symmetry

MDPI

Article

Learning Augmented Memory Joint Aberrance Repressed Correlation Filters for Visual Tracking

Yuanfa Ji [1], Jianzhong He [2], Xiyan Sun [3,*], Yang Bai [4,*], Zhaochuan Wei [2,*] and Kamarul Hawari bin Ghazali [5,*]

[1] National & Local Joint Engineering Research Center of Satellite Navigation Positioning and Location Service, Guilin University of Electronic Technology, Guilin 541004, China; jiyuanfa@guet.edu.cn
[2] School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China; hjz17805972248@163.com
[3] Guangxi Key Laboratory of Precision Navigation Technology and Application, Guilin University of Electronic Technology, Guilin 541004, China
[4] GUET-Nanning E-Tech Research Institute Co., Ltd., Nanning 530031, China
[5] Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang, Pekan 26600, Malaysia
* Correspondence: sunxiyan@guet.edu.cn (X.S.); by@guet.edu.cn (Y.B.); weizc@guet.edu.cn (Z.W.); kamarul@ump.edu.my (K.H.b.G.)

Abstract: With its outstanding performance and tracking speed, discriminative correlation filters (DCF) have gained much attention in visual object tracking, where time-consuming correlation operations can be efficiently computed utilizing the discrete Fourier transform (DFT) with symmetric properties. Nevertheless, the inherent issues of boundary effects and filter degradation, as well as occlusion and background clutter, degrade the tracking performance. In this work, we proposed an augmented memory joint aberrance repressed correlation filter (AMRCF) for visual tracking. Based on the background-aware correlation filter (BACF), we introduced adaptive spatial regularity to mitigate the boundary effect. Several historical views and the current view are exploited to train the model together as a way to reinforce the memory. Furthermore, aberrance repression regularization was introduced to suppress response anomalies due to occlusion and deformation, while adopting the dynamic updating strategy to reduce the impact of anomalies on the appearance model. Finally, extensive experimental results over four well-known tracking benchmarks indicate that the proposed AMRCF tracker achieved comparable tracking performance to most state-of-the-art (SOTA) trackers.

Keywords: visual object tracking; discriminative correlation filter; augmented memory; aberrance repression

Citation: Ji, Y.; He, J.; Sun, X.; Bai, Y.; Wei, Z.; Ghazali, K.H.b. Learning Augmented Memory Joint Aberrance Repressed Correlation Filters for Visual Tracking. *Symmetry* **2022**, *14*, 1502. https://doi.org/10.3390/sym14081502

Academic Editor: Antonio Palacios

Received: 17 June 2022
Accepted: 20 July 2022
Published: 22 July 2022

1. Introduction

Visual object tracking, as a fundamental task of computer vision and pattern recognition, is widely applied in biological vision, autonomous driving, video surveillance and other fields. The task of visual tracking is to predict the spatial position and scale size of an arbitrary target in a video or continuously associated image sequence given limited information (i.e., initial target position and scale information). Due to the variable tracking environment, the tracked object is easily suffered from various factors, such as illumination variation, deformation, partial/full occlusion, causing lost target and tracking failure. Therefore, achieving stable and accurate tracking of an object under complex environments is still a challenge.

Since Bolme [1] et al. first proposed the minimum output sum of the squared error (MOSSE) tracker, which achieved promising accuracy and very fast speed (615 FPS) only using the single-channel greyscale features, DCF-based tracking methods have become a hot topic in visual tracking. By means of DFT with conjugate symmetry, the time-consuming correlation operation can be efficiently solved in the frequency domain. Inspired

by the MOSSE, CSK [2] creatively adopts circular shift instead of random sampling to increase the training samples and introduced kernel functions to improve the computational speed. While the CSK-based multi-channel kernel correlation filter (KCF) [3] achieved a great improvement in tracking performance, the follow-up trackers have made significant advancements over scale adaptation [4–6], feature representation and fusion scheme [7–9], models innovation [10–12], etc., however, several inherent problems remain. First, the training samples based on the periodic assumption are low-quality synthetic samples that produce unwanted boundary effects, making the filter's discriminatory power less than optimal. Second, to accommodate the appearance changes, most trackers maintain and update an appearance model frame by frame with a pre-determined fixed learning rate. This ignores the appearance variations and lacks the integration of historical appearances information, leading to model degradation and reducing its effectiveness. Third, in the case of challenging scenarios such as partial/full occlusion and background clutter, the object appearance changes drastically while the output confident maps become distorted, i.e., response anomalies, which may cause the target drift and tracking failure.

