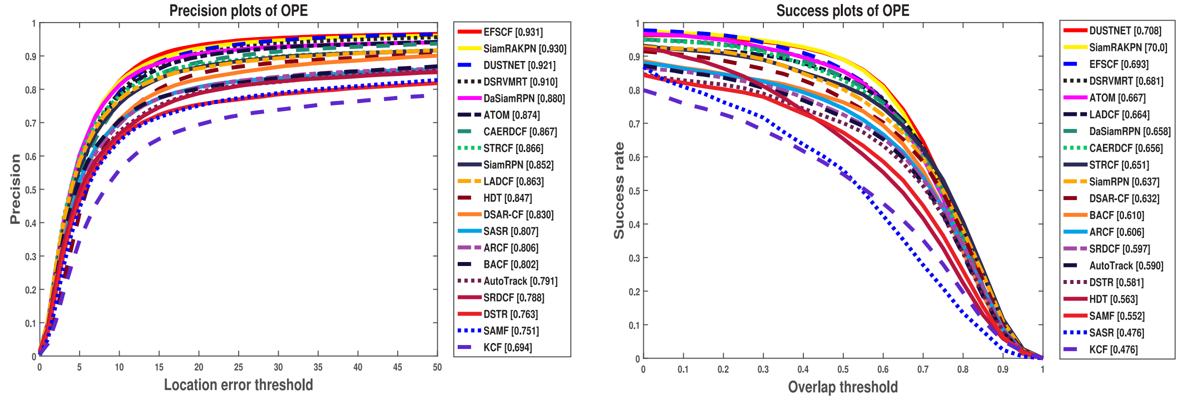
Methods	Metrics	DSRVMRT	EFSCF	SiamRAKPN	DUSTNET	DaSiamRPN	ATOM	CAERDCF	STRCF	SiamRPN	LADCF	HDT	DSAR-CF	SASR	ARCF	BACF
OTB-2015	DP	91.0	93.1	93.0	91.0	88.0	87.4	86.7	86.6	85.2	86.3	84.7	83.0	80.7	80.6	80.2
	AUC	68.1	69.3	70.0	70.8	65.8	66.7	65.6	65.1	63.7	66.4	56.3	63.2	47.6	60.6	61.0
FPS		**33**	–	**7.8**	**49**	**2.7**	**19.0**	–	**5.5**	**5.8**	**10.0**	**1.5**	**16.0**	**25.4**	–	**26.2**



(a)

(b)

Fig. 6. Precision and success plots on OTB2013.



(a)

(b)

Fig. 7. Precision and success plots on OTB2015.

4.4.2. *Experiments on OTB2015 dataset*

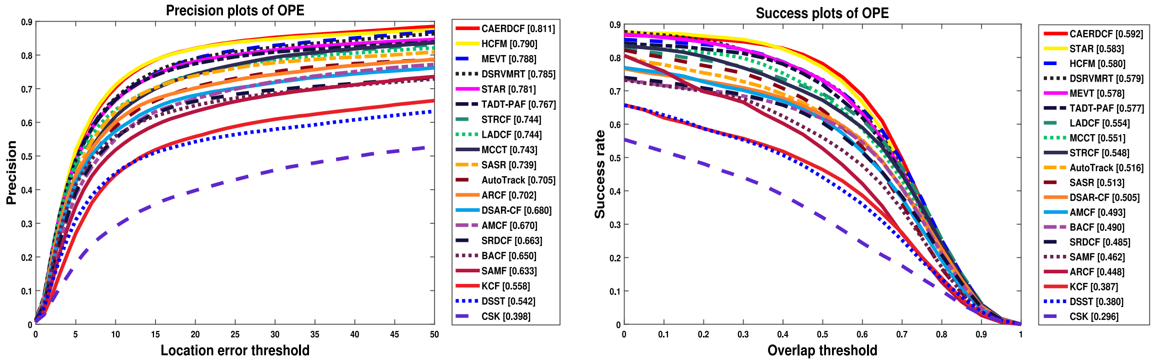
Figs. 7(a) and 7(b) showcase the assessment of the proposed DSRVMRT tracker on the OTB2015 dataset. Particularly, the proposed tracker reaches (91%/68.1%) in DP/AUC on OTB2015 and achieves the optimum performance compared to all trackers. In contrast to other top-performing methods, such as DaSiamRPN, ATOM, and SiamRPN, our method achieves significant developments in both DP scores (3%, 3.6%, 5.8%) and AUC scores (2.3%, 1.4%, 4.4%). Furthermore, when compared to other DCF-based trackers that aim to address the model drift problem by reducing boundary effects, such as ARCF, SRDCF, and BACF, our proposed tracker outperforms them with increases of (10.3%, 12.1%, 10.6%) in DP scores, and (7.4%, 8.3%, 7%) in AUC scores as shown in Table 7. Technically, the DSRVMRT algorithm outperforms STRCF with a gain of (3.0%/2.1%) in both metrics. Specifically, the presented tracker demonstrates superior performance compared to recent tracking methods including CAERDCF, and DUSTNET by (4.3%, 5.8%) and (2.5%, 4.4%) in precision and overlap. Similarly, our proposed DSRVMRT method achieves a notable performance enhancement compared to the AutoTrack and KCF trackers in terms of DP and AUC scores of (79.2%, 69.6%) and (59.0%, 47.7%). To summarize, the mechanism we have proposed is capable of dynamically addressing different obstacles encountered in the target during tracking. Regarding tracking speed, while the Siamese trackers SiamRAKPN and DaSiamRPN demonstrate clear advantages, achieving speeds of 7.8 and 2.7 frames per second respectively, they excel in both precision and accuracy metrics. Although ATOM and LADCF operate at similar speeds to our proposed tracker, their performance in other metrics is more limited. Our DSRVMRT method, which utilizes multiple features, runs at a speed of 33 FPS, making it competitive with other correlation filter-based tracking algorithms that use deep features on the OTB2015 dataset (see Table 6).

4.4.3. *Experiments on TempleColor128 dataset*

Figs. 8(a) and 8(b) illustrate the performance of different trackers on the TempleColor128 dataset. As depicted in Fig. 8(a), the DSRVMRT method obtains the best precision and success scores of 78.6% and 57.9%, respectively. According to Table 7, the CAERDCF and TAD-PAF trackers achieve 81.1%/59.2% and 76.7%/57.7% in DP and AUC scores. More specifically, our DSRVMRT achieves better results than the ensemble tracking methods including MCCT-H, MEEM, and HDT. Conversely, our method exhibits a substantial increment over the BACF tracker, with enhancements of (13.6%) and (8.9%) in precision and overlap, respectively. Encouragingly, our presented approach maintains superior performance compared to many existing methods with boundary effect suppression, including ARCF (70.2%/44.8%), SRDCF (66.3%/48.5%), and STRCF (74.4%/46.9%). The proposed method achieves comparable performance over the recent CAERDCF and TADT-PAF trackers with the DP amd AUC scores of 81.1%/59.2 and 76.7%/57.7% respectively. Specifically, these outcomes demonstrate that our tracking method outperformed well when compared to modern trackers.

