Learning Multi-Regularized Mutation-Aware Correlation Filter for Object Tracking via an Adaptive Hybrid Model

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.*

Abstract

Discriminative Correlation Filters (DCF) have emerged as a popular and effective approach in object tracking. With promising performance and efficiency, DCF-based trackers achieved impressive attention and reliable tracking results in several challenging scenarios. Although DCF-based trackers improve tracking performance, they still suffer from unexpected factors such as appearance mutations, filter degradation, and target distortion, which leads to decreased tracker performance. To address these challenges, a novel Multi-Regularized Mutation-Aware Correlation Filter (MRMACF) approach is presented. To do this, we propose a mutation-aware strategy with an adaptive hybrid model that employs the mutation threat mechanism technique to effectively handle the appearance mutations and filter degradation issues when the filter deviates from the target location. The mutation threat mechanism identifies sudden and significant changes in the target object's appearance, which is achieved by an adaptive hybrid model approach that compares the current appearance with recent historical models. Following that, we introduce an improved sparse spatial feature selection approach that incorporates row and column-based feature selection methods into the sparse spatial technique, which aims to identify crucial features within the target region and successfully address the problem of target distortion. Moreover, the surrounding-aware approach is presented that extracts the surrounding samples of the target region to utilize the context information, which prevents the filter deviation from the target and improves the discriminative ability. Specifically, the adaptive hybrid model approach is proposed to mitigate both tracking drift and the mutation threat of target by incorporating target position information from previous frames. Finally, we showcase the efficiency of the proposed MRMACF approach against existing modern trackers using the OTB-2013, OTB-2015, TempleColor-128, UAV-123, UAVDT, VOT-2018, LaSOT, and GOT-10K benchmark datasets.

Keywords: Correlation Filter, Object Tracking, Surrounding-Aware Method, Temporal Regularization, Mutation-Aware Approach, and Adaptive Hybrid Model.

*Corresponding author, E-mail address : yhjoo@kunsan.ac.kr (Young Hoon Joo).

1. Introduction

Visual tracking is an essential task in computer applications that involves estimating the location and movement of a target object in a video sequence over time. It holds a vital role in numerous applications, such as surveillance, robotics, autonomous vehicles, augmented reality, sports analytics, and other domains [1, 2, 3, 4]. The main objective of object tracking is to continuously monitor the location of a specific target in a dynamic visual environment without distractions. Despite the notable tracking performance across various applications, existing trackers continue to encounter challenges in accurately determining the target's appropriate location due to factors such as illumination changes, occlusion, cluttered backgrounds, camera motion, and variations in appearance. To address these challenges, several tracking approaches, such as correlation filters and deep learning-based trackers have been developed in object tracking.

Recently, discriminative correlation filters have gained much popularity among researchers because of their excellent tracking results and computational performance. In particular, DCF-based trackers manipulate the frequency domain transformed from the time domain to obtain significant tracking speed with the help of Fast Fourier Transform (FFT). Thereafter, positive and negative samples are trained in the frequency domain to improve tracking performance. Despite the frequency domain enhancing the computational efficiency, the negative samples acquire the boundary effects due to the circulant shift process, which significantly decreases the tracker performance. In recent times, several tracking approaches such as BACF [5], SRDCF [6], and BSTCF [7] have been proposed to overcome the boundary effect issue. In this regard, the authors in [5] have presented the BACF tracker that adapts to changes in the appearance of the target object and its surroundings to track the target object accurately. Following that, the BACF tracker keeps the non-zero value within the target region, which significantly reduces the boundary effect issue. Specifically, the authors in [6] have introduced the SRDCF tracker that utilizes spatial-regularization techniques, which can helps to reduce the impact of boundary effects by capturing the object's appearance and motion more accurately. Moreover, the authors in [7] presented the BSTCF tracking approach, which helps to alleviate the impact of the boundary effect by taking spatial and temporal information into account. Even though the correlation filter-based trackers enhanced tracking performance, the tracker may lose its precise target location as the tracking approach becomes more complex. To address this issue, the spatio-temporal technique and surrounding-aware approach have been taken into account in several analyses [8], [9]. The authors in [8] have introduced the STRCF tracker, which integrates both spatial regularization and temporal techniques that helps to handle complex regions by modeling the spatial variations in the object appearance. Following that, the authors in [9] have introduced the SASR tracker, which incorporates context information and selective spatial regularization approaches. These approaches helps to handle the complex region, which improves the

tracking accuracy by reducing the effect of the boundary condition. Despite DCF-based trackers improving their performance by using the hand-crafted feature, they still suffer from unavoidable problems such as object rotation, object deformation, and occlusion. As a result of the above discussions, deep convolutional feature-based trackers can solve these issues by providing robust feature extraction ability.

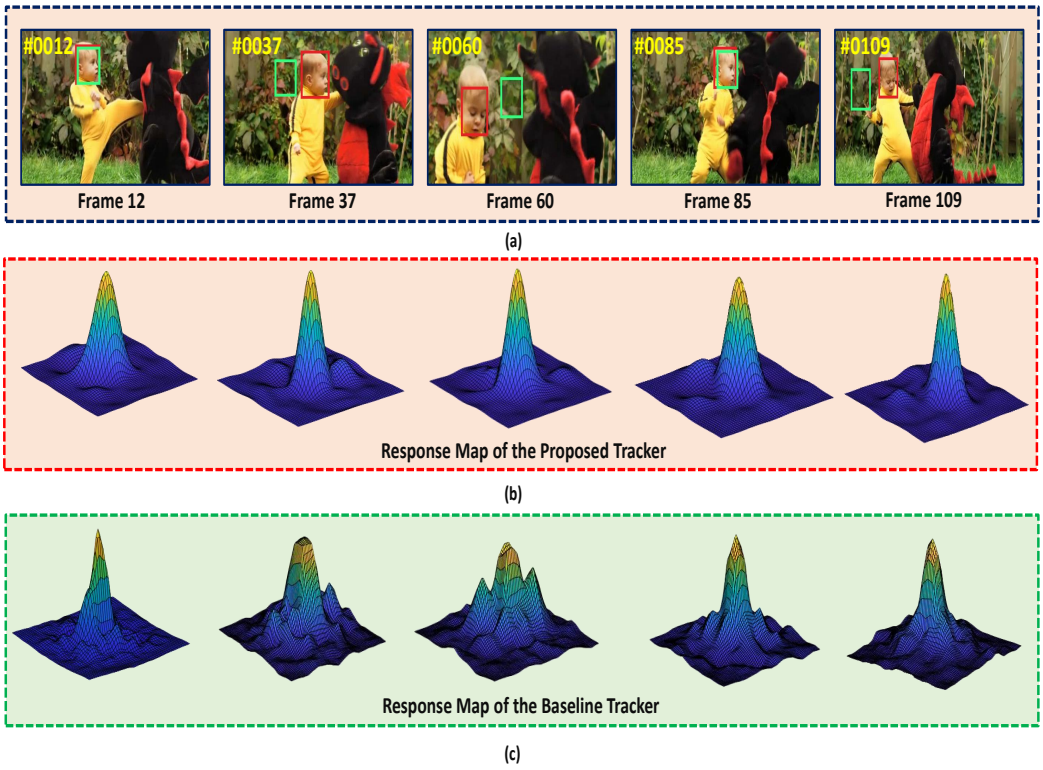


Figure 1: We conducted a performance comparison between the proposed tracker (highlighted in red) and the baseline tracker (highlighted in green) using the DraganBaby sequence on the OTB-2015 dataset. As shown in Fig. 1(a), the proposed MRMACF tracker demonstrates superior performance against the baseline tracker. Also, the response map of the proposed tracker is exhibited in Fig. 1(b). Following that, the response map of the baseline tracker is illustrated in Fig. 1(c). From these analyses, our proposed tracker performs well when the tracker undergoes fast motion and sudden changes in the target.

On the other hand, deep feature-based approaches have obtained remarkable performance in object tracking due to their robust feature extraction ability. In addition, the deep trackers utilize Convolutional Neural Networks (CNNs) to learn features from the object and the surrounding context, which improves the discriminative power of the tracker. In recent times many researchers have focused the deep feature-based methods such as SiamRPN [10], DaSiamRPN [11], VGGNet [12], and ResNet [13]. In this respect, the authors in [10] have presented the SiamRPN tracking approach, which is an efficient and effective offline tracking method that can achieve significant performance in object-tracking tasks. To be more specific, the Siamese network is trained to learn a correlation filter that can accurately predict the target location in subsequent frames. Furthermore, the authors in [11] have provided the DaSiamRPN tracker that uses a combination of correlation filters and a Region Proposal Network (RPN), which helps to track the objects

more accurately in video sequences. Besides, the authors in [12] employed a pre-trained VGGNet model for extracting the multiple convolutional layers that enable the recognition of the target from various scenarios. Also, the VGGNet model incorporates the responses of the Conv-3, Conv-4, and Conv-5 layers to achieve accurate target locations and improve tracking performance. Although the VGGNet model increased tracker performance by using feature extraction, there are some potential issues with using it, including computational cost, irrelevant features, and limited adaptability to changes in the target object's appearance. To address these concerns, the authors in [13] have exploited a pre-trained ResNet architecture, which is used to extract features more quickly than traditional methods and obtain adequate target representation. These features are used as inputs to the correlation filter algorithm, which estimates the position and scale of the target object in the subsequent frames. To take advantage of the rich target representation, we exploited a ResNet model, which is used for robust feature extraction and accurate target prediction during the tracking process.

Motivated by the above discussions, we aim to present a Multi-Regularized Mutation-Aware Correlation Filter (MRMACF) for object tracking. Moreover, the main contribution of the proposed approach is described as follows:

1. We proposed a mutation-aware strategy with an adaptive hybrid model approach that integrates the Mutation Threat Mechanism (MTM) technique, which helps to redetect the target region when the target appearance is affected by the mutation.
2. The sparse spatial feature selection approach is presented that integrates row and column-based feature selection methods into the sparse spatial technique, which enables the identification of critical features within the target region to effectively address target distortion issues.
3. A surrounding-aware approach is proposed to enhance discriminative ability by taking into account the spatial relationships between the target object and its surrounding context. This consideration of spatial relationships assists in more accurate identification of the target object when the target is affected by the fast motion.
4. An adaptive hybrid model with a temporal technique is introduced to prevent the tracker deviations from the target region, effectively addressing both tracking drift and mutation issues. By using this approach, the filter efficiently differentiates the target from the background region, improving the overall tracking accuracy.
5. Finally, the extensive experiments on the public datasets OTB-2013, OTB-2015, TempleColor-128, UAV-123, UAVDT, VOT-2018, LaSOT, and GOT-10K demonstrate the superiority of our proposed approach over other state-of-the-art trackers.