To deal with the above challenges, the existing SOTA trackers mainly focus on two aspects to improve tracking performance: (a) by imposing spatial or temporal regularization constraints whilst constructing the model, and (b) by analysing the inherent information of the confidence map to construct feedback loops to optimize the entire tracking framework. Spatially regularized discriminative correlation filter (SRDCF) [13] introduces a spatial regularization that penalizes the filter coefficients at the boundaries, allowing more energy to be concentrated in the central region. CSR-DCF [14] distinguishes the foreground from the background by colour segmentation in the search area, while the foreground mask matrix is utilized to select the filter coefficients. A background-aware correlation filter (BACF) [15] utilizes a binary mask matrix to crop real samples from the search window and achieve the suppression of background information. However, it is not reasonable to impose fixed constraints on the filter, which does not reflect well the changing characteristics and appearance of the tracking target. Furthermore, by introducing temporal consistency constraints [16–18], the aim is that the model does not over-change during the tracking of the same target, thus mitigating the model's degradation. Nevertheless, the existence of non-ideal factors such as occlusion and background interference make this assumption difficult to achieve. Since minimizing the impact of response anomalies, on the one hand, Huang et al. [19] directly introduced a regularization constraint via comparing the Euclidean distance between the current and the previous output response maps to limit the response distortion. Meanwhile, many works [20–22] have developed high-confidence updating strategies to construct feedback loops, which have effectively improved the tracking performance. However, these trackers hardly exploit the intrinsic information provided by different historical views, increasing the risk of model drift.

In this work, we proposed an augmented memory joint aberrance repressed correlation filter (AMRCF) tracking method to address the above limitation. Based on the excellent BACF tracker, we incorporate the adaptive spatial regularization constraint to mitigate the boundary effects. To mitigate the model degradation, several historical views are selected to train the filter model together with the current view, making the trained model adapt to the new target appearance and remember the previous ones. To adapt with the complex tracking environment, we introduce an aberrance repression regularization constraint to limit the drastic response changes, and propose a high-confidence updating strategy to optimize the overall tracking framework. The main contributions are summarized in the following three-folds:

- The AMRCF method is presented by simultaneously introducing adaptive spatial regularization, augmented memory regularization and aberrance repression regularization into the DCF framework. Combined with a dynamic appearance model update strategy, the overall tracking performance is improved in the case of partial/full occlusion, deformation and background clutter.

- Using the alternating direction method of multipliers (ADMM) [23] algorithm enables the model closure solution to be efficiently calculated.
- The overlap success and distance precision scores on four extensive benchmarks (i.e., OTB50, OTB100, TC128 and UAV123) verified that the proposed AMRCF has an excellent tracking performance comparable to state-of-the-art (SOTA) trackers.

2. Related Works

In this section, we mainly review the three categories of DCF tracking methods that are most relevant to our tracker, including the boundary effect-aware trackers, refined tracking models and high-confident updating schemes.

Due to the periodic assumption of the training samples generated by circular shifting not fully reflecting the sampling information, the tracking model is prone to over-fitting. To overcome the unwanted boundary effects, there are two main improvement directions for DCF. One is based on SRDCF, with the addition of spatial regularization, feature dimensionality reduction, feature interpolation and confidence map fusion [24–26]. Another is to adopt the BACF strategy of labelling around the target region, performing spatial feature restriction and adding constraint terms to the model [27–30]. Huang et al. [29] proposed a background suppressed correlation filter (BSCF) that incorporates all global background patches to enhance the tracking performance. ASRCF [31] incorporates SRDCF and BACF by employing the adaptive spatial regularization term, with the model adaptively penalizing the filter coefficients in the event of occlusion. A context-aware correlation filter CACF [32] takes into account global information, explicitly learning the background information around the target. Zha et al. [33] proposed the semantic-aware spatial regularization correlation filter by using spatial semantic maps to model regularization and feature selection.

To minimize the negative impacts of the model degradation and other issues, DCF has developed numerous improvements. The spatial–temporal regularized correlation filter (STRCF) [16] incorporates a temporal regularity term to minimize overfitting and reduce tracking failures due to target occlusion and distortion. Li et al. [34] proposed the augmented memory for correlation filter (AMCF) for jointly training the model with several frames of historical views and the current view to improve the stability of the model. LADCF [35] incorporates temporal consistency constraints into the model for enhanced tracker robustness and reduced model degradation. In recent work, many trackers incorporated spatial–temporal regularization [36–38], while the tracking performance has been significantly improved. Yu et al. [36] proposed a second-order spatial–temporal correlation filter (SSCF), which incorporates both the first-order and second-order data-fitting terms into the DCF framework. Hu et al. [37] merged the context-aware model into the STRCF tracker, thus expanding the target search domain and obtaining more discriminative information.

The output response maps are well reflective of the target state. In order to make better use of the information in the response maps, much work is performed in confidence map evaluation [1,39] and model updating strategies [20–22,40,41]. Bolme et al. [1] proposed the peak-to-sidelobe ratio (PSR) as the basis for the detection of confidence maps. Wang et al. [39] proposed the average peak-to-correlation energy (*APCE*) to predict the target state by comparing the change of *APCE* values for confidence maps. Fu et al. [20] proposed to control the update to the tracker by verifying the consensus score. Gan et al. [21] proposed a long-term correlation tracker, which can perform a long-term memory function by activating an online random fern classifier for re-detection when the PSR value of the confidence map is below a threshold. MUSTer [22] employs the Atkinson–Shiffrin memory model, which is divided into a long-term memory module for key points and a short-term memory module: when tracking fails or is obscured, the short-term module is updated. Ma et al. [40] proposed the bidirectional incongruity-aware correlation filter (BiCF), which predicts the tracking state by means of forward detection and backward relocation, and obtains a more robust model by suppressing bidirectional incongruity errors. Liu et al. [41]

proposed a long-term tracker with multiple features and saliency redetection (MSLT), which consists of tracking-by-detection and redetection parts, and effectively improving the performance of the model.