4.4.4. *Experiments on UAV123 dataset*

Figs. 9(a) and 9(b) illustrate comprehensive evaluation results on the UAV123 dataset. Our outcomes indicate that the presented approach excels state-of-the-art trackers in both metrics, achieving scores

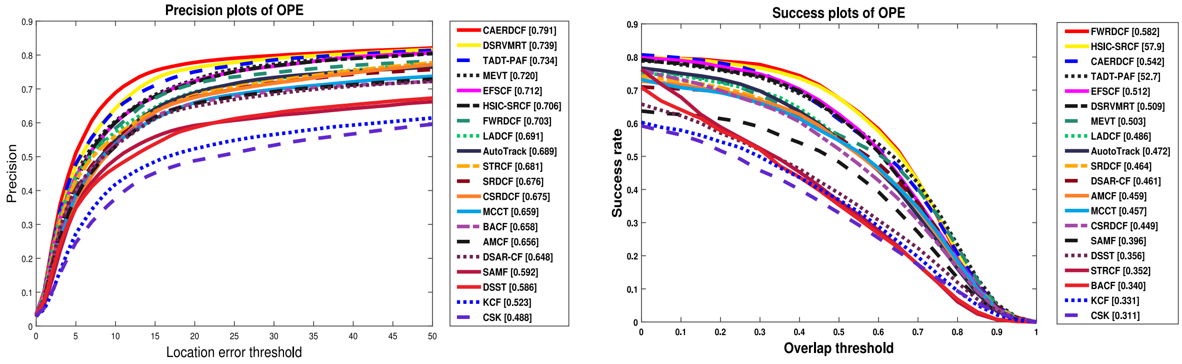


(a) (b)

Fig. 8. Precision and success plots on TempleColor128.

Table 6
Comparative results of conventional trackers with the proposed tracker on the TempleColor128 dataset.

Methods	Metrics	DSRVMRT	CAERDCF	HCFM	MEVT	STAR	TADT-PAF	STRCF	LADCF	MCCT	SASR	AutoTrack	ARCF	DSAR-CF	AMCF	SRDCF
TempleColor128	DP	78.5	81.1	79.0	78.8	78.1	76.7	74.4	74.4	74.3	73.9	70.5	70.2	68.0	67.0	66.3
	AUC	57.9	59.2	58.0	57.8	58.3	57.7	54.8	55.4	55.1	51.3	51.6	44.8	50.5	49.3	48.5



(a) (b)

Fig. 9. Precision and success plots on UAV123.

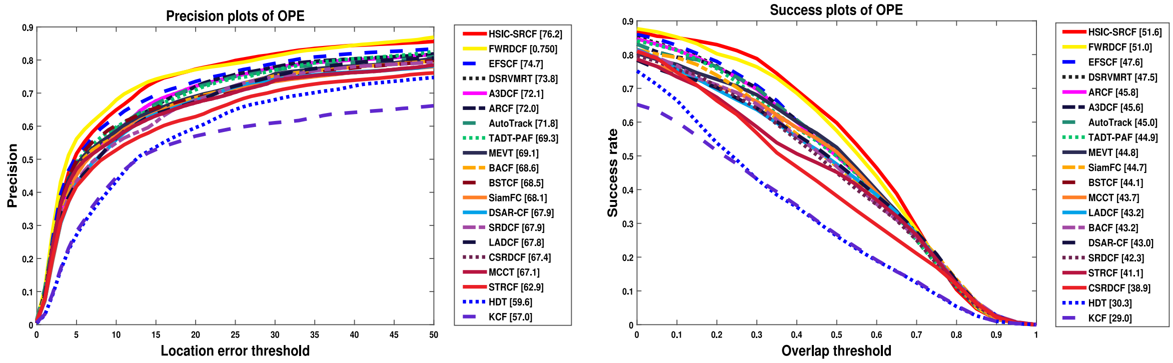
Table 7
Comparative results of conventional trackers with the proposed tracker on the UAV123 dataset.

Methods	Metrics	DSRVMRT	CAERDCF	TADT-PAF	MEVT	EFSCF	HSIC-SRCF	FWRDCF	LADCF	AutoTrack	STRCF	SRDCF	CSRDCF	MCCT	BACF	AMCF
UAV123	DP	73.9	79.1	73.4	72.0	71.2	70.6	70.3	69.1	68.9	68.1	67.6	67.5	65.9	65.8	65.6
	AUC	57.9	54.2	52.7	50.3	51.2	57.9	58.2	48.6	47.2	35.2	46.4	44.9	45.7	34.0	45.5

of 73.9% for precision and 50.9% for success. Compared to the EFSCF, HSIC-SRCF and FWRDCF trackers, the proposed DSRVMRT demonstrated an enhancement of 3.1%, 4.6% and 4.9% in precision rate. It is worth noting that CAERDCF is an excellent method which achieves best performances in both precision and success scores 79.1% and 54.2%. Furthermore, in UAV123, DSRVMRT outperformed the baseline BACF, exhibiting an increase of 7.9%/5% in DP/AUC scores. Specifically, the proposed method obtains a gain of (5.8%/2.8%), and (6%/4.5%), with the STRCF and SRDCF approaches in DP and AUC scores. Furthermore, DSRVMRT demonstrates superior performance compared to the CF trackers, AMCF and KCF, with relative increases of (8.3%) and (5%) in AUC scores, and (21.6%) and (17.8%) in precision scores, respectively. Notably, our tracker surpasses the performance of MCCT-H and MEEM, which utilize multi-layer features and tracker ensembles in all cases. The DSRVMRT tracker, which leverages multiple historical frames, has demonstrated its competence in effectively handling significant appearance changes.