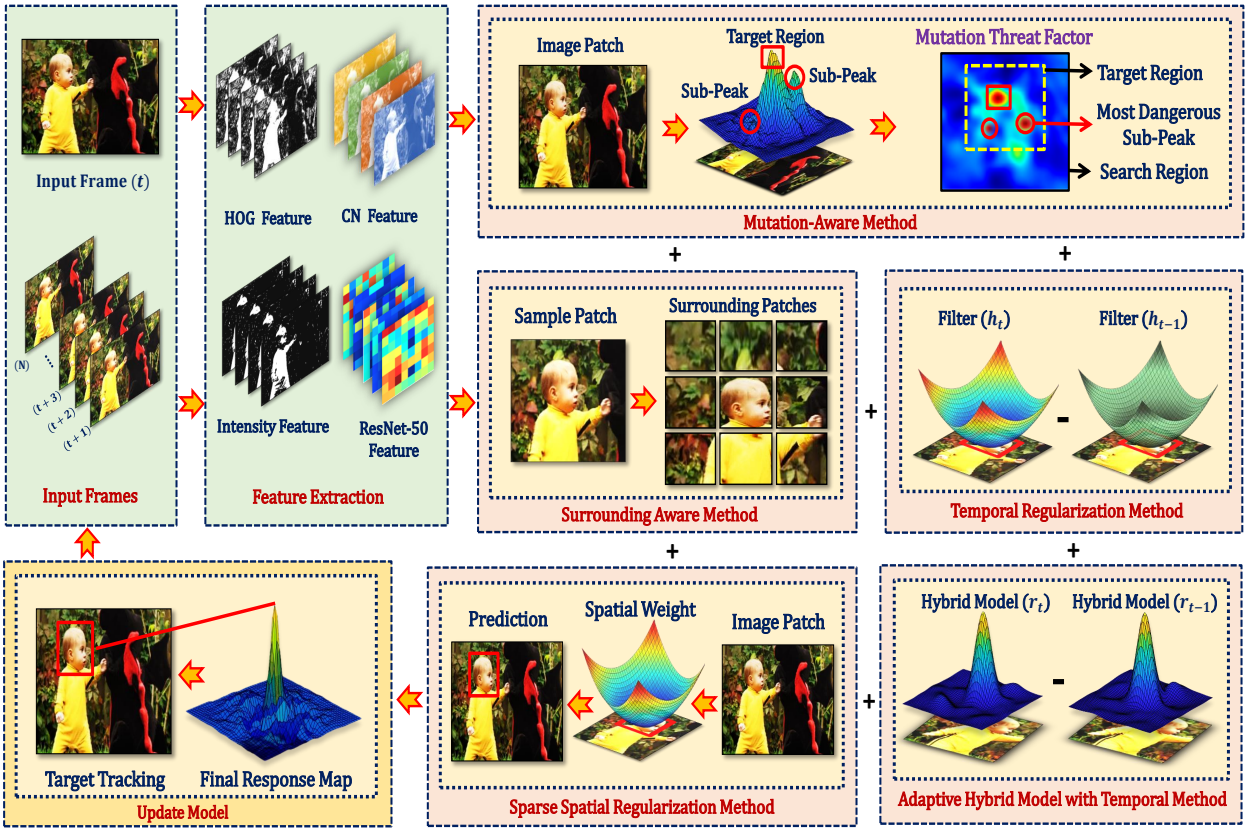


Figure 2: The overall schematic diagram of the proposed MRMACF method.

2. Related Works

In this section, we briefly discuss the conventional correlation filter-based approaches. We first analyze the correlation filter-based methods in subsection 2.1. Then, we investigate the deep feature-based tracking approaches in subsection 2.2. Following that, we examine the transformer-based methods in subsection 2.3. The comparative results of the baseline and proposed trackers are illustrated in Fig. 1, and the schematic diagram of the proposed method is exhibited in Fig. 2.

2.1. Tracking based on Correlation Filter Approaches

In recent times, DCFs have attained promising results among the research community due to their accuracy, robustness, and efficiency. Accordingly, many researchers have focused the DCF-based tracking approaches such as BACF [5], SRDCF [6], STRCF [8], SASR [9], and LADCF [14]. In this regard, the authors in [5] have presented the BACF tracker that helps to effectively handle the boundary effect issue by integrating the image patch and binary mask in the target region. Also, the BACF tracker has the ability to track the target object at different scale variations. Therefore, the BACF approach increases the robustness and reliability of the filter and improves the tracker's performance as well. Besides, the authors

in [6] have introduced the SRDCF tracker, which can effectively suppress background clutter and improve the discriminative power of the filter by regularizing the filter coefficients in the spatial domain. The main encumbrance of the SRDCF tracker is its limited ability to handle fast-moving objects. For this reason, it may struggle to track the target objects precisely. To overcome this encumbrance, the STRCF [8] tracker is designed to handle fast-moving objects by integrating both spatial and temporal approaches into the tracking model. Also, the STRCF tracker uses a fixed-size filter to represent the object's appearance. Thus, the tracker may not be able to track the target object properly during the scale changes. To address this limitation, the authors in [14] presented the LADCF tracker, which better handles deformable objects and scale variation by using a multi-resolution feature representation. Moreover, DCF-based tracking methods are typically impacted by object rotation, which affects tracking performance. To overcome these challenges, the authors in [9] have presented the SASR tracking approach, which can effectively handle object rotation by incorporating an adaptive model update strategy based on the estimated rotation angle. Although these tracking methods have enhanced tracking performance, tracking drift makes the process more difficult during tracking. To avoid this problem, we presented the multi-regularized mutation-aware correlation filter with an adaptive hybrid model approach, which can effectively handle the tracking drift issue.

2.2. Tracking based on Deep Feature Approaches

In recent times, deep feature-based approaches have obtained remarkable attention, which has enhanced popularity among the tracking community due to their superior tracking performance and feature extraction ability in object tracking. In this regard, many researchers have integrated the hand-crafted and deep features into the DCF-based tracking methods such as BSTCF [7], A3DCF [15], STAR [16], MEVT [17], and HCFM [18]. The authors in [7] have provided the BSTCF tracker, which incorporates spatial and temporal information into the filter to obtain better tracking performance. Further, the BSTCF tracker utilized the VGG networks to enhance the tracker's accuracy and robustness. Despite the BSTCF tracker's enhanced tracker performance, it can suffer from sudden changes caused by background appearance such as moving objects or changes in lighting conditions. To avoid this problem, the authors in [15] have introduced the A3DCF tracker, which employs several mechanisms to handle sudden changes in background appearance. Specifically, the A3DCF tracker leverages attribute-aware feature representation, which captures various visual attributes of the target object, such as color, texture, shape, or motion. By considering multiple attributes, the tracker becomes more robust to background changes and can distinguish the target object from the background clutter. Although the A3DCF tracker has the ability to handle background changes, the tracker may still face challenges when it comes to maintaining accurate tracking over long durations. To address these constrain, the authors in [16] have presented the STAR tracker, which employed adaptive

model update mechanisms to continuously adapt the correlation filter and suppress potential drift. It updates the model using a combination of the current frame's information and the accumulated knowledge from previous frames. By adapting the model over time, the tracker can better handle appearance changes and maintain accurate tracking. Likewise, the authors in [17] have introduced the MEVT tracker, which utilized an ensemble tracking strategy via a multi-layer convolutional feature fusion technique. Also, the MEVT tracker incorporates the different convolutional layer features from Conv3-4, Conv4-3, Conv4-4, Conv5-4, and HOG features to track the target object more accurately. Moreover, the authors in [18] have provided the HCFM tracker, which used the VGGNet-19 deep learning network for robust feature extraction. To be more specific, the HCFM tracker reduces the unwanted background region response weights by providing a distractor-aware map. Therefore, the HCFM tracker focuses on the target region to enhance the tracker's efficiency. Inspired by the above discussion, we present a multi-feature fusion approach that integrates hand-crafted and deep convolutional features such as the HOG, Intensity, ColorName, and ResNet models for robust feature extraction ability.

2.3. Tracking based on Transformer Approaches

Recently, transformers-based architecture introduced for natural language processing tasks has demonstrated remarkable success in various computer vision domains, prompting researchers to explore their potential for object tracking. The transformer's ability to capture long-range dependencies and contextual information makes it well-suited for addressing the challenges posed by object appearance variations, occlusions, and scale changes. In the context of object tracking, transformer-based methods [19, 20, 21] leverage the self-attention mechanism to efficiently model the relationships between different regions of an image. The authors in [19] introduced the RPformer architecture, specifically designed to effectively incorporate the feature relationship between the template and the search region. This design facilitates to capture the rich contextual information, thereby minimizing information loss and enhancing the utilization of global feature information. Following that, SIFT [20] proposes a transformer-based tracker that effectively combines spatial and temporal features for robust object tracking, which utilizes a spatial attention mechanism to capture the target object's appearance in the current frame and a temporal attention mechanism to learn the object's motion patterns across frames. Moreover, the authors in [21] propose a novel transformer-based object tracker that effectively utilizes a deformable transformer module and spatio-temporal information to handle object deformations and long-range appearance dependencies, which helps to enhance the tracker's efficiency and robustness.

3. Proposed Method and Implementation

In this part, we mainly explore the proposed multi-regularized mutation-aware correlation filter via an adaptive hybrid model. First, the baseline tracker BACF approach is examined in subsection 3.1. Then, we discussed the sparse spatial feature selection technique in subsection 3.2. In addition, we presented the surrounding-aware approach in subsection 3.3. Further, the temporal regularization strategy is investigated in subsection 3.4. Then, we discuss the mutation-aware technique in subsection 3.5. Moreover, the optimal adaptive hybrid model and an adaptive hybrid model with temporal regularization techniques are presented in subsections 3.6 and 3.7. Subsequently, we explore the implementation intricacies of the proposed approach in the remaining sections.

3.1. Background-Aware Correlation Filter

Initially, we utilize the BACF [5] tracker as our baseline tracker, which aims to maximize the correlation response of the target region while minimizing the response from the background region. Furthermore, the BACF tracker often utilizes FFT for efficient correlation calculation between the filter and the input image. Moreover, by cropping the more true negative sample using a binary matrix, which effectively handles the boundary effect issues. Hence, the objective function can be described in the following manner:

$$E(h,r) = \frac{1}{2}\sum_{k=1}^{C}\left\| y^k - \sum_{k=1}^{C} X^k \circledast (B^{\mathsf{T}} h^k)\right\|_{\mathcal{F}}^2 + \frac{\lambda_1}{2}\sum_{k=1}^{C}\left\|h^k\right\|_{\mathcal{F}}^2. \tag{1}$$

where $X \in \mathbb{R}^T$ denotes the training sample, y represents the correlation output, and $h \in \mathbb{R}^{\mathcal{D}}$ denotes the filter. Besides, B indicates the $\mathcal{D} \times T$ binary matrix and C denotes the feature channels. Further, $\circledast$ represents the circular convolution operation, T indicates the transpose matrix, λ_1 is a penalty factor, and $\mathcal{F}$ denotes the Frobenius norm.

3.2. Sparse Spatial Feature Selection

The correlation filter-based trackers typically suffer from boundary effects and target distortion issues, which considerably degrade the tracking efficiency. To overcome this problem, we present the $\ell_{2,1}$ norm-based sparse spatial feature selection approach, which effectively utilizes the row and column-based feature selection techniques to obtain the prominent features in the target region. The comprehensive explanation of the sparse spatial feature selection method is described in Remark 2. Further, we described the objective function as follows:

$$E(h,r) = \frac{1}{2}\sum_{k=1}^{C}\left\| r^k - \sum_{k=1}^{C} X^k \circledast (B^{\mathsf{T}} h^k)\right\|_{\mathcal{F}}^2 + \frac{\lambda_1}{2}\sum_{i=1}^{M}\left\|h_{i:}^k\right\|_{\mathcal{F}} + \frac{\lambda_2}{2}\sum_{j=1}^{N}\left\|h_{:j}^k\right\|_{\mathcal{F}}, \tag{2}$$

where r^k denotes the adaptive hybrid model. Specifically, the detailed description of the r^k is illustrated in Remark 1.

In Eq. (2), the second and third terms can be described as follows:

$$\sum_{i=1}^{M} \left\| h_{i:}^{k} \right\|_{\mathcal{F}} = \sum_{i=1}^{M} \sqrt{\sum_{j=1}^{N} \sum_{k=1}^{C} (h_{ij}^{k})^2}, \tag{3}$$

$$\sum_{j=1}^{N} \left\| h_{:j}^{k} \right\|_{\mathcal{F}} = \sum_{j=1}^{N} \sqrt{\sum_{i=1}^{M} \sum_{k=1}^{C} (h_{ij}^{k})^2}. \tag{4}$$

Eq. (3) denotes the row-based $\ell_{2,1}$-norm and Eq. (4) represents the column-based $\ell_{2,1}$-norm, respectively. In particular, the key principle of row and column-based methods is to implement structural sparsity that enhances the discriminative ability of the filter.

3.3. Surrounding-Aware Approach

A surrounding-aware tracking technique is presented, which takes advantage of context information surrounding the target region to reduce tracking drift and improve tracking performance. Moreover, the surrounding-aware approach incorporates online updating mechanisms, allowing the filters to adapt and learn from new information during the tracking process. By continuously updating the filters based on the target object and its surrounding context, the approach can effectively handle appearance changes, occlusions, and other challenges encountered in real-time tracking scenarios. This adaptability enhances the robustness of the filter by leveraging contextual information effectively. Therefore, the formulation of the objective function is defined in the following way:

$$E(h, r) = \frac{1}{2} \sum_{k=1}^{C} \left\| r^k - \sum_{k=1}^{C} X^k \circledast (B^{\mathsf{T}} h^k) \right\|_{\mathcal{F}}^2 + \frac{\lambda_1}{2} \sum_{i=1}^{M} \left\| h_{i:}^{k} \right\|_{\mathcal{F}} + \frac{\lambda_2}{2} \sum_{j=1}^{N} \left\| h_{:j}^{k} \right\|_{\mathcal{F}} + \frac{1}{2} \sum_{p=1}^{P} \left\| \sum_{k=1}^{C} \alpha^k X_p^k \circledast (B^{\mathsf{T}} h^k) \right\|_{\mathcal{F}}^2, \tag{5}$$

where $X_p^k (p = 1, 2, ..., P)$ denotes vectorized frame, and p represents the number of context patches that contain some target objects in the background region. Besides, α^k indicates the weight of the p^{th} context sample.