3. Proposed Method

3.1. Review the BACF Tracker

Unlike traditional KCF [3] tracker, the BACF tracker expands the detection region and crops the real training samples using the binary mask matrix P. This allows BACF to obtain more high-quality negative samples, which greatly mitigates the boundary effect and achieves greater performance. The objective function of BACF is as follows:

$$\varepsilon(h_c) = \frac{1}{2}\|y_c - \sum_{d=1}^{D} x_c^d * \left(P^T h_c^d\right)\|_2^2 + \frac{\lambda}{2}\sum_{d=1}^{D}\|h_c^d\|_2^2 \quad (1)$$

where the subscript c indicates the current frame; $x \in \mathbb{R}^{M\times N\times D}$ denotes the D-dimensional feature maps of size $M \times N$ extracted from the input image; and $y \in \mathbb{R}^{M\times N}$ is represented as the ideal output of the filter. The feature maps x are correlated with the trained filter $h \in \mathbb{R}^{M\times N\times D}$ to obtain the final output $\sum_{d=1}^{D} x_c^d * h_c^d$. λ is a regularization parameter.

Despite the good tracking performance of the BACF, there are still several issues that need to be further addressed: (1) the traditional DCF tracker, which utilizes a fixed learning rate to maintain and update the appearance model frame by frame, does not take into account the current object state, which made it difficult to adapt the complex tracking environment changes; (2) there is no reaction tactic to cope with anomalies and the fact that tracking targets can be easily lost; and (3) further optimization can be made by dealing with boundary effects.

3.2. Objective Function of AMRCF

Aiming at the weaknesses of the BACF tracker, we propose an augmented memory joint aberrance repressed correlation filter (AMRCF) to improve in terms of model stability and accuracy. The overall workflow and framework of the proposed AMRCF are represented in Figures 1 and 2. The proposed AMRCF tracker based on the BACF tracker, introduces adaptive spatial regularization, augmented memory regularization and aberrance repression regularization to improve the overall tracking performance. Meanwhile, combining the update strategy of the appearance model, AMRCF can adapt to the changing tracking scenarios, and can then achieve more accurate and robust tracking results.

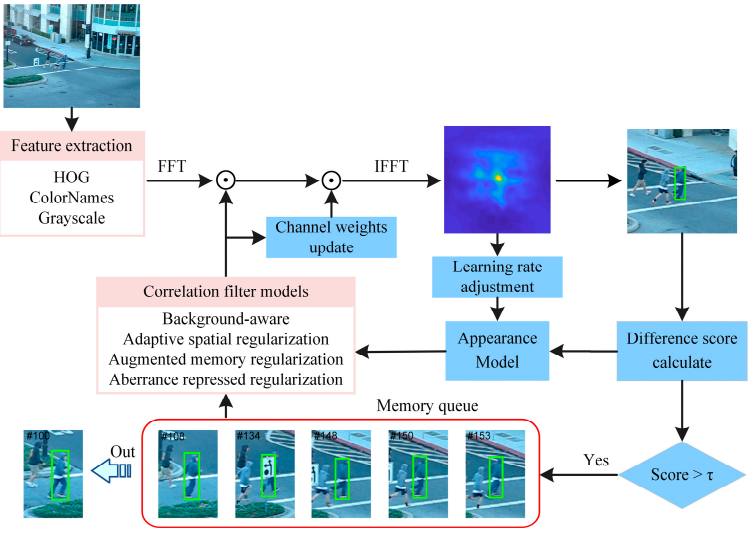

Figure 1. The flowchart of the proposed AMRCF tracker.