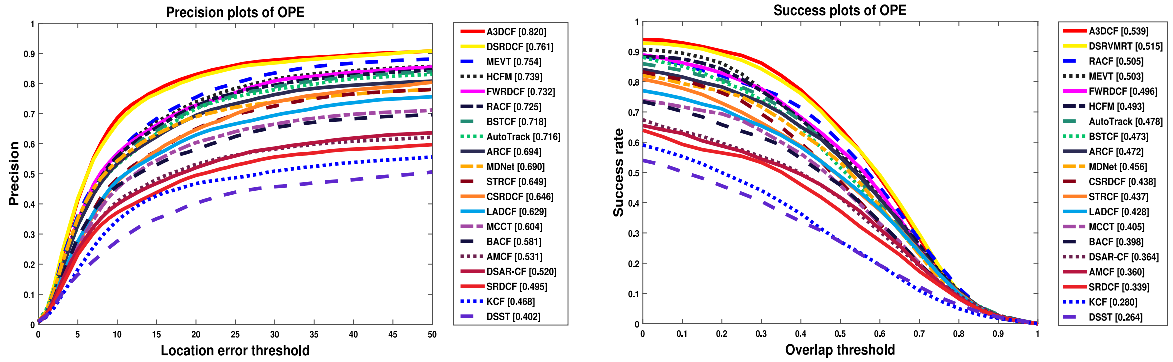
4.4.5. Experiments on UAVDT dataset

In Fig. 10, we exhibit the precision and success plots of the proposed DSRVMRT tracker on UAVDT, corresponding to the DP and AUC scores indicated in the figure legends. Our method outperforms all of the evaluated methodologies, obtaining a DP score of (73.8%) and an AUC value of (47.5%), respectively. Specifically, the EFSCF, HSIC-SRCF, and FWRDCF trackers achieve the best DP/AUC scores of 74.7%/47.6%, 76.2%/51.6% and 75.0%/51.0% on the UAVDT dataset. Moreover, ARCF was (1.8%/1.7%) lower than our method, and obtained the second-highest score, as seen in Figs. 10(a) and 10(b). Meanwhile, the DSRVMRT tracker surpasses the ensemble trackers MEVT, CF2, and HDT, with relative increases in the AUC score of (2.3%, 12% and 17.2%), correspondingly. Conversely, the ADNet and SiamFC trackers exhibit excellent performance, achieving DP and AUC scores of (68.3%/42.9%) and (68.1%/44.7%), respectively. Specifically, DSRVMRT shows superior performance compared to the BACF tracker, with enhancements in DP/AUC scores of (5.2%/4.3%), respectively. In particular, LADCF, SRDCF, and STRCF trackers achieve (67.8%/43.2%),



(a) (b)

Fig. 10. Precision and success plots on UAVDT.



(a) (b)

Fig. 11. Precision and success plots on DTB70.

(67.9%/42.3%), and (62.9%/41.1%) in precision and overlap scores. Additionally, the DSRVMRT approach enhances the AUC and DP scores of (16.8%) and (18.5%) with the KCF tracker. Our method exhibits notable advantages over conventional approaches, as illustrated in Fig. 10. The integration of multiple regularizers and deep networks has proven to be an effective means of enhancing the tracker's performance. In general, our model performs better and maintains greater stability.

4.4.6. Experiments on DTB70 dataset

To estimate the proposed tracker performance, we conducted experiments on the DTB70 dataset. On this dataset, the proposed DSRVMRT ranks first in terms of DP/AUC scores (76.1%/51.5%) as shown in Figs. 11(a) and 11(b). Technically, the FWRDCF tracker achieves best performance of 73.2% and 49.3% on the DTB70 dataset. Out of the 19 trackers that were compared, the AutoTrack, ARCF, and SRDCF trackers which reduce the boundary effects achieved the precision and AUC scores of (71.6%/47.8%), (69.4%/47.2%), and (49.5%/33.9%), respectively. On the other hand, DSRVMRT obtains the best performance increment of (0.7%/1.2%), (15.7%/11%) and (18%/15%) in the precision and success against the MEVT, MCCT-H, and MEEM methods. Furthermore, when compared to MDNet, LADCF, and DSAR-CF, our tracker exhibits comparable tracking outcomes. In conclusion, the proposed DSRVMRT algorithm has achieved tracking accuracy comparable to that of the most advanced trackers.

4.4.7. Experiments on GOT10K dataset

The GOT-10k dataset is a significant short-term tracking dataset, comprising over 10,000 videos for training and 180 videos for testing, with more than 1.5 million bounding boxes manually annotated for benchmarking. In compliance with the evaluation rules, models are only trained on the GOT-10k training dataset, and there are no overlapping object classes between the training set and the testing machine. To ensure fair evaluation, the model's performance is evaluated only on the GOT-10k test set after being trained on the GOT-10k training set. A comparison of our tracker's performance with other state-of-the-art trackers on the GOT-10k benchmark is presented in Fig. 12. The proposed tracker demonstrates a negligible decrease of only 0.38 in an average overlap (AO) score compared to the ATOM tracker. Fig. 12 further reveals that the proposed tracker outperforms the 3rd and 4th best-performing trackers (SiamRPN and DaSiamRPN trackers) by 0.35 and 0.74 in terms of AO score, respectively. Additionally, our tracker shows superior performance compared to SiamFC and MDNet by 0.17 and 0.219, respectively. Moreover, compared to the baseline BACF model, our DSRVMRT model remarkably improves the AO score by 0.258 points. Finally, our tracker shows significant improvements of 0.282 and 0.315 in the AO score compared to SRDCF and KCF models.

4.4.8. Experiments on LaSOT dataset

The LaSOT (Large-Scale Single Object Tracking) dataset is a substantial collection of video sequences that presents a wide range of challenges, establishing it as a thorough benchmark for assessing visual object tracking algorithms in complex, real-world environments. In addition, our study includes an evaluation of the DSRVMRT method, leveraging the LaSOT dataset alongside modern tracking algorithms. Recognized as a prominent benchmark, the LaSOT dataset is particularly valued for its extensive portrayal of various real-world scenarios and the difficulties commonly faced in object-tracking tasks. Fig. 13 illustrates the overall performance in terms of precision and success rate on the LaSOT testing set. At first glance, our method may appear average. However, it achieves a notable 15.5% improvement over the BACF method, which serves as the baseline for our approach.

Table 8
Comparative results of conventional trackers with the proposed tracker on the LaSOT dataset.