3.4. Temporal Regularization Approach

The BACF tracker is mostly affected by occlusion and object deformation. For instance, when the target object is partially or completely occluded by other objects or obstacles, the tracker may encounter difficulties in accurately estimating the object's position and appearance. Hence, the BACF tracker may struggle to differentiate between the target object and occluding objects, leading to tracking failures or drift. Moreover, the temporal regularization approach assists in handling occlusions during object tracking. It allows the

tracker to maintain the target's estimated position even when it is temporarily occluded by other objects or obstacles. The temporal information aids in predicting the probable state of the target object following occlusion and maintains consistent tracking performance. Specifically, the temporal regularization technique takes into account the relationship between the present and prior frames of the target region to boost the performance of the tracker. Hence, the formulation of our objective function is described as follows:

$$\begin{aligned} E(h,r) =& \frac{1}{2}\sum_{k=1}^{C}\left\| r^k - \sum_{k=1}^{C} X^k \circledast (B^{\mathsf{T}} h^k)\right\|_{\mathcal{F}}^2 + \frac{\lambda_1}{2}\sum_{i=1}^{M}\left\|h_{i:}^k\right\|_{\mathcal{F}} + \frac{\lambda_2}{2}\sum_{j=1}^{N}\left\|h_{:j}^k\right\|_{\mathcal{F}} \\ &+ \frac{1}{2}\sum_{p=1}^{P}\left\|\sum_{k=1}^{C}\alpha^k X_p^k \circledast (B^{\mathsf{T}} h^k)\right\|_{\mathcal{F}}^2 + \frac{\lambda_3}{2}\sum_{k=1}^{C}\left\|h_t^k - h_{t-1}^k\right\|_{\mathcal{F}}^2, \end{aligned} \tag{6}$$

where λ_3 indicates the penalty factor, h_t^k denotes the current frame information, and h_{t-1}^k represents the previous frame information of the target region.

3.5. Mutation-Aware Technique

In this division, we present a brief analysis of the mutation-aware technique. When the target appearance is affected by mutations during the tracking, the tracker deviates from the target region to sub-peaks in the response map, which impacts the tracking performance of the filter. To address this issue, we presented the Mutation Threat Mechanism (MTM) technique that effectively handles the sub-peaks in the response map. Following that, we employed the distance matrix (Π) to evaluate the MTM technique, which is used in the adaptive hybrid model strategy. Moreover, the fundamental principle of mutation-aware is to identify the location of the most critical sub-peak in the response map, illustrating the intensity of the current mutation. The main workflow of the MTM technique is illustrated in Fig.3. Finally, the computation of sub-peaks in the MTM is determined as follows:

$$M^k = \frac{R^k \odot \Psi^k}{R_{max}^k} \odot \Pi, \tag{7}$$

where k represents the $k-th$ frame of the image, R denotes the search region response map, and R_{map}^k indicates the response map peak value (R). Furthermore, Ψ indicates the binary matrix of the response map's sub-peaks. M^k denotes the MTM of the $k-th$ frame. Moreover, Π indicates the distance matrix, which is defined as follows:

$$\Pi = \begin{cases} \frac{\vartheta}{1+\delta.exp(d_{(i,j)})} & , \ d_{(i,j)} > d_{min} \\ 0 & , \ d_{(i,j)} \leq d_{min}, \end{cases} \tag{8}$$

where ϑ represents the weights on the distance matrix, δ denotes the influence degree of the distance. The distance to the center of the matrix along the $i-th$ row and $j-th$ column is denoted by $d_{(i,j)}$. Also, the

center region weights are set to 0 with a threshold d_{min}. Moreover, the MTM is employed in the optimal adaptive hybrid model.

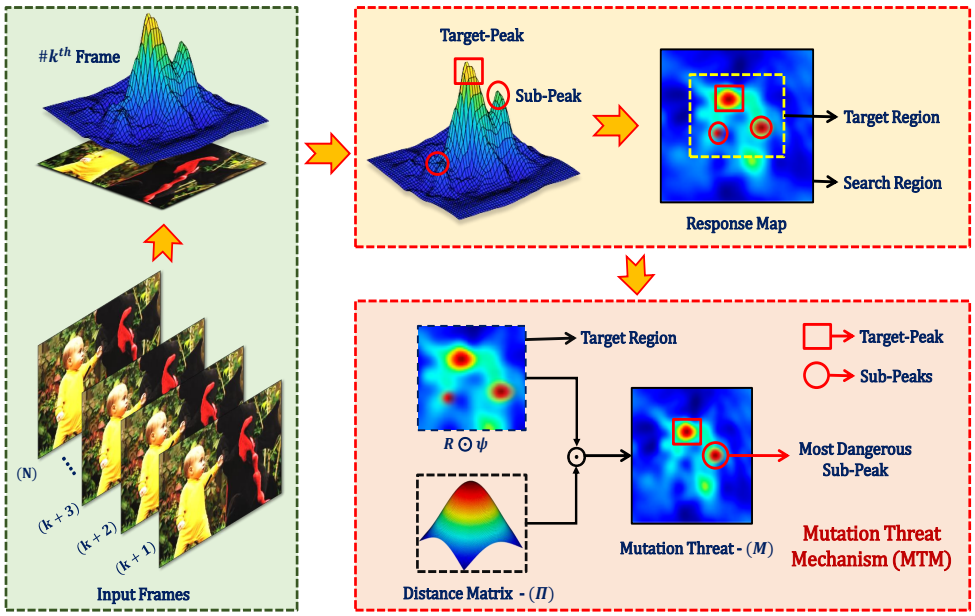


Figure 3: The comprehensive diagram of mutation threat mechanism techniques aims to pinpoint the most perilous sub-peaks, illustrating the intensity of the current mutation within the response map.

3.6. Optimal Adaptive Hybrid Model

The optimal adaptive hybrid model utilizes the MTM technique that helps to tackle the mutation problem during tracking. Thereby, the tracker effectively tracks the target location without distractions. Hence, the formulation of our objective function is calculated as follows:

$$\begin{aligned} E(h,r) =& \frac{1}{2}\sum_{k=1}^{C}\left\| r^k - \sum_{k=1}^{C} X^k \circledast (B^{\mathsf{T}} h^k)\right\|_{\mathcal{F}}^2 + \frac{\lambda_1}{2}\sum_{i=1}^{M}\left\|h_{i:}^k\right\|_{\mathcal{F}} + \frac{\lambda_2}{2}\sum_{j=1}^{N}\left\|h_{:j}^k\right\|_{\mathcal{F}} \\ &+ \frac{1}{2}\sum_{p=1}^{P}\left\|\sum_{k=1}^{C}\alpha^k X_p^k \circledast (B^{\mathsf{T}} h^k)\right\|_{\mathcal{F}}^2 + \frac{\lambda_3}{2}\sum_{k=1}^{C}\left\|h_t^k - h_{t-1}^k\right\|_{\mathcal{F}}^2 + \frac{\lambda_4}{2}\sum_{k=1}^{C}\left\|\Omega_t^k - r_t^k\right\|_{\mathcal{F}}^2, \end{aligned} \tag{9}$$

where Ω_t^k indicates the optimal adaptive hybrid model, r_t^k denotes the adaptive hybrid model, and λ_4 is a regularization parameter. Moreover, the optimal adaptive hybrid model Ω_t^k is described in Remark 1.

3.7. Adaptive Hybrid Model with Temporal Regularization

To prevent the tracker deviates from the target region due to mutations, we present the adaptive hybrid model with a temporal technique. More specifically, the adaptive hybrid model utilizes the previous frame

information to redetect the target region after mutations that effectively increases the tracker performance. Therefore, the finalist objective function is expressed in the following way:

$$E(h,r) = \frac{1}{2}\sum_{k=1}^{C}\left\|r^k - \sum_{k=1}^{C} X^k \circledast (B^{\mathsf{T}}h^k)\right\|_{\mathcal{F}}^2 + \frac{\lambda_1}{2}\sum_{i=1}^{M}\left\|h_{i:}^k\right\|_{\mathcal{F}} + \frac{\lambda_2}{2}\sum_{j=1}^{N}\left\|h_{:j}^k\right\|_{\mathcal{F}} + \frac{1}{2}\sum_{p=1}^{P}\left\|\sum_{k=1}^{C}\alpha^k X_p^k \circledast (B^{\mathsf{T}}h^k)\right\|_{\mathcal{F}}^2 + \frac{\lambda_3}{2}\sum_{k=1}^{C}\left\|h_t^k - h_{t-1}^k\right\|_{\mathcal{F}}^2 + \frac{\lambda_4}{2}\sum_{k=1}^{C}\left\|\Omega_t^k - r_t^k\right\|_{\mathcal{F}}^2 + \frac{\Phi}{2}\sum_{k=1}^{C}\left\|r_t^k - r_{t-1}^k\right\|_{\mathcal{F}}^2. \tag{10}$$

where Φ is a penalty factor, r_t^k represents the current hybrid model information, and r_{t-1}^k denotes the previous hybrid model information of the target.

3.8. Frequencey Domain Transformation

Fourier transform enables efficient implementation of correlation filters through element-wise multiplication. Moreover, the Eq. (10) is transformed into the frequency domain to optimize computational efficiency. Hence, the objective function is formulated as follows:

$$\begin{aligned} E(h,\hat{r},\hat{g}) =& \frac{1}{2}\sum_{k=1}^{C}\left\|\hat{r}^k - \sum_{k=1}^{C}\hat{X}^k \circledast \hat{g}^k\right\|_{\mathcal{F}}^2 + \frac{\lambda_1}{2}\sum_{i=1}^{M}\left\|h_{i:}^k\right\|_{\mathcal{F}} + \frac{\lambda_2}{2}\sum_{j=1}^{N}\left\|h_{:j}^k\right\|_{\mathcal{F}} + \frac{1}{2}\sum_{p=1}^{P}\left\|\sum_{k=1}^{C}\alpha^k \hat{X}_p^k \circledast \hat{g}^k\right\|_{\mathcal{F}}^2 \\ &+ \frac{\lambda_3}{2}\sum_{k=1}^{C}\left\|\hat{g}_t^k - \hat{g}_{t-1}^k\right\|_{\mathcal{F}}^2 + \frac{\lambda_4}{2}\sum_{k=1}^{C}\left\|\hat{\Omega}_t^k - \hat{r}_t^k\right\|_{\mathcal{F}}^2 + \frac{\Phi}{2}\sum_{k=1}^{C}\left\|\hat{r}_t^k - \hat{r}_{t-1}^k\right\|_{\mathcal{F}}^2, \\ s.t.\ \hat{g}^k =& \sqrt{T}(FB^{\mathsf{T}} \otimes I^k)h^k, \end{aligned} \tag{11}$$

where $\hat{g}^k$ represents the auxiliary variable, $\wedge$ indicates the discrete Fourier transform, and $\otimes$ denotes the Kronecker product. Besides, F indicates the $B * B$ Fourier matrix and I^k denotes the identity matrix.