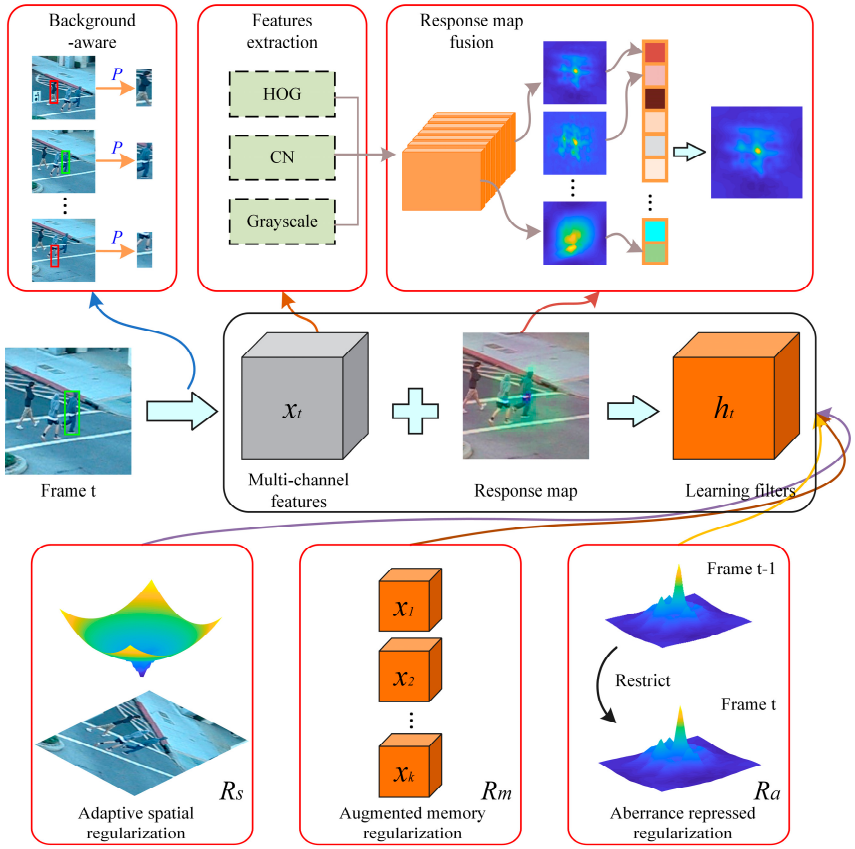

Figure 2. Overall framework of the proposed AMRCF tracker.

The objective function $\varepsilon(h)$ of AMRCF is as follows:

$$\varepsilon(h) = \frac{1}{2}\|y_c - \sum_{d=1}^{D} x_c^d \times \left(P^T h_c^d\right)\|_2^2 + \mathcal{R}_s + \mathcal{R}_m + \mathcal{R}_a \quad (2)$$

where $\mathcal{R}_s$, $\mathcal{R}_m$ and $\mathcal{R}_a$ denote the adaptive spatial regularization term, the augmented memory regularization term and the aberrance repressed regularization term, respectively.

3.2.1. Adaptive Spatial Regularization

The SRDCF model introduces the spatial regularization constraint, which is a negative Gaussian-shaped spatial weight vector, making it possible to have a strong response around the centre of the tracked object. Nevertheless, the fixed regularity fails to reflect the changing appearance of the target. Therefore, the proposed AMRCF introduced adaptive spatial regularization that penalizes the filter coefficients at the unreliable parts while approximating the spatial weights w to the a priori reference weights w_r, preventing model degradation.

$$\mathcal{R}_s = \frac{\lambda_1}{2}\sum_{d=1}^{D}\|w \odot h_c^d\|_2^2 + \frac{\lambda_2}{2}\|w - w^r\|_2^2 \quad (3)$$

where λ_1 and λ_2 denote the regularization coefficients of the respective regularization terms.

3.2.2. Augmented Memory Regularization

Due to the frame-by-frame update strategy of the appearance model, the historical view will be forgotten at an exponential rate as the number of iterations increases. The model will focus more on the most recent views, which reduce the anti-interference capability of the model. In this work, we introduced an augmented memory regularization constraint which utilized the perceptual hashing algorithm (PHA) [42] to select K-frame

historical views with distinct differences and make them train the filter model together with the current view for enhancing memory and mitigating model degradation:

$$\mathcal{R}_m = \frac{\lambda_3}{2} \sum_{k=1}^{K} \| y_k - \sum_{d=1}^{D} x_k^d \times \left(P^T h_c^d \right) \|_2^2 \tag{4}$$

where x_k^d and y_k denote the d-channel training sample of the k-th view in the memory sequence and its ideal output, respectively. h_c denotes the filter model of the current frame. λ_3 denotes the regularization coefficient. K indicates the capacity of the memory queue for storing the history view.

PHA generated a unique "fingerprint" for each view by comparing the fingerprint differences between the before and after frame views to decide whether to update and maintain the memory queue.

The initial view is converted to grayscale and transformed to the frequency domain using the discrete cosine transform (DCT). Only the low frequency region $A \in \mathbb{R}^{W \times W}$. with high energy density is retained. Then, each point $h_{i,j}$ is compared with the average of all elements to obtain the hash matrix P:

$$p_{i,j} = \begin{cases} 1, & h_{i,j} > \frac{1}{W^2} \sum_{i=1}^{W} \sum_{j=1}^{W} h_{i,j} \\ 0, & others \end{cases} \tag{5}$$

By XOR operation, the difference score S between the current frame P^c and the latest view P^k in the memory queue is derived:

$$S = \frac{1}{W^2} \sum_{i=1}^{W} \sum_{j=1}^{W} \left(p_{ij}^c \oplus p_{ij}^k \right) \tag{6}$$

When the obtained score is greater than a certain threshold τ, indicating that there is a significant difference between the current frame object and the latest frame in the memory queue, then the memory queue will be updated according to the "first-in-first-out" principle.