Methods	Metrics	DSRVMRT	HiFT	TCTrack++	MDNet	SiamFC	DSiam	STRCF	CFNet	BACF	CSRDCF	SRDCF	SAMF	DSST
LaSOT	DP	**46.2**	49.8	48.4	48.1	44.9	43.2	35.3	33.0	30.7	27.9	27.9	27.1	23.9
	AUC	**43.1**	47.0	46.5	45.6	40.9	39.2	33.1	28.6	28.5	25.2	27.9	24.7	23.2

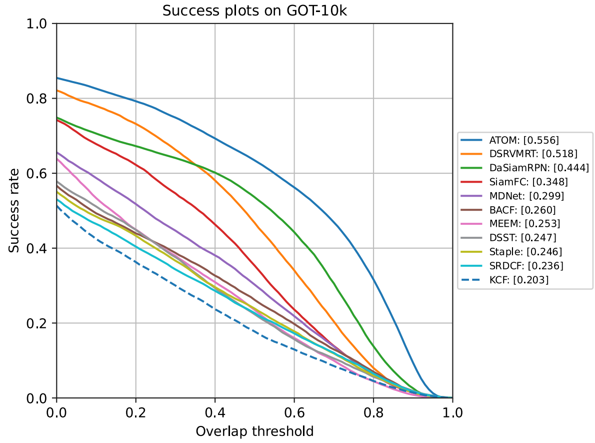


Fig. 12. Comparison with trackers on GOT-10k dataset.

The tracking performance of our method can be further enhanced by incorporating a more robust baseline method and an improved feature extraction model. Thanks to the advantages of multi-feature fusion, our method outperforms the SiamFC, Dsiam, and STRCF methods, increasing the DP score by 1.3%, 3%, and 10.9%, respectively. In comparison to the SRDCF tracker, the DSRVMRT tracker demonstrates a significant 18.3% improvement in precision. As shown in Table 8, our tracking algorithm surpasses scale trackers like SAMF and DSST, with improvements of 19.1% and 22.3% in distance precision scores, respectively. Overall, as shown in Fig. 13, our method consistently outperforms contemporary trackers.

4.4.9. Attribute evaluation

To delve deeper into the tracking performance of various methods across different scenarios, we carried out attribute-based experiments.

1. **OTB2013:** To comprehensively assess the effectiveness of our proposed method across various attributes, we analyzed different

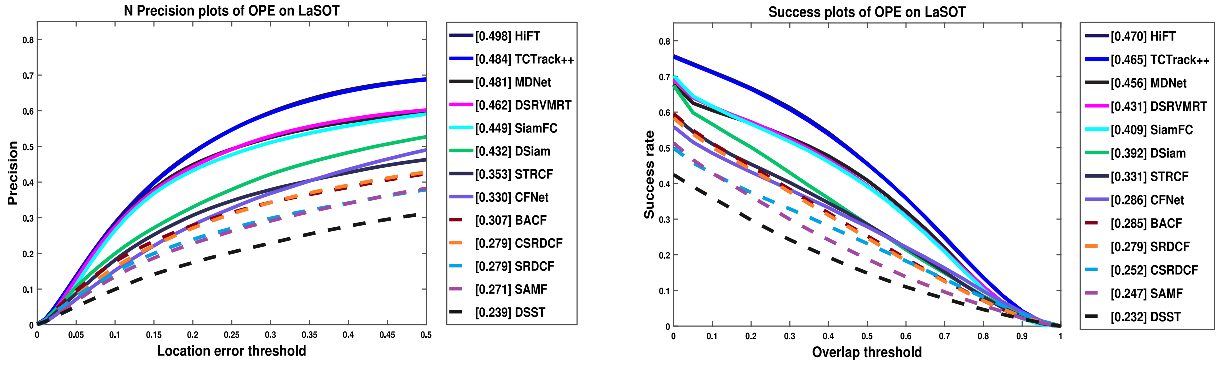
   attributes using the OTB 2013 benchmark. Fig. 14 displays the bar chart representation for various trackers across 11 attributes. DSRVMRT exhibits superior performance over conventional methods, particularly in the most challenging scenarios involving deformation, in-plane rotation, out-of-plane rotation, scale variation, and illumination variation in terms of distance precision. In addition, DSRVMRT continues to outperform HDT and CF2 in scenarios of occlusion, background clutters, and motion blur. This suggests that the presented method is effective in handling a variety of difficult situations. The superiority of DSRVMRT over BACF and KCF is evident from its significantly better performance across most attributes, achieving favorable precision and success rates. The success of our tracker can be attributed to the innovative strategies we have developed, including our channel reliability-aware approach and interval-based response inconsistency method.
2. **OTB2015:** In Table 9, we showcase the precision and success scores of 10 CF-based tracking methods, which include the DSRVMRT approach. Specifically, our DSRVMRT scheme obtains outstanding results for varying illumination, scale variation, fast motion, in-plane/out-of-plane rotation, deformation, and low-resolution attributes. Furthermore, our proposed approach demonstrates high effectiveness across various scenarios, adeptly managing complex scenes. In particular, the DSRVMRT method exhibits favorable outcomes across diverse challenging attributes, outperforming other methods within the 11 attributes outlined in the OTB2015 dataset. This result proves the proposed method effectively mitigates tracking failures caused by background distractions.
3. **Templecolor128 dataset:** Fig. 15 depicts a comparative analysis of tracking algorithms across different challenging attributes using the TempleColor128 dataset. In particular, our tracking approach surpasses the other tracking methods in scenarios of cluttered environments, target deformation, and rotation, which highlights its effectiveness in addressing changes in illumination and object deformation. Moreover, our presented approach maintains strong performance even when encountering challenges such as motion blur, fast motion, and low resolution. Finally, we enhance the tracking approach by integrating different feature response maps to ensure optimal efficiency.

4.4.10. Qualitative comparisons

To ensure the presented tracker performance, we conducted the qualitative analysis in various challenging sequences of the OTB2015 dataset as depicted in Fig. 16. In particular, the proposed tracker exhibited outstanding performance in situations characterized by occlusion and fast motion. For instance, when the fast motion in the Biker sequence at the $88th$ frame, the presented tracker performed well compared to the ARCF, AutoTracker, and BACF trackers. Specifically, when the tracker undergoes the target rotation in the MotorRolling sequence at the $97th$ frame, the ARCF, BACF, SRDCF, STRCF, and KCF trackers fail to track the target location. Despite the scale changes and target rotation in the Box and MotorRolling sequences, the proposed DSRVMRT tracker performed well compared to other trackers. In cases of complete or partial occlusion in the Lemming and Singer1 sequences, the ARCF, CF2, and KCF trackers encounter difficulties maintaining tracking of the target object. However, the proposed DSRVMRT tracker demonstrated superior performance in the scenario of occlusion. In conclusion, the DSRVMRT tracker exhibits excellent tracking efficiency in several challenging scenarios against state-of-the-art trackers.