3.9. Augmented Lagrangian Method

We employ the ADMM technique to improve the computational speed, which helps to guide the optimization process toward better convergence by gradually adjusting the penalty parameters. To solve Eq. (11), we utilize the ALM technique in the frequency domain.

$$\begin{aligned} \mathcal{L}(h,\hat{r},\hat{g},\hat{\Gamma}) =& \frac{1}{2}\sum_{k=1}^{C}\left\|\hat{r}^k - \sum_{k=1}^{C}\hat{X}^k \circledast \hat{g}^k\right\|_{\mathcal{F}}^2 + \frac{\lambda_1}{2}\sum_{i=1}^{M}\left\|h_{i:}^k\right\|_{\mathcal{F}} + \frac{\lambda_2}{2}\sum_{j=1}^{N}\left\|h_{:j}^k\right\|_{\mathcal{F}} + \frac{1}{2}\sum_{p=1}^{P}\left\|\sum_{k=1}^{C}\alpha^k \hat{X}_p^k \circledast \hat{g}^k\right\|_{\mathcal{F}}^2 \\ &+ \frac{\lambda_3}{2}\sum_{k=1}^{C}\left\|\hat{g}_t^k - \hat{g}_{t-1}^k\right\|_{\mathcal{F}}^2 + \frac{\lambda_4}{2}\sum_{k=1}^{C}\left\|\hat{\Omega}_t^k - \hat{r}_t^k\right\|_{\mathcal{F}}^2 + \frac{\Phi}{2}\sum_{k=1}^{C}\left\|\hat{r}_t^k - \hat{r}_{t-1}^k\right\|_{\mathcal{F}}^2 + \frac{\mu}{2}\sum_{k=1}^{C}\left\|\hat{g}^k - h^k + \frac{\hat{\Gamma}^k}{\mu}\right\|_{\mathcal{F}}^2, \end{aligned} \tag{12}$$

where $\hat{\Gamma}$ denotes the Lagrangian term and μ is the regularization parameter. The following sub-problems are optimized using the ADMM approach.

3.10. Sub-Problem $\hat{g}^*_{t+1}$

To solve the sub-problem $\hat{g}^*_{t+1}$ and obtain better convergence, we utilize the ALM technique. Moreover, the sub-problem $\hat{g}^*_{t+1}$ can be solved as follows:

$$\begin{aligned}\hat{g}^*_{t+1} =& \frac{1}{2}\sum_{k=1}^{C}\left\|\hat{r}^k - \sum_{k=1}^{C}\hat{X}^k \circledast \hat{g}^k\right\|^2_{\mathcal{F}} + \frac{1}{2}\sum_{p=1}^{P}\left\|\sum_{k=1}^{C}\alpha^k\hat{X}^k_p \circledast \hat{g}^k\right\|^2_{\mathcal{F}} \\ &+ \frac{\lambda_3}{2}\sum_{k=1}^{C}\left\|\hat{g}^k_t - \hat{g}^k_{t-1}\right\|^2_{\mathcal{F}} + \frac{\mu}{2}\sum_{k=1}^{C}\left\|\hat{g}^k - h^k + \frac{\hat{\Gamma}^k}{\mu}\right\|^2_{\mathcal{F}}.\end{aligned} \tag{13}$$

After solving Eq. (13), the closed-form solution can be obtained as follows:

$$\hat{g}^*_{t+1} = \frac{\hat{X}^\intercal\hat{r} + \lambda_3\hat{g}_{t-1} + \mu h_t - \hat{\Gamma}}{\hat{X}^\intercal\hat{X} + \alpha\hat{X} + \lambda_3 + \mu}. \tag{14}$$

3.11. Sub-Problem h^*_{t+1}

The sub-problem h^*_{t+1} is formulated as follows:

$$h^*_{t+1} = \frac{\lambda_1}{2}\sum_{i=1}^{M}\|h_{i:}\|_{\mathcal{F}} + \frac{\lambda_2}{2}\sum_{j=1}^{N}\|h_{:j}\|_{\mathcal{F}} + \frac{\mu}{2}\sum_{k=1}^{K}\left\|g^k - h^k + \frac{\Gamma^k}{\mu}\right\|^2_{\mathcal{F}}. \tag{15}$$

To facilitate the sub-problem h, Eq. 15 can be segregated into two sub-problems as follows.

3.11.1. Sub-Problem h_i

The solution of the sub-problem h_i is solved as follows:

$$h_i = \frac{\lambda_1}{2}\sum_{i=1}^{M}\left\|h^k_{i:}\right\|_{\mathcal{F}} + \frac{\mu}{2}\sum_{i=1}^{M}\left\|g^k_{i:} - h^k_{i:} + \frac{\Gamma^k_{i:}}{\mu}\right\|^2_{\mathcal{F}}. \tag{16}$$

After solving Eq. (16), the closed-form solution of sub-problem h_i can be calculated as follows:

$$h_i = max\left(0, 1 - \frac{\lambda_1}{\mu\left\|g_{i:} + \frac{\Gamma_{i:}}{\mu}\right\|_{\mathcal{F}}}\right)\left(g_{i:} + \frac{\Gamma_{i:}}{\mu}\right). \tag{17}$$

3.11.2. Sub-Problem h_j

The solution of the sub-problem h_j is solved as follows:

$$h_j = \frac{\lambda_2}{2}\sum_{j=1}^{N}\left\|h^k_{:j}\right\|_{\mathcal{F}} + \frac{\mu}{2}\sum_{j=1}^{N}\left\|g^k_{:j} - h^k_{:j} + \frac{\Gamma^k_{:j}}{\mu}\right\|^2_{\mathcal{F}}. \tag{18}$$

After solving Eq. (18), the closed-form solution of sub-problem h_j can be derived as follows:

$$h_j = max\left(0, 1 - \frac{\lambda_2}{\mu \left\| g_{:j} + \frac{\Gamma_{:j}}{\mu} \right\|_{\mathcal{F}}}\right)\left(g_{:j} + \frac{\Gamma_{:j}}{\mu}\right). \tag{19}$$

Eq. (17) denotes the filter weight in the row direction and Eq. (19) represents the filter weight in the column direction. Following that, the Eq. (17) and Eq. (19) are combined as follows:

$$h_{ij} = max\left(0, 1 - \frac{\lambda_1}{\mu \left\| g_{i:} + \frac{\Gamma_{i:}}{\mu} \right\|_{\mathcal{F}}} - \frac{\lambda_2}{\mu \left\| g_{:j} + \frac{\Gamma_{:j}}{\mu} \right\|_{\mathcal{F}}}\right)\mathcal{Q}. \tag{20}$$

where $\mathcal{Q} = \left(g_{ij} + \frac{\Gamma_{ij}}{\mu}\right)$. Besides, h_{ij} denotes the $i-th$ row and $j-th$ column of the filter in the spatial domain.

3.12. Sub-Problem $\hat{r}^*_{t+1}$

The solution of the sub-problem $\hat{r}^*_{t+1}$ is solved as follows:

$$\hat{r}^*_{t+1} = \frac{1}{2}\sum_{k=1}^{C}\left\|\hat{r}^k - \sum_{k=1}^{C}\hat{X}^k \circledast \hat{g}^k\right\|^2_{\mathcal{F}} + \frac{(1+\Psi^2)\lambda_4}{2}\sum_{k=1}^{C}\left\|\hat{\Omega}^k_t - \hat{r}^k_t\right\|^2_{\mathcal{F}} + \frac{(1+\Psi^2)\Phi}{2}\sum_{k=1}^{C}\left\|\hat{r}^k_t - \hat{r}^k_{t-1}\right\|^2_{\mathcal{F}}. \tag{21}$$

After solving Eq. (21), we obtain the closed-form solution for the sub-problem $\hat{r}^*_{t+1}$, which can be expressed as follows:

$$\hat{r}^*_{t+1} = \frac{\hat{X}\hat{g} + (1+\Psi^2)\lambda_4\hat{\Omega} + (1-\Psi^2)\Phi\hat{r}_{t-1}}{1 + (1+\Psi^2)\lambda_4 + (1-\Psi^2)\Phi}, \tag{22}$$

where Ψ represents the penalty factor.

3.13. Lagrangian Multiplier Update

The Lagrangian multiplier is updated as follows:

$$\begin{aligned} \hat{\Gamma}_{t+1} &\leftarrow \hat{\Gamma}_t + \mu(\hat{g}^*_{t+1} - h^*_{t+1}), \\ \mu_{t+1} &= min(\mu_{max}, \beta\mu_t), \end{aligned} \tag{23}$$

where $\hat{\Gamma}$ represents the Lagrangian parameter, β denotes the scale factor, and μ_{max} is the predefined maximum value of μ. Besides, the t and $t+1$ indicate the iteration of the filter.

3.14. Model Update

We mainly focus on the online model update technique, which is an essential component of object tracking. Moreover, the online model update mechanism helps the tracker to handle challenging scenarios where occlusions frequently occur, enhancing its tracking performance and robustness. Therefore, the online model update technique is defined as follows:

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

where $\hat{X}_t^{model}$ represents the newly updated model and $\hat{X}_{t-1}^{model}$ denotes the previous model of the input frame. Besides, γ denotes the learning rate of the target model, and $\hat{X}_t$ indicates the weight of the current filter. Finally, t and $t-1$ denote the current frame and the learned model of the previous frame, respectively.

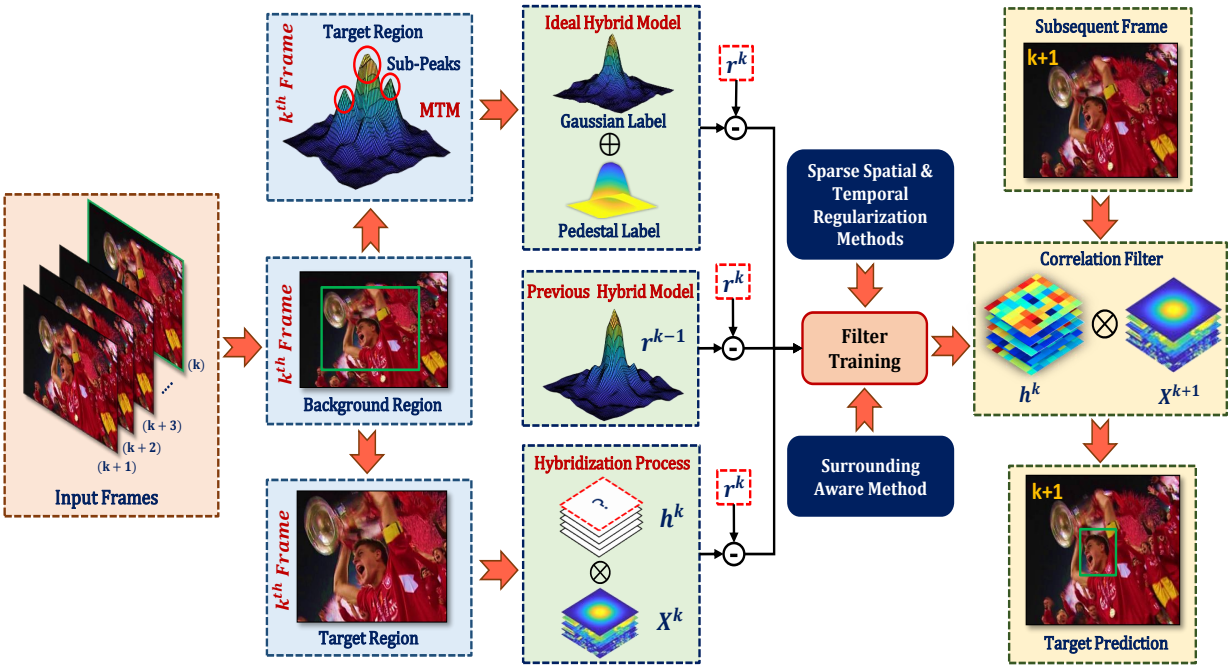


Figure 4: The schematic diagram of the proposed method. The dotted red boxes indicate the variables that need to be resolved in the primary regression analysis.

Remark 1: The presented adaptive hybrid model r^k consists of three key components: The ideal hybrid model, the previous hybrid model, and the hybridization process. The detailed architecture is illustrated in Fig. 4.

Ideal Hybrid Model: The ideal hybrid model contains two sections: Gaussian label, and pedestal label, which are described in the following manner:

Gaussian Label: The Gaussian label trains the filter to generate a peak response at the target's location. This peak response is used to locate the target in subsequent frames. Notably, the Gaussian label can be more robust to changes in the appearance, scale, and orientation of the target object.