3.2.3. Aberrance Repression Regularization

Ideally (i.e., without much change in the object appearance), the response of two adjacent frames does not tend to change much. However, abrupt changes in appearance caused by background clutter, target occlusion, etc., will cause response anomalies. As aberration occur, the similarity between the output response maps M_1 and M_2 suddenly drops, while the Euclidean distance between M_1 and M_2 will become larger. In this work, we adopted aberrance-suppressed regularization to restrict the response variation, which can effectively limit the response variation:

$$\mathcal{R}_a = \frac{\lambda_4}{2} \| \sum_{d=1}^{D} x_{c-1}^d \times \left(P^T h_{c-1}^d \right) \left[\psi_{p,q} \right] - \sum_{d=1}^{D} x_c^d \times \left(P^T h_c^d \right) \|_2^2 \tag{7}$$

where p and q denote the difference in the peak position between the response maps and the shift operation $\left[\psi_{p,q} \right]$ to make the peak points of two response maps overlap each other. λ_4 denotes the regularization coefficient.

3.3. Optimization

Introducing auxiliary variables $\hat{G} = [\hat{g}^1, \hat{g}^2, \ldots, \hat{g}^D] (\hat{g}^d = \sqrt{T} F P^T h^d,\ d = 1, 2, \ldots, D)$, (where the superscript $\hat{}$ denotes the discrete Fourier operator and F denotes the standard

orthogonal matrix), which translated Equation (2) into the frequency domain representation according to Parseval's theorem:

$$\begin{aligned} E(H,\hat{G},w) = \tfrac{1}{2T}\|\hat{y}_c - \sum_{d=1}^{D}\hat{x}_c^d \odot \hat{g}_c^d\|_2^2 + \tfrac{\lambda_1}{2}\sum_{d=1}^{D}\|w\odot h_c^d\|_2^2 + \tfrac{\lambda_2}{2}\|w-w^r\|_2^2 \\ + \tfrac{\lambda_3}{2T}\sum_{k=1}^{K}\left\|\hat{y}_k - \sum_{d=1}^{D}\hat{x}_k^d\odot\hat{g}_c^d\|_2^2 + \tfrac{\lambda_4}{2}\right\|M_{c-1}\left[\psi_{p,q}\right] - \sum_{d=1}^{D}\hat{x}_c^d\odot\hat{g}_c^d\|_2^2 \end{aligned} \tag{8}$$

where the output response of the $(c-1)$-th frame is rewritten as $M_{c-1} = \sum_{d=1}^{D} x_{c-1}^d \times \left(P^T h_{c-1}^d\right)$.

Rewrite Equation (8) using the augmented Lagrange multiplier method (ALM) as:

$$\mathcal{L}(H,\hat{G},w,\hat{\zeta}) = E(H,\hat{G},w) + \hat{\zeta}^T(\hat{g}_c^d - \sqrt{T}FP^T h_c^d) + \frac{\mu}{2}\|\hat{g}_c^d - \sqrt{T}FP^T h_c^d\|_2^2 \tag{9}$$

where the penalty factor μ and the matrix of auxiliary variables $\hat{\zeta}^T = [\hat{\zeta}^{1T}, \hat{\zeta}^{2T}, \ldots, \hat{\zeta}^{DT}]^T$ are introduced.

In order to obtain closed solutions, Equation (9) is solved in steps via ADMM, for which we decompose it into three sub-problems to solve, and all three sub-problems have closed solutions.

Subproblem h_{c+1}^*: If $\hat{G}$, w and $\hat{\zeta}$ are given, then h yields:

$$\begin{aligned} h_{c+1}^* &= \left(T\mu + \lambda_1 w^T w\right)^{-1}\left(\sqrt{T}PF^T\hat{\zeta} + \mu\sqrt{T}PF^T\hat{g}_c\right) \\ &= \left(\mu I_c + \frac{\lambda_1 w^T w}{T}\right)^{-1}(\zeta + \mu g_c) \end{aligned} \tag{10}$$

where ζ and g can be easily obtained via the Fourier inverse operation on $\hat{\zeta}$ and $\hat{g}_c$.

$$\begin{cases} \zeta = \frac{1}{\sqrt{T}}PF^T\hat{\zeta} \\ g_c = \frac{1}{\sqrt{T}}PF^T\hat{g}_c \end{cases} \tag{11}$$

Subproblem $\hat{G}^*$: If given w, $\hat{\zeta}$ and h_{c+1}^*, since each channel information is relatively independent for the sample $\hat{x}_c^d$, the subproblem $\hat{G}^*$ can be further split into $d = \{1, 2, \ldots, D\}$ smaller problems, each of which can be denoted as:

$$\hat{G}^*(d) = \underset{\hat{G}^*(d)}{argmin}\left\{\begin{array}{c} \frac{1}{2T}\|\hat{y}_c - \hat{x}_c^T\hat{g}_c\|_2^2 + \frac{\lambda_3}{2T}\sum_{k=1}^{K}\|\hat{y}_k - \hat{x}_k^T\hat{g}_c\|_2^2 \\ + \frac{\lambda_4}{2}\|M_{c-1}\left[\psi_{p,q}\right] - \hat{x}_c^T\hat{g}_c\|_2^2 \\ + \hat{\zeta}^T\left(\hat{g}_c - \sqrt{T}FP^T h_c\right) + \frac{\mu}{2}\|\hat{g}_c^d - \sqrt{T}FP^T h_c\|_2^2 \end{array}\right\} \tag{12}$$