4.5. Limitations and future work

As depicted in Fig. 17, the proposed tracker encounters challenges in specific scenarios. In the first and second rows, the tracked object is entirely obscured for an extended period. During this time, training patches lacking the object's appearance information are used for filter learning. Consequently, the interval response inconsistency offers limited contribution and fails to guide accurate tracking in cases of prolonged, complete occlusion. In the third row, the tracker struggles with large deformations and rapid rotations of the object amidst background clutter. These factors result in chaotic response maps, making it difficult to suppress distractions and accurately track the target. These shortcomings hinder the overall tracking the performance improvements. To address the performance limitations of DSRVMRT, future research will focus on strategies such as supervised tracking training and scale adjustments. These approaches aim to improve its ability to handle challenging conditions like motion blur and poor lighting. A key focus will be on verifying the generalizability of the proposed saliency features. Future studies will test their effectiveness not only



(a) (b)

Fig. 13. Precision and success plots on LaSOT.

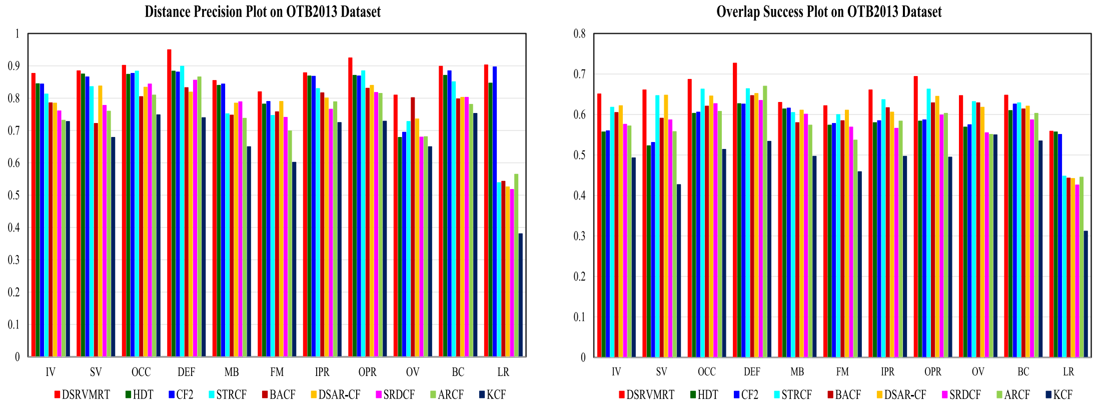


Fig. 14. Distance precision and overlap success plots on the OTB2013 dataset with 11 attributes.

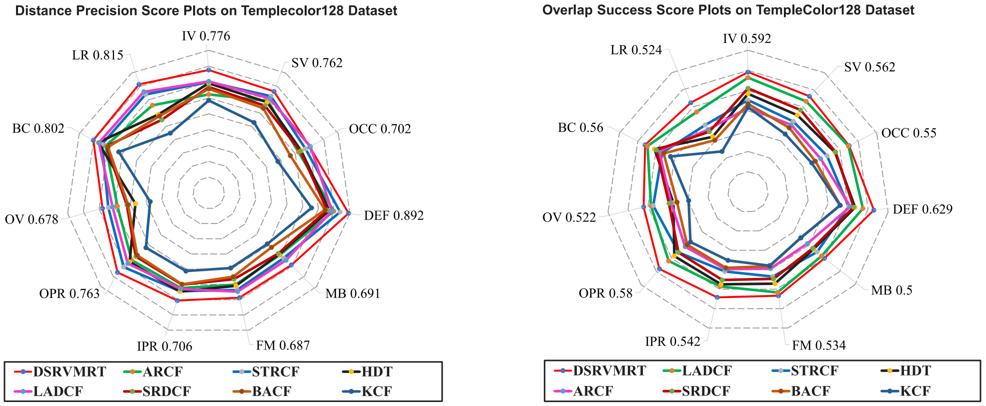


Fig. 15. Distance precision and overlap success plots on the TempleColor128 dataset with 11 attributes.

Table 9
Attribute-based evaluation of different trackers on OTB2015. The best results are highlighted in different color fonts.

Methods	Ours	STRCF	HDT	DSAR-CF	CF2	BACF	SRDCF	AutoTrack	ARCF	KCF
IV	90.0/68.7	84.1/65.3	82.0/53.5	81.8/64.9	81.7/54.0	81.4/63.2	79.2/61.3	78.7/60.1	76.6/59.9	71.9/47.9
SV	87.4/64.5	84.3/63.6	82.4/62.7	80.6/48.6	79.7/48.5	76.8/56.0	74.4/56.2	74.3/55.9	74.3/54.4	63.3/39.3
OCC	86.6/65.6	82.8/62.4	77.5/60.5	76.9/52.0	76.2/52.0	73.6/55.4	73.0/55.3	70.3/55.1	72.5/54.1	62.2/43.3
DEF	89.7/65.9	84.4/60.7	79.3/59.3	82.1/54.3	79.1/53.0	76.9/58.2	73.4/54.4	76.8/57.9	73.2/55.3	61.7/43.6
MB	87.3/66.7	84.5/67.0	81.9/64.6	79.4/56.3	79.7/57.3	77.4/61.8	78.2/61.0	73.9/57.9	77.6/61.2	61.8/45.6
FM	84.5/64.2	79.9/63.0	78.9/61.6	80.2/55.0	79.2/55.2	76.4/59.0	76.2/59.5	76.4/59.0	75.4/58.5	62.0/44.8
OPR	90.0/66.2	85.6/62.7	81.3/61.3	80.8/53.1	81.0/53.2	77.4/56.0	74.4/54.9	77.4/57.9	75.7/54.5	67.7/45.2
IPR	87.3/63.6	81.2/60.0	78.3/58.0	84.1/55.0	85.1/55.4	78.2/55.7	73.9/53.8	76.2/56.5	77.6/54.9	69.5/46.3
OV	89.2/66.0	75.7/57.8	72.7/57.5	68.6/50.0	69.9/50.1	69.6/52.9	62.4/48.9	70.9/54.0	72.4/54.7	53.4/42.0
BC	90.0/67.2	87.3/64.8	81.5/63.7	84.4/57.8	84.3/58.5	76.0/58.9	77.5/58.3	77.8/60.0	75.4/56.0	71.3/49.8
LR	82.0/59.6	75.6/55.8	75.8/57.1	76.6/42.0	78.7/42.4	71.4/49.4	63.1/48.0	69.0/50.6	72.9/51.5	54.6/30.7