Pedestal Label: When the target is affected by the mutation, the filter deviates from the target location. To address this problem, the ideal hybrid model employs the pedestal label, which adaptively adjusts the response value to redetect the target region. Specifically, the pedestal label is initialized based on the target scale and predefined value. Moreover, the optimal adaptive hybrid model Ω is determined as follows:

$$\Omega = y1 + (1 - \Theta.max(M^k))y2. \tag{25}$$

where $y1$ denotes the Gaussian label and $y2$ represents the pedestal label. Also, Θ represents the predefined value, and $max(M^k)$ indicates the maximum value of MTM that is discussed in the previous section 3.5.

Previous Hybrid Model: To ensure the stability of the adaptive hybrid model and avoid the excessive impact of the specific mutations, we updated the adaptive hybrid model by utilizing the previous frame information. In our objective function, we also consider the consistency of the adaptive hybrid model, which is illustrated in Eq. 3.7.

Hybridization Process: Unlike the conventional methods, the adaptive hybrid model r and the correlation filter h are trained alternately in the presented method. In particular, the interaction between the filter and adaptive hybrid model during the training of the k-th frame leads to a rapid convergence process.

Remark 2: In contrast to traditional sparse spatial methods [22, 23], we propose the $\ell_{2,1}$ norm-based sparse spatial feature selection approach, which effectively leverages the row and column-based feature selection techniques to obtain the prominent features in the target region. Specifically, the row and column-wise sparse methods are computationally more efficient. By taking advantage of the sparsity in both rows and columns independently, the sparse spatial technique minimizes computational complexity.

4. Experimental Results and Analysis

To estimate the performance and robustness of the proposed MRMACF tracker, we perform extensive experiments on 8 various challenging benchmark datasets such as OTB-2013 [24], OTB-2015 [24], TempleColor-128 [25], UAV-123 [26], UAVDT [27], VOT-2018 [28], LaSOT [29], and GOT-10K [30]. We first analyze the experimental setup in subsection 4.1. Following that, we conduct the parameter analysis and ablation study with various components of our proposed technique in subsection 4.2. Finally, we examine the comparative analysis of our proposed approach with other state-of-the-art-trackers in subsection 4.3.

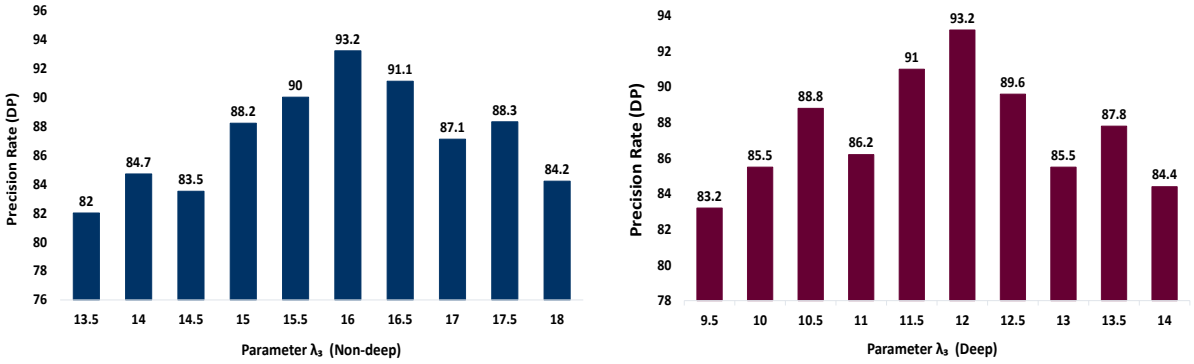
4.1. Experimental Setup

In this section, we discuss our experimental setup as well as feature fusion details. The presented technique is performed in MATLAB and the experiments are implemented on a PC with an Intel(R) Core(TM) i5-12400 CPU at 2.50GHz, 16 GB RAM, and NVIDIA GeForce RTX 3060 GPU. Specifically, the proposed technique utilizes both hand-crafted and deep features, including HOG, Intensity, ColorName, and ResNet. These features are exploited to extract robust information from the image sequences.

4.2. Parameter and Ablation Analysis

4.2.1. Parameter Analysis

We conduct the parameter analysis for our proposed approach on the OTB-2015 dataset. In this regard, we examine the λ_3 parameter Eq. (13) and set the value of λ_3 = [16 12] for hand-crafted and deep convolutional features. Also, the experimental results are illustrated in Fig. 5. Moreover, we examine the regularization parameters $\lambda_1, \lambda_2, \lambda_3$, and λ_4 in Eqs. (16), (18), (13), and (21).



(a) (b)

Figure 5: Parameter Analysis

The parameters λ_1, λ_2 and λ_4 are fixed value, such that $\lambda_1 = 10, \lambda_2 = 1$, and $\lambda_4 = 1.5$, respectively. When the λ_3 (Non-deep) parameter varies from 13.5 to 18, the precision score improves continuously and obtains the maximum value at $\lambda_3 = 16$. However, when the value exceeds 16, the precision score gradually decreased as shown in Fig. 5*a*. Likewise, when the λ_3 (Deep) parameter varies from 9.5 to 14, the precision score increases gradually and achieves the maximum value at $\lambda_3 = 12$. Nevertheless, when the value exceeds 12, the precision score is gradually reduced as shown in Fig. 5*b*. Specifically, the λ_3 is a key parameter to improve the tracking efficiency.

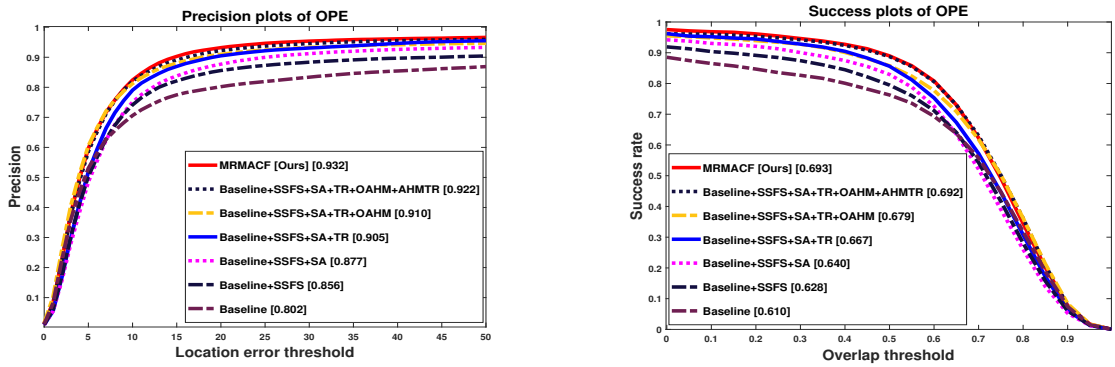
Table 1: The parameters of the proposed tracker are listed below.

Parameters	Values	Parameters	Values
λ_1	10	λ_2	1
λ_3	[16 12]	λ_4	1.5
α	0.05	ϑ	1
ϕ	1	δ	0.01
μ	1	μ_{max}	0.1
γ	[0.6 0.05]	β	1.5

In (13), the surrounding-aware approach uses the α parameter and we set the value of $\alpha = 0.05$. Besides, the Φ denotes the penalty factor of the adaptive hybrid model in Eq. (21), which is a fixed value $\Phi = 1$. Also, Ψ represents the penalty factor, we fixed the value of $\Psi = 0.8$. Moreover, we set the learning rate γ parameters 0.6 and 0.05 for hand-crafted and deep convolutional features. In the end, the parameter execution of our proposed approach is listed in Table 1.

4.2.2. Ablation Analysis

We conduct an ablation analysis for our proposed approach on the OTB-2015 dataset. We first compared our proposed approach with the baseline tracker and different components of the proposed technique such as Sparse Spatial Feature Selection (SSFS), Surrounding-Aware (SA), Temporal Regularization (TR), Optimal Adaptive Hybrid Model (OAHM), and Adaptive Hybrid Model with Temporal Regularization (AHMTR). Also, the ablation study results are exhibited in Fig. 6. As shown in Fig. 6, we confirm that the baseline tracker achieved the Distance Precision (DP) score (80.2%) and Area Under Curve (AUC) score (61.0%).



(a) (b)

Figure 6: The precision and success plots of the proposed method on the OTB-2015 dataset are illustrated in Fig. 6a and Fig. 6b.

Further, we can see that by adding the SSAS approach to our baseline tracker ($Baseline + SSAS$), the tracking efficiency significantly improved in DP / AUC scores of (5.4% / 1.8%). Additionally, we observed that the DP / AUC scores of (7.5% / 3.0%) greatly increased after adding the SA approach with our baseline tracker ($Baseline + SSAS + SR$). Also, by integrating the TR approach with our baseline

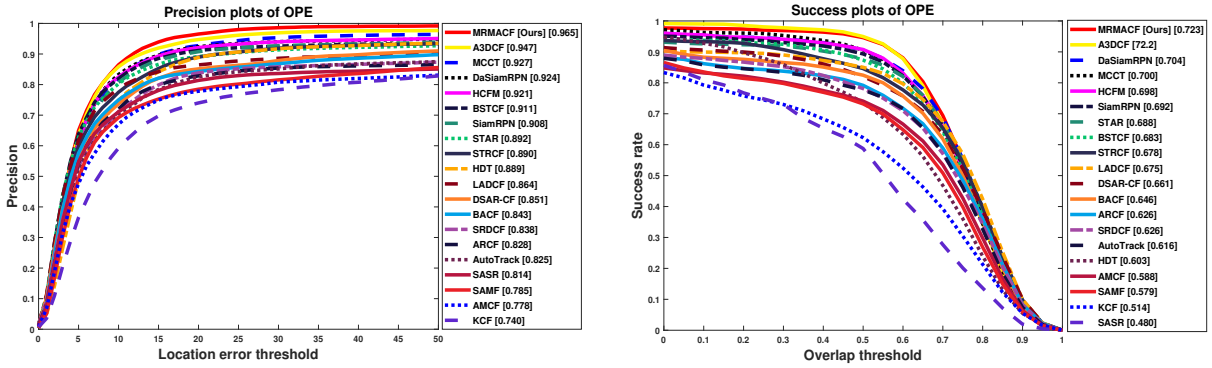
tracker ($Baseline + SSAS + SR + TR$), we confirm that the tracking performance enhanced in DP / AUC scores of (10.3% / 5.7%). In particular, we confirm that by adding the OAHM technique to our baseline tracker ($Baseline + SSAS + SR + TR + OAHM$), the tracking performance considerably increased in DP / AUC scores of (10.8% / 6.9%). Moreover, by incorporating the AHMTR strategy with our baseline tracker ($Baseline + SSAS + SR + TR + OAHM + AHMTR$), we observed that the tracking efficiency is improved in DP / AUC scores of (12% / 8.2%), respectively. Specifically, we have proven that the proposed approach achieved great performance in DP / AUC scores of (93.2% / 69.3%) by incorporating all components of the proposed approach.

4.3. Experimental Evaluation

We conduct the evaluation process of our proposed technique with 30 various trackers such as BACF [5], SRDCF [6], BSTCF [7], STRCF [8], SASR [9], SiamRPN [10], DaSiamRPN [11], LADCF [14], A3DCF [15], STAR [16], MEVT [17], HCFM [18], MCCT [31], ARCF [32], DSAR-CF [33], AMCF [34], SAMF [35], KCF [36], AutoTrack [37], DSST [38], HDT [39], CSK [40], CSRDCF [41], SiamFC [42], MDNet [43], RACF [44], FDSiamFC [45], ATOM [46], AFSN [47], and GOTURN [48].

4.3.1. Evaluation on OTB-2013 Dataset

We assess the performance of the proposed MRMACF method against 19 other state-of-the-art trackers using the OTB-2013 dataset. The comparison results and speed of the conventional and proposed tracker's precision and success scores are exhibited in Fig. 7 and Table 2. As shown in Fig. 7*a* and Fig. 7*b*, we can see that our presented method obtains the better DP and AUC scores of (96.5%) and (72.3%). Also, compared with the baseline method our proposed approach demonstrates superior performance in DP / AUC scores of (12.2% / 7.7%).



(a) (b)

Figure 7: The distance precision and overlap success results of the OTB-2013 dataset are exhibited in Fig. 7a and Fig. 7b, respectively.