The closed solution of $\hat{G}^*(d)$ can be solved as:

$$\begin{aligned} \hat{G}^*(d) = \left[\left(\lambda_4 + \tfrac{1}{T}\right)\hat{x}_c\hat{x}_c{}^T + \tfrac{\lambda_3}{T}\textstyle\sum_{k=1}^{K}\hat{x}_k\hat{x}_k{}^T + \mu\right]^{-1} \\ \left(\tfrac{1}{T}\hat{x}_c\hat{y}_c + \tfrac{\lambda_3}{T}\textstyle\sum_{k=1}^{K}\hat{x}_k\hat{y}_k + \lambda_4\hat{x}_c M_{c-1} - \hat{\zeta} + \mu\hat{h}_c\right) \end{aligned} \tag{13}$$

Further reducing the computational effort by using the Sherman–Morrison formula, i.e., $\left(uv^T + A\right)^{-1} = A^{-1} - \dfrac{A^{-1}uv^TA^{-1}}{1+v^TA^{-1}u}$ (here u and v are two column vectors and uv^T is

a rank one matrix), here set $u = v^T = \hat{x}_c$, $A = \frac{\lambda_3}{T}\sum_{k=1}^{K}\hat{x}_k\hat{x}_k{}^T + \mu$, $B = \lambda_4 + \frac{1}{T}$. $\hat{G}^*(d)$ is further derived as:

$$\begin{aligned}\hat{G}^*(d) = \frac{1}{A}\left(\frac{1}{T}\hat{x}_c\hat{y}_c + \frac{\lambda_3}{T}\sum_{k=1}^{K}\hat{x}_k\hat{y}_k + \lambda_4\hat{x}_c M_{c-1} - \hat{\zeta} + \mu\hat{h}_c\right) \\ - \frac{\hat{x}_c}{Ab}\left[\frac{\hat{S}_{xx}\hat{y}_c}{T} + \frac{\lambda_4\hat{S}_{xc}}{T} + \lambda_3\hat{S}_{xx}M_{c-1} - \hat{S}_{x\zeta} + \mu\hat{S}_{xh}\right]\end{aligned} \tag{14}$$

where $b = \frac{A}{B} + \hat{x}_c{}^T\hat{x}_c$, $\hat{S}_{xx} = \hat{x}_c{}^T\hat{x}_c$, $\hat{S}_{xc} = \hat{x}_c{}^T\sum_{k=1}^{K}\hat{x}_k\hat{y}_k$, $\hat{S}_{x\zeta} = \hat{x}_c{}^T\hat{\zeta}$, $\hat{S}_{xh} = \hat{x}_c{}^T\hat{h}_c$.

Subproblem w^*: If h is given, then w can be solved for:

$$\begin{aligned}w^* &= \left(\lambda_1\sum_{d=1}^{D}h_c^d \odot h_c^d + \lambda_2 I\right)^{-1}\lambda_2 w^r \\ &= \frac{\lambda_2 w^r}{\lambda_1\sum_{d=1}^{D}h_c^d \odot h_c^d + \lambda_2 I}\end{aligned} \tag{15}$$

Update of Lagrange multipliers $\hat{\zeta}$: Updates to $\hat{\zeta}$ are made by:

$$\hat{\zeta}_{i+1} = \hat{\zeta}_i + \mu * \left(\hat{G}_{i+1} - \hat{h}_{i+1}\right) \tag{16}$$

where the subscripts i and $i+1$ denote the i-th and $(i+1)$-th iterations, respectively. $\hat{G}_{i+1}$ and $\hat{h}_{i+1}$ denote the subproblem solutions obtained in the i + 1-th iteration, respectively. Here, the canonical constant $\mu^{i+1} = \min(\mu_{max}, \beta\mu^i)$.

Complexity Analysis: Since each pixel is relatively independent, we need to solve the *D*MN* subproblems, where *D* represents the number of channels. Since the subproblem G requires the FFT and inverse FFT transformation in each iteration, so the computational complexity is *O*(*DMN*log*(MN)*). Meanwhile, the computational complexity of both subproblems *h* and *w* is *O(DMN)*. Therefore, the overall complexity of model is *O(LDMN*log*(MN))* when the number of iterations is *L*. It is worth remarking that $\sum_{k=1}^{K}\hat{x}_k\hat{y}_k$ and $\sum_{k=1}^{K}\hat{x}_k\hat{y}_k$ in Equation (14) represent the actual view information stored in the memory queue, and are constant terms that do not require additional computational resources.