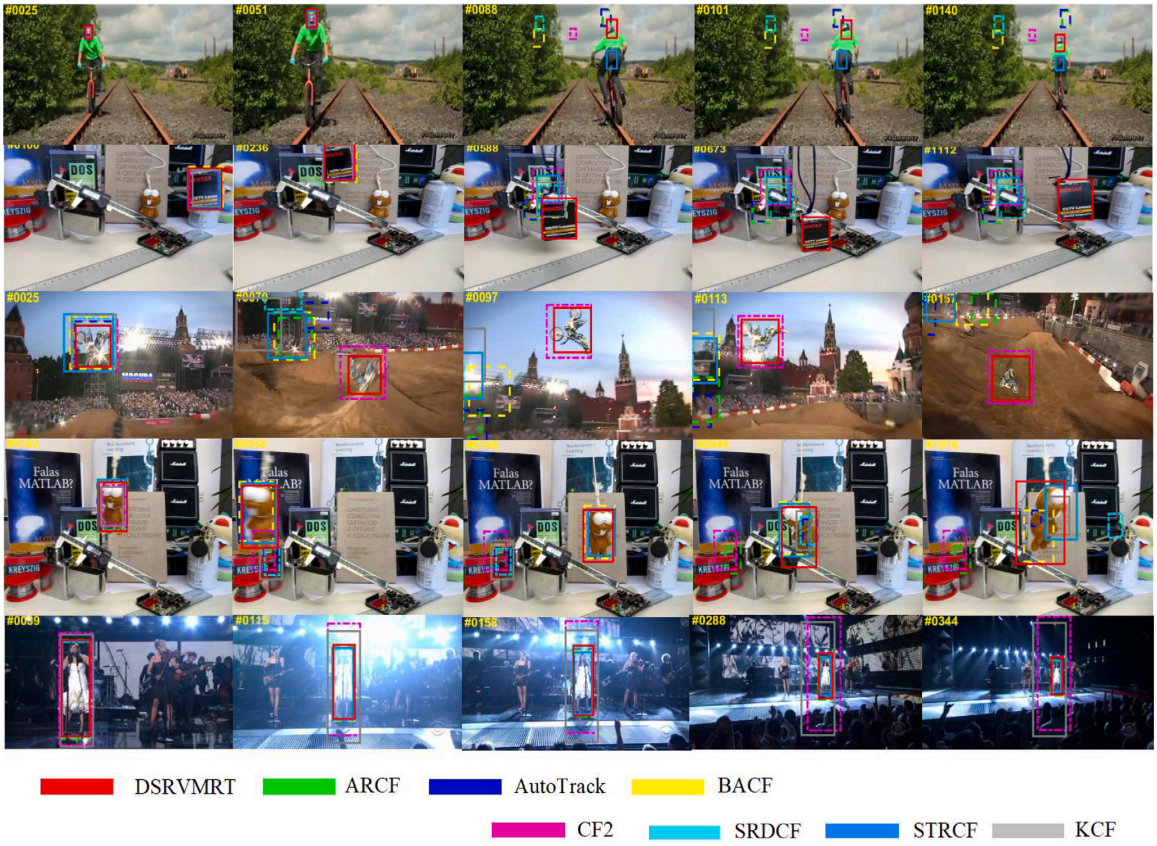


Fig. 16. Visual comparisons of tracking results between our proposed tracker and other state-of-the-art trackers, including BACF, CF2, SRDCF, KCF, ARCF, STRCF, and AutoTrack are provided in the OTB2015 dataset. The sequences are arranged in the following order from top to bottom: Biker, Box, MotorRolling, Lemming, and Singer1.

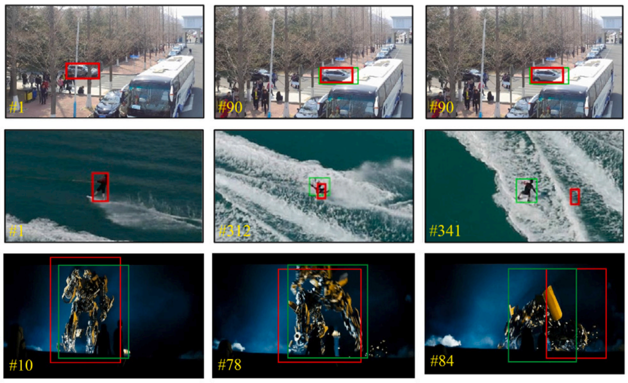


Fig. 17. Examples of failure cases in our tracking system. The green color box represents ground truth and the red color box represents the tracking result of the DSRVMRT method.

in DCF-based trackers but also in Siamese-based trackers to expand their applicability. Another area of improvement involves developing methods to more effectively integrate historical information. By incorporating this data adaptively, the tracker's performance in complex scenarios, such as those involving occlusion or dynamic object changes, can be significantly enhanced.

5. Conclusions

In this paper, we have presented a disruptor-suppressed response variation-aware multi-regularized correlation filter for robust visual tracking. Specifically, a response variation regularization approach has been proposed, which makes the filter change smoothly and prevents the over-fitting problem at the current frame. Moreover, the channel reliability weight distributions enhance the target localization and improve the tracking performance of the target with large appearance variations. Regarding the target temporal state diversity, the disruptor-suppressed regularization was designed to accurately reflect the changes in the appearance of the target during tracking. The experimental results of the seven challenging datasets have demonstrated the excellent performance of the proposed tracker compared to other modern trackers. To enhance the accuracy of monitoring, it will be necessary to explore and identify additional representative features in future work. In addition to promoting the object-sensing capabilities of the tracker, it is imperative to explore methods for generating effective regularizers that can further increase its accuracy.

CRediT authorship contribution statement

Sathishkumar Moorthy: Writing – original draft, Validation, Methodology, Investigation, Conceptualization. **Sachin Sakthi K.S.:** Writing – review & editing, Visualization, Validation, Investigation. **Sathiyamoorthi Arthanari:** Writing – original draft, Visualization, Validation. **Jae Hoon Jeong:** Writing – review & editing, Supervision, Project administration. **Young Hoon Joo:** Writing – review & editing, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the Basic Science Research Program under Grant NRF-2016R1A6A1A03013567 and Grant NRF-2021R1A2B5B01001484 and by the framework of the International Co-operation Program under Grant NRF-2022K2A9A2A06045121 through the National Research Foundation of Korea (NRF) funded by the Ministry of Education.

Data availability

Data will be made available on request.

References

[1] S. Moorthy, J.Y. Choi, Y.H. Joo, Gaussian-response correlation filter for robust visual object tracking, Neurocomputing 411 (2020) 78–90.