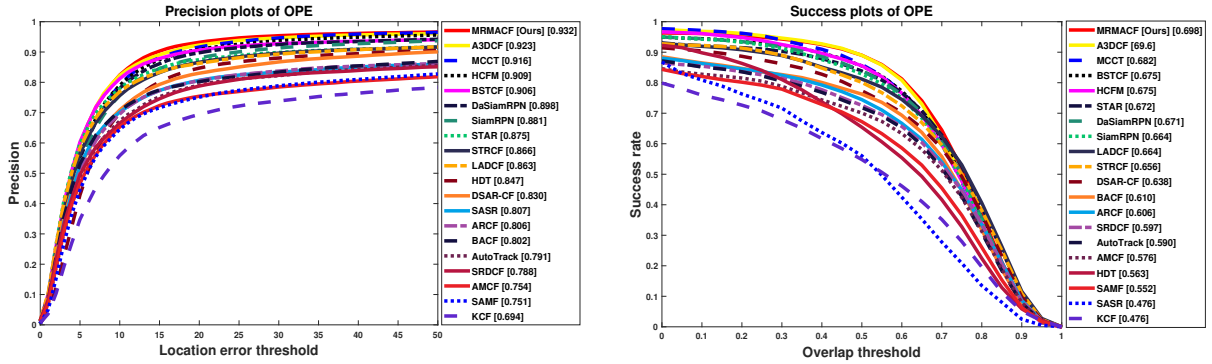
Table 2: Comparative results of conventional trackers with the proposed tracker on the OTB-2013 dataset with 50 sequences. ⋆ denotes the hand-crafted and deep features-based trackers. The results of the top 15 trackers are presented as follows:

Methods	Metrics	Ours⋆	A3DCF⋆	MCCT⋆	DaSiamRPN⋆	HCFM⋆	BSTCF⋆	SiamRPN⋆	STAR	STRCF	HDT⋆	LADCF	DSAR-CF	BACF	SRDCF	ARCF
OTB-2013	DP	96.5	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	72.3	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		10.8	-	7.8	160.0	2.7	19.0	160.0	5.5	5.8	10.0	1.5	16.0	25.4	-	26.2

Moreover, when compared to the hand-crafted feature-based methods such as STRCF [8], LADCF [14], DSAR-CF [33], and SRDCF [6], the proposed approach obtains better improvement in the DP and AUC scores of (7.5% / 4.5%), (10.1% / 4.8%), (11.4% / 6.2%), and (12.7% / 9.7%), respectively. Besides, compared to deep features-based approaches such as A3DCF [15], MCCT [31], DaSiamRPN [11], and SiamRPN [10], we demonstrate that the proposed approach achieves significant gain in DP and AUC scores of (1.8% / 0.1%), (3.8% / 2.3%), (4.1% / 1.9%), and (5.7% / 3.1%), respectively. Finally, we show that the presented method attains outstanding efficiency when compared to the hand-crafted and deep feature-based approaches.

4.3.2. Evaluation on OTB-2015 Dataset

We evaluate the tracking performance of our presented MRMACF method by conducting experimental evaluations on the OTB-2015 datasets. The comparison of tracking speed between the proposed method and conventional approaches is demonstrated in Table 3. Also, the experimental results are exhibited in Fig. 8a and Fig. 8b. From these results, we ensure that our presented technique obtains outstanding tracking efficiency in terms of precision and success scores of (93.2% / 69.8%), which is better than other conventional trackers. Also, the performance of the proposed approach is compared with the best top 15 trackers with 11 Attributes, as exhibited in Table 4.



(a) (b)

Figure 8: The distance precision and overlap success results of the OTB-2015 dataset are exhibited in Fig. 8a and Fig. 8b, respectively.

Table 3: Comparative results of conventional trackers with the proposed tracker on the OTB-2015 dataset with 100 sequences. ⋆ denotes the hand-crafted and deep features-based trackers. The results of the top 15 trackers are presented as follows:

Methods	Metrics	Ours⋆	A3DCF⋆	MCCT⋆	HCFM⋆	BSTCF⋆	DaSiamRPN⋆	SiamRPN⋆	STAR	STRCF	LADCF	HDT⋆	DSAR-CF	SASR⋆	ARCF	BACF
OTB-2015	DP	93.2	92.3	91.6	90.9	90.6	89.8	88.1	87.5	86.6	86.3	84.7	83.0	80.7	80.6	80.2
	AUC	69.8	69.6	68.2	67.5	67.5	67.1	66.4	67.2	65.6	66.4	56.3	63.8	47.6	60.6	61.0
FPS		10.2	4.2	7.8	2.3	15.6	160.0	160.0	2.5	24.3	0.9	2.7	6.1	-	10.2	26.7

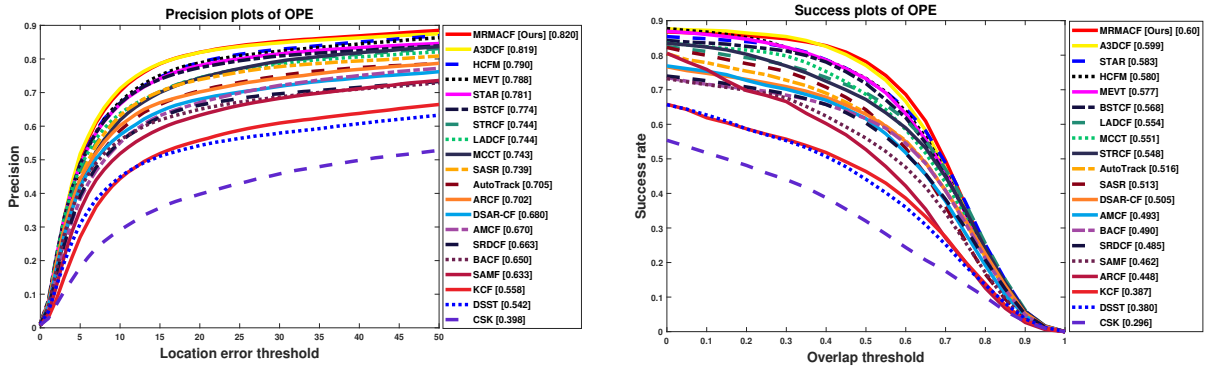
Table 4: The proposed method undergoes attribute-based evaluation on the OTB-2015 dataset, where the most favorable results are indicated in red, blue, and green fonts.

Method	MRMACF	A3DCF	MCCT	HCFM	BSTCF	DaSiamRPM	SiamRPN	STAR	STRCF	LADCF	HDT	DSAR-CF	SASR	ARCF	BACF
IV	**95.3/72.3**	**93.8/72.1**	**90.7/69.3**	89.1/69.0	88.4/67.9	85.8/66.0	86.8/67.0	86.0/68.2	83.8/65.2	80.5/64.6	81.7/53.2	81.3/64.8	79.1/48.6	76.0/59.6	80.9/63.0
OPR	**93.1/68.5**	**92.6/68.4**	**91.0/66.4**	89.9/65.6	89.8/65.5	90.4/66.3	86.1/64.1	87.3/65.3	85.5/62.7	83.4/63.0	80.6/52.9	81.0/61.1	79.3/43.7	77.1/55.7	77.1/57.7
SV	**91.6/67.8**	**91.3/67.5**	**88.6**/64.6	87.6/63.6	88.5/**65.2**	85.6/62.5	84.5/62.4	84.0/63.9	84.3/63.6	83.5/63.8	80.6/48.6	82.4/62.7	77.5/42.5	76.8/56.0	74.3/55.9
OCC	**89.0/67.4**	**88.7/66.7**	87.2/64.9	86.4/64.0	**88.3/66.6**	84.4/63.4	80.6/61.0	83.6/64.5	82.6/62.3	82.4/63.9	76.6/51.7	77.1/60.3	73.2/38.7	73.1/55.1	69.7/54.8
DEF	**89.9/67.0**	**89.8/66.1**	**89.6/64.3**	89.5/63.8	87.5/63.7	86.7/62.6	87.1/63.9	83.8/62.8	84.1/60.5	81.3/59.7	81.9/54.0	78.9/59.0	76.7/44.1	76.5/58.0	76.3/57.7
MB	**92.8/70.8**	**91.5/70.4**	**89.0/68.9**	88.0/68.0	86.1/67.2	85.5/66.2	83.5/65.4	82.3/66.8	84.1/67.0	81.4/65.9	78.9/56.0	81.3/64.4	76.3/45.6	76.8/61.5	73.1/57.6
FM	**92.1/69.1**	**89.9/68.2**	**88.0/65.6**	87.5/64.9	84.1/64.6	85.9/64.5	85.4/64.5	81.3/65.1	79.5/62.9	77.8/62.0	79.8/54.7	78.4/61.4	77.9/46.8	75.9/58.8	75.8/58.7
IPR	**94.7/68.0**	**94.3/67.9**	**92.7/65.9**	91.8/65.3	87.2/63.1	90.1/65.0	87.4/62.8	83.8/62.1	80.9/59.8	80.1/59.6	84.0/54.7	77.9/57.8	78.8/44.8	77.8/55.4	75.7/56.3
OV	**90.0/66.3**	84.8/62.3	87.2/**65.5**	**87.8/65.6**	**87.5**/65.4	85.3/63.3	80.5/59.0	77.0/60.0	75.7/57.8	82.8/63.0	68.6/50.0	72.7/57.5	72.2/35.4	69.6/52.9	70.9/54.0
BC	**93.0/70.0**	89.6/67.5	**92.5/69.9**	**91.4/68.7**	88.0/65.7	89.6/68.5	89.1/68.0	85.9/66.0	87.3/64.8	84.4/64.9	84.4/57.8	81.5/63.7	79.1/48.1	76.0/58.9	77.8/60.0
LR	**95.1**/62.0	**94.5/66.1**	**89.6/62.8**	89.4/62.6	87.3/**63.3**	81.9/58.3	82.8/58.0	81.2/60.3	75.6/55.8	75.4/56.5	76.6/42.0	75.8/57.1	81.2/38.5	71.4/49.4	69.0/50.6

Especially, when compared with the baseline method, our presented approach demonstrates notable performance improvement in terms of DP / AUC scores of (13.0% / 8.8%). Moreover, compared to other hand-crafted feature-based approaches like STRCF [8], LADCF [14], DSAR-CF [33], and SRDCF [6], we confirm that our proposed tracker improves the better results in DP / AUC scores of (6.6% / 4.2%), (6.9% / 3.4%), (10.2% / 6.0%), and (14.4% / 10.1%), respectively. Similarly, compared to deep convolutional feature-based methods such as A3DCF [15], MCCT [31], HCFM [18], DaSiamRPN [11], and SiamRPN [10], we ensure that the presented method enhances the superior tracking results in DP / AUC scores of (0.9% / 0.2%), (1.6% / 1.6%), (2.3% / 2.3%), (3.4% / 2.7%) and (5.1% / 3.4%), respectively. Overall, the proposed MRMACF tracker improves the better tracking performance when compared to the state-of-the-art-trackers.

4.3.3. Evaluation on TempleColor-128 Dataset

We conducted experiments on the TC-128 benchmark dataset, comprising 128 sequences. Specifically, we confirm that the proposed approach attained the best tracking outcomes in terms of precision score (82.0%) and success score (60.0%), which is better than other conventional trackers. Also, the precision and success results are exhibited in Fig. 9a and Fig. 9b. Moreover, we observe that the presented approach improves the best tracking result in terms of DP and AUC scores of (17.0% / 11.0%) when compared to the baseline tracker.



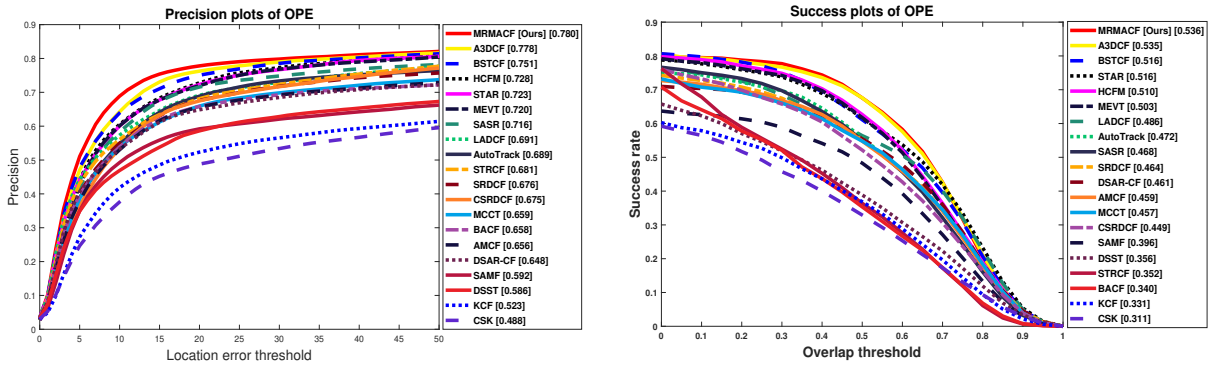
(a) (b)

Figure 9: The distance precision and overlap success results of the TempleColor-128 dataset are exhibited in Fig. 9a and Fig. 9b, respectively.