3.4. Detection

To strengthen the effects of the reliable channels on the final response results, different channels are assigned corresponding channel weights. The PSR value, as the evaluation criterion for the confident map [1], is adopted as the reference for the performance representation of the different channels, and the channel weights are determined and updated as follows:

$$C_{c+1}^d = (1-\gamma)C_c^d + \gamma\frac{PSR\left(\hat{h}_c^{d*} \odot \hat{x}_c^d\right)}{\sum_{d=1}^{D}PSR\left(\hat{h}_c^{d*} \odot \hat{x}_c^d\right)} \tag{17}$$

where γ denotes the learning rate of the channel weights. $PSR(\cdot) = \frac{(\max(R)-\mu)}{\sigma}$, R denotes the output response map, μ and σ denote the mean and mean squared deviation of the response map, respectively.

The final output response R_c can be derived as follows:

$$R_c = \mathcal{F}^{-1}\left(\sum_{d=1}^{D}C_c^d\hat{h}_{c-1}^{d*} \odot \hat{x}_c^d\right) \tag{18}$$

where $\mathcal{F}^{-1}$ denotes the Fourier inverse operation. The individual channel output responses are multiplied by the corresponding channel weights to the final output response map, and the maximum point is the target position.

3.5. Model Update

To adapt to the changing tracking environment, we use the maximum peak and *APCE* values as a basis for determining the reliability of the output response. As shown in Figure 3, ideally, the output response approximates a two-dimensional normal distribution with a relatively flat response and prominent main peak, and the corresponding *APCE* values are in the higher range. When anomalies occur such as occlusion, blurring, and illumination variation, the output response map fluctuates dramatically and the maximum response value decreases, while the *APCE* value will also decrease. Furthermore, the maximum peak position may not be the target position. Therefore, the tracking results are considered reliable when both the maximum peak and *APCE* values are above a certain ratio of the respective historical averages, as shown in Figure 4.

$$\begin{cases} F_{t,max} \geq \beta_1 \frac{1}{t-1} \sum_{i=1}^{t-1} F_{i,max} \\ APCE_t \geq \beta_2 \frac{1}{t-1} \sum_{i=1}^{t-1} APCE_i \end{cases} \tag{19}$$

where $F_{t,max}$ and $APCE_t$ denote the maximum peak value and average *APCE* value of the output response map at frame t, respectively. $APCE = \frac{|F_{max} - F_{min}|^2}{\text{mean}\left(\sum_{w,h}\left(F_{w,h} - F_{min}\right)^2\right)}$, F_{max} and F_{min} indicate the maximum and minimum value of the response map, respectively.

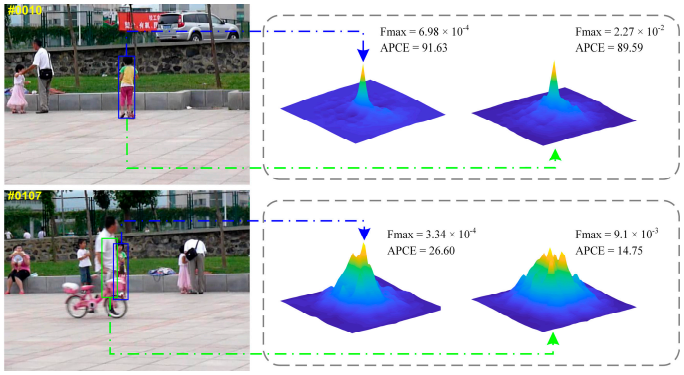

Figure 3. Comparison between the baseline BACF tracker and the proposed AMRCF. Blue and green boxes indicate the tracking results of the proposed and baseline, respectively. Since the tracked object is occluded by the background, the target appearance changes drastically. The response map also presents multiple peaks, and the *APCE* value and the maximum peak are extremely reduced.

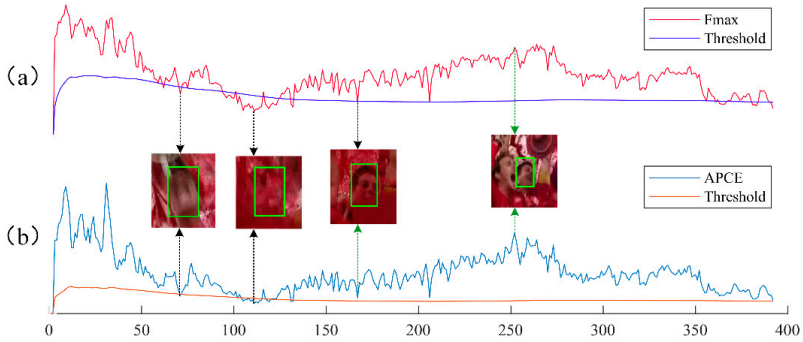

Figure 4. Figure (**a**) represents the maximum value with its threshold variation curve, and Figure (**b**) represents the *APCE* value with its threshold variation curve. The target appearance is significantly clearer when one of the *APCE* or maximum values are above the threshold (as the target pointed by the green arrow). Additionally, when both are above the threshold, we consider the target results reliable.

Depending on the object state, we selected different learning rates to maintain the appearance model. Considering that the target may be in an unreliable state for a long time, in this case, the learning rate should be further reduced. The learning rate is determined as follows:

$$lr_{new} = \begin{cases} \alpha - \beta * lr_{old} & \text{Reliable state} \\ 0.012 & \text{Unreliable state} \\ 0.003 & * \end{cases} \tag{20}$$

where α and β are set to 0.0175 and 0.1, respectively, as empirical constants. $*$ indicates the condition of "three consecutive frames judged to be unreliable state". As the target state changes from unreliable to reliable, the learning rate forms a small pulse and quickly stabilizes to $\frac{\alpha}{1+\beta}$.