[2] S. Moorthy, Y.H. Joo, Formation control and tracking of mobile robots using distributed estimators and a biologically inspired approach, J. Electr. Eng. Technol. 18 (3) (2023) 2231–2244.

[3] D.S. Bolme, J.R. Beveridge, B.A. Draper, Y.M. Lui, Visual object tracking using adaptive correlation filters, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2010, pp. 2544–2550.

[4] J.F. Henriques, R. Caseiro, P. Martins, J. Batista, High-speed tracking with kernelized correlation filters, IEEE Trans. Pattern Anal. Mach. Intell. 37 (3) (2015) 583–596.

[5] J. Zhang, T. Yuan, Y. He, J. Wang, A background-aware correlation filter with adaptive saliency-aware regularization for visual tracking, Neural Comput. Appl. (2022) 1–18.

[6] M. Danelljan, G. Hager, F. Shahbaz Khan, M. Felsberg, Learning spatially regularized correlation filters for visual tracking, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 4310–4318.

[7] H. Kiani Galoogahi, A. Fagg, S. Lucey, Learning background-aware correlation filters for visual tracking, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 1135–1143.

[8] Z. Huang, C. Fu, Y. Li, F. Lin, P. Lu, Learning aberrance repressed correlation filters for real-time uav tracking, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 2891–2900.

[9] B. Kitt, A. Geiger, H. Lategahn, Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme, in: 2010 Ieee Intelligent Vehicles Symposium, IEEE, 2010, pp. 486–492.

[10] C. Fu, J. Ye, J. Xu, Y. He, F. Lin, Disruptor-aware interval-based response inconsistency for correlation filters in real-time aerial tracking, IEEE Trans. Geosci. Remote Sens. 59 (8) (2020) 6301–6313.

[11] D. Elayaperumal, Y.H. Joo, Visual object tracking using sparse context-aware spatio-temporal correlation filter, J. Vis. Commun. Image Represent. 70 (2020) 102820.

[12] M. Danelljan, G. Häger, F. Khan, M. Felsberg, Accurate scale estimation for robust visual tracking, in: British Machine Vision Conference, Nottingham, September 1-5, 2014, BMVA Press, 2014.

[13] M. Mueller, N. Smith, B. Ghanem, Context-aware correlation filter tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1396–1404.

[14] K. Dai, D. Wang, H. Lu, C. Sun, J. Li, Visual tracking via adaptive spatially-regularized correlation filters, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 4670–4679.

[15] S. Moorthy, Y.H. Joo, Adaptive spatial-temporal surrounding-aware correlation filter tracking via ensemble learning, Pattern Recognit. 139 (2023) 109457.

[16] D. Elayaperumal, Y.H. Joo, Robust visual object tracking using context-based spatial variation via multi-feature fusion, Inform. Sci. 577 (2021) 467–482.

[17] D. Yuan, X. Shu, N. Fan, X. Chang, Q. Liu, Z. He, Accurate bounding-box regression with distance-iou loss for visual tracking, J. Vis. Commun. Image Represent. 83 (2022) 103428.

[18] D. Yuan, X. Shu, Z. He, TRBACF: Learning temporal regularized correlation filters for high performance online visual object tracking, J. Vis. Commun. Image Represent. 72 (2020) 102882.

[19] A. Lukezic, T. Vojir, L. Cehovin Zajc, J. Matas, M. Kristan, Discriminative correlation filter with channel and spatial reliability, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6309–6318.

[20] T. Xu, Z.-H. Feng, X.-J. Wu, J. Kittler, Joint group feature selection and discriminative filter learning for robust visual object tracking, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 7950–7960.

[21] K. Nai, Z. Li, H. Wang, Learning channel-aware correlation filters for robust object tracking, IEEE Trans. Circuits Syst. Video Technol. 32 (11) (2022) 7843–7857.

[22] M. Jain, A. Tyagi, S. Denman, S. Sridharan, C. Fookes, Channel graph regularized correlation filters for visual object tracking, IEEE Trans. Circuits Syst. Video Technol. 32 (2) (2021) 715–729.

[23] X. Lu, C. Ma, B. Ni, X. Yang, Adaptive region proposal with channel regularization for robust object tracking, IEEE Trans. Circuits Syst. Video Technol. 31 (4) (2019) 1268–1282.

[24] L. Fan, L. Zhang, Multi-system fusion based on deep neural network and cloud edge computing and its application in intelligent manufacturing, Neural Comput. Appl. 34 (5) (2022) 3411–3420.

[25] S. Moorthy, Y.H. Joo, Learning dynamic spatial-temporal regularized correlation filter tracking with response deviation suppression via multi-feature fusion, Neural Netw. 167 (2023) 360–379.

[26] C. Ma, J.-B. Huang, X. Yang, M.-H. Yang, Hierarchical convolutional features for visual tracking, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3074–3082.

[27] S. Moorthy, Y.H. Joo, Multi-expert visual tracking using hierarchical convolutional feature fusion via contextual information, Inform. Sci. 546 (2021) 996–1013.

[28] S.S. KS, Y.H. Joo, J.H. Jeong, Keypoint prediction enhanced siamese networks with attention for accurate visual object tracking, Expert Syst. Appl. (2024) 126237.

[29] Y. Wu, J. Lim, M.-H. Yang, Object tracking benchmark, IEEE Trans. Pattern Anal. Mach. Intell. 37 (9) (2015) 1834–1848.

[30] P. Liang, E. Blasch, H. Ling, Encoding color information for visual tracking: Algorithms and benchmark, IEEE Trans. Image Process. 24 (12) (2015) 5630–5644.

[31] L. Huang, X. Zhao, K. Huang, Got-10k: A large high-diversity benchmark for generic object tracking in the wild, IEEE Trans. Pattern Anal. Mach. Intell. 43 (5) (2019) 1562–1577.

[32] M. Mueller, N. Smith, B. Ghanem, A benchmark and simulator for uav tracking, in: Proceedings of the European Conference on Computer Vision, ECCV, Springer, 2016, pp. 445–461.

[33] D. Du, Y. Qi, H. Yu, Y. Yang, K. Duan, G. Li, W. Zhang, Q. Huang, Q. Tian, The unmanned aerial vehicle benchmark: Object detection and tracking, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 370–386.

[34] H. Fan, L. Lin, F. Yang, P. Chu, G. Deng, S. Yu, H. Bai, Y. Xu, C. Liao, H. Ling, Lasot: A high-quality benchmark for large-scale single object tracking, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 5374–5383.