In particular, compared to the hand-crafted feature-based approaches like STRCF [8], LADCF [14], ARCF [32], DSAR-CF [33], and SRDCF [6], we show that our presented method achieves the large gain in terms of DP and AUC scores of (7.6% / 5.2%), (7.6% / 4.6%), (11.8% / 15.2%), (4.0% / 9.2%), and (15.7% / 11.5%), respectively. Besides, compared to deep feature-based trackers A3DCF [15], HCFM [18], MEVT [17], BSTCF [7], and SASR [9], we can see that our proposed method enhances the gain of (0.1% / 0.1%), (3.0% / 2.0%), (3.2% / 2.3%), (4.6% / 3.2%), and (8.1% / 8.7%) in DP / AUC scores, respectively. From these analyses, we confirm that our method demonstrates significant performance when compared to the hand-crafted and deep feature-based approaches.

4.3.4. Evaluation on UAV-123 Dataset

To estimate the tracker performance, we conduct the experimental evaluation on the UAV123 dataset. Following that, the comparative results of the proposed MRMACF approach are exhibited in Fig. 10. As shown in Fig. 10a and Fig. 10b, we ensure that our presented technique achieves better improvement in DP / AUC scores of (78.0% / 53.6%).



(a) (b)

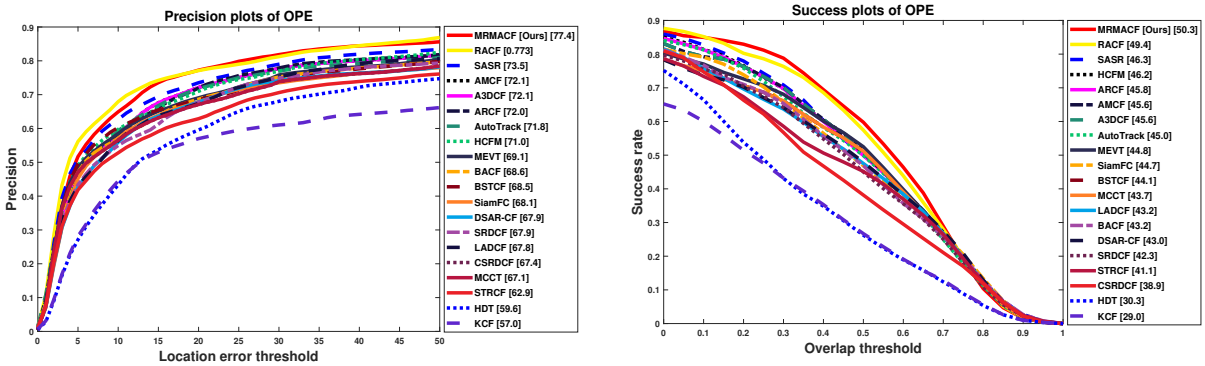
Figure 10: The distance precision and overlap success results of the UAV-123 dataset are exhibited in Fig. 10a and Fig. 10b, respectively.

Specifically, we observe that the presented approach obtains great tracking improvement in DP and AUC scores of (12.2% / 19.6%), compared to the baseline tracker. Further, when compared to hand-crafted feature-based trackers such as LADCF [14], AutoTrack [37], STRCF [8], SRDCF [6], and DSAR-CF [33], we ensure that our presented MRMACF tracker enhances the better results in precision and success scores of (8.9% / 5.0%), (9.1% / 6.4%), (9.9% / 18.4%), (10.4% / 7.2%), and (13.2% / 7.5%), respectively. In particular, when compared to deep convolutional feature-based methods likes A3DCF [15], BSTCF [7], HCFM [18], and SASR [9], we confirm that our presented method improves the remarkable tracking performance in terms of DP and AUC scores of (0.2% / 0.1%), (2.9% / 2.0%), (5.2% / 2.6%), (6.0% / 3.3%), and (6.4% / 6.8%),

respectively. In the end, our presented approach outperformed superior tracking efficiency when compared to other conventional trackers.

4.3.5. Evaluation on UAVDT Dataset

To estimate the tracking performance of the proposed method, we conduct an experimental analysis on the UAVDT dataset. The DP and AUC scores are illustrated in Fig. 11. As shown in Fig. 11a and Fig. 11b, our proposed approach obtains outstanding performance in the DP score (77.4%) and AUC score (50.3%). More specifically, we observe that compared to the baseline method, the presented tracker increases efficiency by achieving (8.8%) in the DP score and (7.1%) in the AUC score. Compared to hand-crafted feature-based methods such as RACF [44], AMCF [34], AutoTrack [37], DSAR-CF [33], and SRDCF [6], we see that, our MRMACF method improves the significant results in terms of DP and AUC scores of (0.1% / 0.9%), (5.3% / 4.7%), (5.6% / 5.3%), (9.5% / 7.3%), and (9.5% / 8.0%), respectively. Furthermore, we verify that the proposed approach enhances tracking efficiency significantly in terms of precision and success scores when compared to deep convolutional feature-based approaches such as SASR [9] (3.9% / 4.0%), A3DCF [15] (5.3% / 4.7%), HCFM [18] (6.4% / 4.1%), MEVT [17] (8.3% / 5.5%), and SiamFC [42] (9.3% / 5.6%), respectively. Overall, our tracking approach outperforms well compared with other conventional trackers.



(a) (b)

Figure 11: The distance precision and overlap success results of the UAVDT dataset are exhibited in Fig. 11a and Fig. 11b, respectively.

4.3.6. Evaluation on VOT-2018 Dataset

To evaluate the efficiency of our proposed approach, we conducted experiments on the VOT-2018 benchmark, comprising 60 challenging video sequences. Following that, we use the Expected Average Overlap (EAO) and Accuracy as our evaluation criteria. Moreover, the evaluation results are exhibited in Table 5. As shown in Table 5, our proposed tracker achieves the best results in the EAO metric of 51.2% and Accuracy of 55.0%, when compared to the other conventional trackers. Specifically, when compared to the baseline method, we observe that our presented tracker enhance the better results in EAO 37.2% and accuracy 4.0%

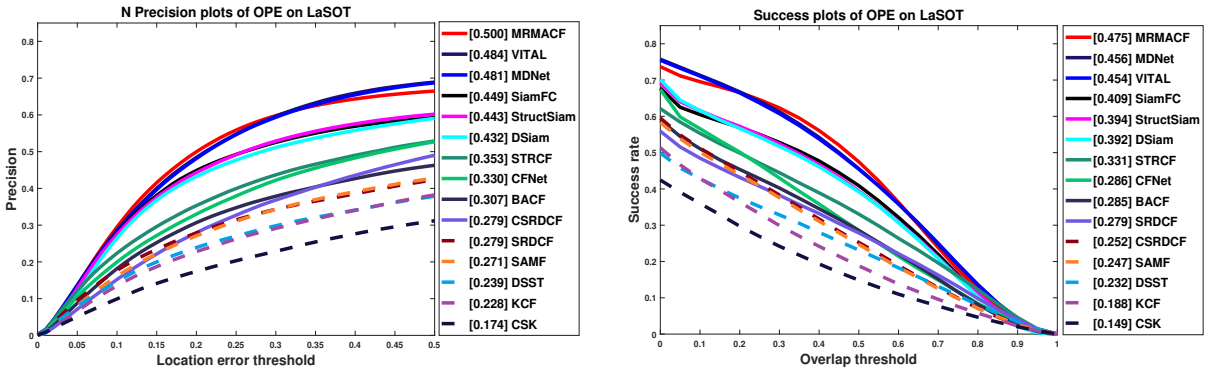
metrics. In particular, when compared to the hand-crafted and convolutional features-based trackers such as SRDCF [6], STRCF [8], SiamFC [42], StructSiam [49], and A3DCF [15], we can see that our presented trackers increase the better results in EAO metric (49.3%, 16.7%, 32.4%, 24.8%, and 10.6%) and Accuracy metric (6.0%, 2.7%, 4.7%, 1.4% and 0.2%), respectively. From these hand-crafted and deep analyses, we observed that the proposed method obtains superior efficiency in terms of EAO and Accuracy metrics.

Table 5: Performance comparison results of the state-of-the-art-trackers on the VOT-2018 dataset.

Tracker	OURS	SAMF	SRDCF	KCF	BACF	SiamFC	CSRDCF	StructSiam	MCCT	DaSiamRPN	STRCF	SiamRPN	LADCF	ATOM	A3DCF
EAO	0.512	0.093	0.119	0.135	0.140	0.188	0.256	0.264	0.274	0.326	0.345	0.383	0.389	0.400	0.406
Accuracy	0.550	0.484	0.490	0.447	0.510	0.503	0.491	0.536	0.530	0.566	0.523	0.586	0.503	0.590	0.548

4.3.7. Evaluation on LaSOT Dataset

We perform experiments on the LaSOT dataset to assess the performance of the proposed tracker. For comparison purposes, we used 15 modern trackers with the proposed approach in the LaSOT dataset. In addition, the LaSOT dataset is a large-scale dataset, which contains 1400 video sequences with 70 different object categories and each category has 20 sequences. In this work, we used the 1400 sequences for our tracker evaluation. Moreover, the experimental results are exhibited in Fig. 12. As shown in Fig. 12a and Fig. 12b, we can see that the presented approach obtains remarkable tracking efficiency in terms of a precision score of 50.0% and success score of 47.5%, respectively.

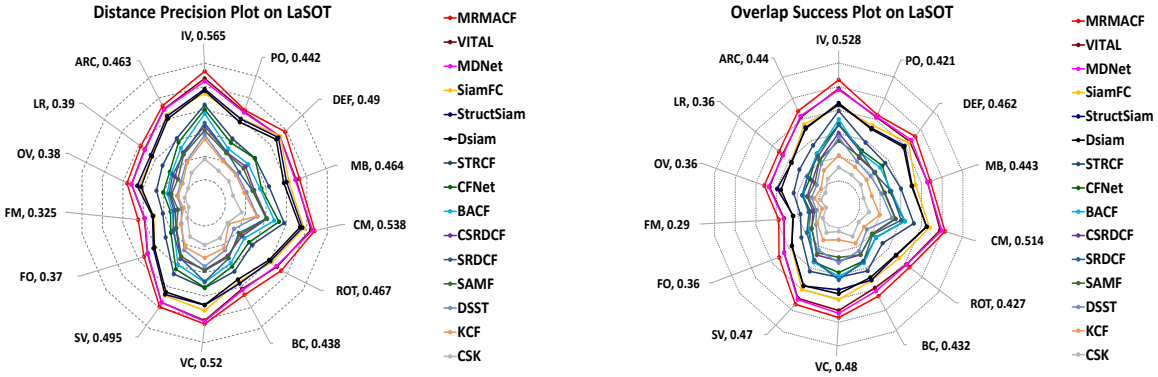


(a) (b)

Figure 12: The distance precision and overlap success results of the LaSOT dataset are exhibited in Fig. 12a and Fig. 12b, respectively.

In particular, we conduct the 14 attribute-based evaluations on the LaSOT dataset to estimate the proposed tracker performance with other conventional trackers. Also, the attribute-based evaluation results are exhibited in Fig. 13a and Fig. 13b. Specifically, we ensure that the presented technique increases the considerable tracking results in DP and AUC scores of (19.3% / 19.0%) when compared to the baseline tracker. Furthermore, compared to the hand-crafted feature-based trackers such as STRCF [8], CSRDCF

[41], SRDCF [6], SAMF [35], and DSST [38], we can see that our tracking approach enhances the outstanding performance in terms of DP / AUC scores of (14.7% / 14.4%), (22.1% / 22.3%), (22.1% / 19.6%), (22.9% / 22.8%), and (26.1% / 24.3%), respectively. Meanwhile, when compared to the deep feature-based trackers such as VITAL [50], MDNet [43], SiamFC [42], StructSiam [49], and CFNet [51], we notice that our tracking approach increases the superior efficiency in terms of DP and AUC scores of (1.6% / 2.1%), (1.9% / 1.9%), (5.1% / 6.6%), (5.7% / 8.1%), and (17% / 18.9%), respectively. In the end, we confirm that our presented approach outperforms other existing trackers.