[35] S. Li, D.-Y. Yeung, Visual object tracking for unmanned aerial vehicles: A benchmark and new motion models, in: Proceedings of the AAAI Conference on Artificial Intelligence, AAAI, 2017, pp. 4140–4146.

[36] T. Xu, Z.-H. Feng, X.-J. Wu, J. Kittler, Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking, IEEE Trans. Image Process. 28 (11) (2019) 5596–5609.

[37] W. Feng, R. Han, Q. Guo, J. Zhu, S. Wang, Dynamic saliency-aware regularization for correlation filter-based object tracking, IEEE Trans. Image Process. 28 (7) (2019) 3232–3245.

[38] Y. Qi, S. Zhang, L. Qin, H. Yao, Q. Huang, J. Lim, M.-H. Yang, Hedged deep tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4303–4311.

[39] J. Zhang, S. Ma, S. Sclaroff, MEEM: robust tracking via multiple experts using entropy minimization, in: Proceedings of the European Conference on Computer Vision, ECCV, Springer, 2014, pp. 188–203.

[40] B. Li, J. Yan, W. Wu, Z. Zhu, X. Hu, High performance visual tracking with siamese region proposal network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8971–8980.

[41] N. Wang, W. Zhou, Q. Tian, R. Hong, M. Wang, H. Li, Multi-cue correlation filters for robust visual tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4844–4853.

[42] F. Li, C. Tian, W. Zuo, L. Zhang, M.-H. Yang, Learning spatial-temporal regularized correlation filters for visual tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4904–4913.

[43] L. Bertinetto, J. Valmadre, J.F. Henriques, A. Vedaldi, P.H. Torr, Fully-convolutional siamese networks for object tracking, in: Computer Vision–ECCV 2016 Workshops: Amsterdam, the Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II 14, Springer, 2016, pp. 850–865.

[44] Y. Li, C. Fu, F. Ding, Z. Huang, G. Lu, AutoTrack: Towards high-performance visual tracking for UAV with automatic spatio-temporal regularization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 11923–11932.

[45] Z. Zhu, Q. Wang, B. Li, W. Wu, J. Yan, W. Hu, Distractor-aware siamese networks for visual object tracking, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 101–117.

[46] M. Danelljan, G. Bhat, F.S. Khan, M. Felsberg, Atom: Accurate tracking by overlap maximization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 4660–4669.

[47] H. Nam, B. Han, Learning multi-domain convolutional neural networks for visual tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4293–4302.

[48] L. Xu, P. Kim, M. Wang, J. Pan, X. Yang, M. Gao, Spatio-temporal joint aberrance suppressed correlation filter for visual tracking, Complex Intell. Syst. (2021) 1–13.

[49] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, P.H. Torr, Staple: Complementary learners for real-time tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1401–1409.

[50] Y. Li, J. Zhu, A scale adaptive kernel correlation filter tracker with feature integration, in: Proceedings of the European Conference on Computer Vision, ECCV, Springer, 2014, pp. 254–265.

[51] X.-F. Zhu, X.-J. Wu, T. Xu, Z.-H. Feng, J. Kittler, Robust visual object tracking via adaptive attribute-aware discriminative correlation filters, IEEE Trans. Multimed. 24 (2021) 301–312.

[52] J. Zhang, Y. Liu, H. Liu, J. Wang, Y. Zhang, Distractor-aware visual tracking using hierarchical correlation filters adaptive selection, Appl. Intell. 52 (6) (2022) 6129–6147.

[53] J. Zhang, Y. He, W. Feng, J. Wang, N.N. Xiong, Learning background-aware and spatial-temporal regularized correlation filters for visual tracking, Appl. Intell. (2022) 1–16.

[54] L. Xu, P. Kim, M. Wang, J. Pan, X. Yang, M. Gao, Spatio-temporal joint aberrance suppressed correlation filter for visual tracking, Complex & Intell. Syst. (2022) 1–13.

[55] C. Fu, W. Xiong, F. Lin, Y. Yue, Surrounding-aware correlation filter for UAV tracking with selective spatial regularization, Signal Process. 167 (2020) 107324.

[56] Y. Li, C. Fu, F. Ding, Z. Huang, J. Pan, Augmented memory for correlation filters in real-time UAV tracking, in: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, IEEE, 2020, pp. 1559–1566.

[57] J. Wen, H. Chu, Z. Lai, T. Xu, L. Shen, Enhanced robust spatial feature selection and correlation filter learning for UAV tracking, Neural Netw. 161 (2023) 39–54.

[58] T. Liu, J. Li, J. Wu, J. Chang, Y. Xiao, Y. Hong, Visual tracking with dumbbell selection network, Neurocomputing 516 (2023) 77–91.

[59] S.S. Kuppusami Sakthivel, S. Moorthy, S. Arthanari, J.H. Jeong, Y.H. Joo, Learning a context-aware environmental residual correlation filter via deep convolution features for visual object tracking, Math. 12 (14) (2024) 2279.

[60] Z. An, X. Wang, B. Li, J. Fu, Learning spatial regularization correlation filters with the hilbert-schmidt independence criterion in rkhs for uav tracking, IEEE Trans. Instrum. Meas. 72 (2023) 1–12.

[61] X. Wang, F. Ma, X. Wang, C. Chen, Learning feature-weighted regularization discriminative correlation filters for real-time UAV tracking, Signal Process. 228 (2025) 109765.

[62] W. Hu, Q. Wang, L. Zhang, L. Bertinetto, P.H. Torr, Siammask: A framework for fast online object tracking and segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 45 (3) (2023) 3072–3089.

[63] S. Li, S. Zhao, B. Cheng, J. Chen, Part-aware framework for robust object tracking, IEEE Trans. Image Process. 32 (2023) 750–763.

[64] Z. Cao, C. Fu, J. Ye, B. Li, Y. Li, Hift: Hierarchical feature transformer for aerial tracking, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 15457–15466.

[65] Z. Cao, Z. Huang, L. Pan, S. Zhang, Z. Liu, C. Fu, Towards real-world visual tracking with temporal contexts, IEEE Trans. Pattern Anal. Mach. Intell. 45 (12) (2023) 15834–15849.

[66] H. Huang, G. Liu, Y. Zhang, R. Xiong, Feature distillation siamese networks for object tracking, Appl. Soft Comput. 132 (2023) 109912.

[67] J. Valmadre, L. Bertinetto, J. Henriques, A. Vedaldi, P.H. Torr, End-to-end representation learning for correlation filter based tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2805–2813.