(a) (b)

Figure 13: Attribute-based evaluation of the distance precision and overlap success scores on the LaSOT dataset.

4.3.8. Evaluation on GOT-10K Dataset

We also estimate our proposed approach on the GOT-10K large-scale benchmark dataset, which contains more than 10000 video sequences. Following that, we employ the 180 test sequence with 84 different objects to evaluate our proposed tracker performance. Moreover, we evaluate the success rate metric on the GOT-10K dataset, which achieves significant performance in the AUC score of 57.6%, as shown in Fig. 14. Besides, when compared to the baseline tracker BACF [5], we observe that our tracking method enhances better tracking performance in the AUC score of 31.6%. Specifically, we ensure that when compared to the hand-crafted feature-based methods such as DSST [38], SAMF [35], CSK [40], and KCF [36] our presented approach improves remarkable tracking performance in the success rate 32.9%, 33.0%, 37.1%, and 37.3%, respectively. Likewise, when compared to the deep feature-based trackers such as DaSiamRPN [11], ATOM [46], FDSiamFC [45], SiamFC [42], and GOTRUN [48], we can see that our tracker obtains significant results in AUC scores 13.2%, 2.0%, 22.6%, 22.8%, and 22.9%, respectively. From these hand-crafted and deep feature analyses, we confirmed that our proposed MRMACF tracker demonstrates superior efficiency on the GOT-10K benchmark dataset.

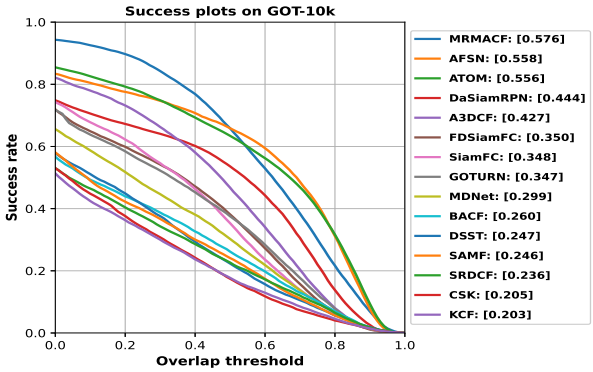


Figure 14: The success plots of the state-of-the-art trackers on the GOT-10K dataset.

4.4. Qualitative Analysis

We conduct the qualitative analysis of our proposed method with other state-of-the-art-trackers such as STAR [16], BACF [5], BSTCF [7], SRDCF [7], SASR [9], and STRCF [8]. Also, the qualitative comparison results are exhibited in Fig. 15. As shown in Fig. 15, we compared the proposed tracker on five challenging video sequences. In addition, the STAR, BACF, BSTCF, and STRCF trackers fail to capture the target location accurately due to the fast motion, object rotation, and occlusion.

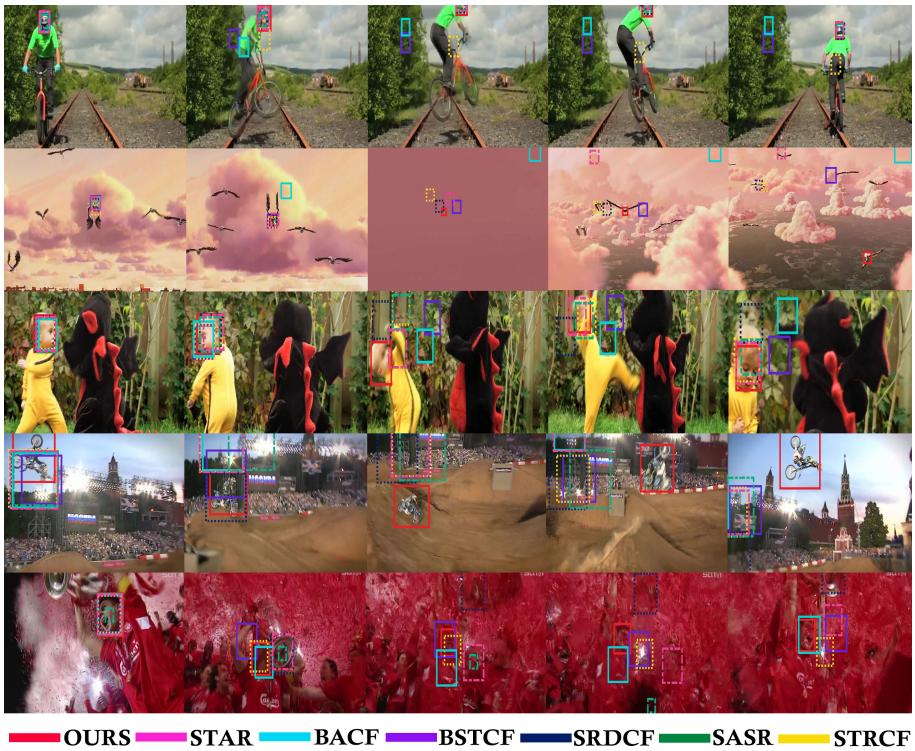


Figure 15: We conducted a comparative analysis of our MRMACF tracker against six modern trackers, namely BACF [5], BSTCF [7], SRDCF [6], SASR [9], STRCF [8], and STAR [16], across five challenging sequences in the OTB-2015 dataset. These sequences are Biker, Bird1, DragonBaby, MotorRolling, and Soccer, listed from top to bottom.

Specifically, when the target undergoes fast motion and object rotation in the Biker and DragonBaby sequences, the BACF, BSTCF, SRDCF, and STRCF trackers failed to monitor the target region. Despite the fast motion and object rotation, our proposed tracker track the target region accurately as shown in Fig. 15. Moreover, when occurring occlusion and object rotation in the Bird1 and Soccer sequences, the proposed tracker performs well until the end of the video sequence compared to the other trackers. In particular, the presented approach demonstrated significant performance compared to the other all trackers, when the target undergoes the object rotation in the MotorRolling sequence. Overall, we can confirm that the presented tracker performs well in all these scenarios compared to the other conventional trackers.

5. Conclusion

In this study, we proposed a novel multi-regularized mutation-aware correlation filter via an adaptive hybrid model. In this perspective, mutation-aware regularization with an adaptive hybrid model approach has been proposed to overcome the mutation issue during tracking. Following that, the sparse spatial feature selection method has utilized the row and column-based feature selection techniques to obtain the prominent features of the target region. Then, a surrounding-aware approach has been proposed to extract the surrounding samples from the target region, utilize the context information, and find the exact target location. Specifically, this study utilized an adaptive hybrid model with a temporal regularization approach that helps to redetect the object during the mutation by leveraging information from the previous frame of the target region. Finally, we have demonstrated that experimental results on the benchmark datasets OTB-2013, OTB-2015, TempleColor-128, UAV-123, UAVDT, VOT-2018, LaSOT, and GOT-10K show the superiority of our proposed method over other state-of-the-art trackers.

Acknowledgement

This work was partially supported by the Basic Science Research Program through the National Research Foundation (NRF) funded by the Ministry of Education of South Korea (NRF-2016R1A6A1A03013567, NRF-2021R1A2B5B01001484).

References

[1] W. Han, C. K. L. Lekamalage, G.-B. Huang, Efficient joint model learning, segmentation and model updating for visual tracking, Neural Networks 147 (2022) 175–185.

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

[3] J. Wang, C. Lai, Y. Wang, W. Zhang, Emat: Efficient feature fusion network for visual tracking via optimized multi-head attention, Neural Networks (2024) 106110.

[4] D. Elayaperumal, Y. H. Joo, Learning spatial variance-key surrounding-aware tracking via multi-expert deep feature fusion, Information Sciences 629 (2023) 502–519.

[5] 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.

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

[8] 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.

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

[10] 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.

[11] 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.

[12] H. Ren, L. Xing, T. Shi, Research on background learning correlation filtering algorithm with multi-feature fusion, IEEE Access 11 (2023) 32895–32906. doi:10.1109/ACCESS.2023.3262726.

[13] K. Nai, Z. Li, H. Wang, Dynamic feature fusion with spatial-temporal context for robust object tracking, Pattern Recognition 130 (2022) 108775.

[14] 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 Transactions on Image Processing 28 (11) (2019) 5596–5609.

[15] 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 Transactions on Multimedia 24 (2021) 301–312.

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

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

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

[19] F. Gu, J. Lu, C. Cai, Rpformer: A robust parallel transformer for visual tracking in complex scenes, IEEE Transactions on Instrumentation and Measurement 71 (2022) 1–14.

[20] H. Zhang, Y. Piao, N. Qi, Stft: Spatial and temporal feature fusion for transformer tracker, IET Computer Vision (2023).

[21] R. Wu, X. Wen, L. Yuan, H. Xu, Y. Liu, Visual tracking based on deformable transformer and spatiotemporal information, Engineering Applications of Artificial Intelligence 127 (2024) 107269.

[22] J. Zhang, Y. He, S. Wang, Learning adaptive sparse spatially-regularized correlation filters for visual tracking, IEEE Signal Processing Letters 30 (2023) 11–15.

[23] Z. Ji, K. Feng, Y. Qian, J. Liang, Sparse regularized correlation filter for uav object tracking with adaptive contextual learning and keyfilter selection, Information Sciences 658 (2024) 120013.

[24] Y. Wu, J. Lim, M.-H. Yang, Object tracking benchmark, IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (9) (2015) 1834–1848. doi:10.1109/TPAMI.2014.2388226.

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

[26] 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.

[27] 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.

[28] M. Kristan, A. Leonardis, J. Matas, M. Felsberg, R. Pflugfelder, L. ˇCehovin Zajc, T. Vojir, G. Bhat, A. Lukezic, A. Eldesokey, et al., The sixth visual object tracking vot2018 challenge results, in: Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018, pp. 0–0.

[29] 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.

[30] L. Huang, X. Zhao, K. Huang, Got-10k: A large high-diversity benchmark for generic object tracking in the wild, IEEE Transactions on Pattern Analysis and Machine Intelligence 43 (5) (2019) 1562–1577.

[31] 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.

[32] 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/CVF International Conference on Computer Vision, 2019, pp. 2891–2900.

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

[34] 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.

[35] 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.

[36] J. F. Henriques, R. Caseiro, P. Martins, J. Batista, High-speed tracking with kernelized correlation filters, IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (3) (2014) 583–596.

[37] 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/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11923–11932.

[38] M. Danelljan, G. Häger, F. S. Khan, M. Felsberg, Discriminative scale space tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (8) (2016) 1561–1575.

[39] 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.

[40] J. F. Henriques, R. Caseiro, P. Martins, J. Batista, Exploiting the circulant structure of tracking-by-detection with kernels, in: Proceedings of the European Conference on Computer Vision (ECCV), Springer, 2012, pp. 702–715.

[41] A. Lukezic, T. Vojir, L. ˇCehovin 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.

[42] 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.

[43] 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.

[44] S. Li, Y. Liu, Q. Zhao, Z. Feng, Learning residue-aware correlation filters and refining scale for real-time uav tracking, Pattern Recognition 127 (2022) 108614.

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

[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] S. Peng, Y. Yu, K. Wang, L. He, Accurate anchor free tracking, arXiv preprint arXiv:2006.07560 (2020).

[48] D. Held, S. Thrun, S. Savarese, Learning to track at 100 fps with deep regression networks, in: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, Springer, 2016, pp. 749–765.

[49] Y. Zhang, L. Wang, J. Qi, D. Wang, M. Feng, H. Lu, Structured siamese network for real-time visual tracking, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 351–366.

[50] Y. Song, C. Ma, X. Wu, L. Gong, L. Bao, W. Zuo, C. Shen, R. W. Lau, M.-H. Yang, Vital: Visual tracking via adversarial learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8990–8999.

[51] 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.