The appearance model $\hat{x}_c{}^{model}$ updates as follows:

$$\hat{x}_{c+1}^{model} = (1 - lr_{new})\hat{x}_c{}^{model} + lr_{new}\hat{x}_c \tag{21}$$

where $\hat{x}_c^{model}$ and $\hat{x}_{c+1}^{model}$ represent the appearance model of the current frame and the next frame, respectively.

The overall workflow of the proposed AMRCF algorithm is as follows (Algorithm 1):

Algorithm 1: Augmented memory joint aberrance repressed correlation filter (AMRCF)

Input: First frame state of the sequence (i.e., target position and scale information);
Output: Target position at frame t;
Initialize tracker model hyperparameters.
for t = 1: end **do**
Training
Extract multi-channel feature maps x_t^d
Calculate the hash matrix p_t
if $t = 1$ **then**
Initialize the FIFO memory sequence.
Initialize the channel weight model $C_t^d = 1/D$
Initialize the appearance model.
else
Calculate the *score* between p_t and p_{t-1}.
if *score* $> \tau$ **then**
Update the FIFO memory queue.
end if
Store the hash matrix p_t.
end if
Optimize the filter model h_t via Equations (10) and (14)–(16) for L iterations
Detecting
Crop multi-scale search regions centered at P_t with S scales based on the bounding box at frame $t + 1$.
Extract multi-channel feature maps x_{t+1}^d
Use Equation (18) to final response output map R_r, $(r = 1, 2, \ldots, S)$.
Estimate the target position T_{t+1} and scale s_{t+1} from the maximum value of the response maps.
Updating
Use Equation (17) to update the channel weight model C_{t+1}
Use Equation (20) to Calculate the learning rate l_{new}.
Use Equation (21) to update the appearance model.
end for

4. Experiments

In this section, we adopt the one-pass evaluation (OPE) criterion on four widely used benchmark datasets (OTB50 [43], OTB100 [44], Temple-Color 128 [45] and UAV123 [46], respectively, with 269 challenge image sequences over 220k images) to evaluate the proposed algorithm and several SOTA trackers, including BACF [15], SRDCF [13], AutoTrack [47],

STRCF [16], ARCF [19], ARCF_H [19], MUSTER [22], BiCF [40], fDSST [6], Staple [9], LADCF_HC [35], AMCF [34], and ECO_HC [25]. The evaluation criteria include the overlap success rate (SR) and distance precision (DP), while the tracking speed is measured in terms of frames per second (FPS).

4.1. Implementation Details

The experiments involved were conducted in the same configuration (CPU Intel i7-7700 3.60 GHz 8.00 GB RAM and GPU NVIDIA GT730). For the reference tracker, the model structure and parameters were obtained from publicly available sources without any modifications. For the proposed AMRCF, the parameters are set as follows: The adaptive spatial regularization coefficients are $\lambda_1 = 0.2$, $\lambda_2 = 0.001$, the augmented memory regularization coefficient $\lambda_3 = 0.05$, the aberrance repression regularization coefficient $\lambda_4 = 0.233$, the learning rate of the channel weight model $\gamma = 0.018$, the initial appearance model learning rate is $lr_{init} = 0.013$, the length of the memory queue $K = 5$, the selection threshold τ for the memory sequence is set to 0.5, and the reliability judgement conditions of β_1 and β_2 are 0.3 and 0.6, respectively. We predict target placement using HOG, CN and Grayscale features, with scale evaluation using only five scales of HOG features.

4.2. Evaluation of the OTB Benchmark

4.2.1. Overall Performance

The OTB benchmark dataset contains 50 and 100 annotated video sequences with 11 different challenge attributes, i.e., motion blur (MB), illumination variation (IV), out-plane rotation (OPR), in-plane rotation (IPR), scale variation SV, deformation (DEF), fast motion (FM), out-of-view (OV), occlusion (OCC), background clutter (BC) and low resolution (LR). The overall results are presented in Figures 5 and 6. It can be seen that the performance of the proposed tracker is fully comparable to most good trackers. The proposed AMRCF gained 88.9% and 84.5% scores for the success rate and DP on the OTB50 benchmark while gained 86.2% and 81.8% scores on the OTB100 benchmark. The proposed AMRCF increased the success rate and DP by 5% and 3.7%, respectively, for OTB50; and by 4.6% and 5% for OTB100 compared to the baseline tracker BACF; while the proposed AMRCF increased the overlap success rate and DP results on the OTB100 benchmark dataset by 8.6% and 9.9%, respectively, against the spatial regularization-based SRDCF tracker; by 9.6% and 9.7%, respectively, against the augmented memory-based AMCF tracker; and by 5.7% and 7.1%, respectively, against the aberrance repression-based ARCF tracker.

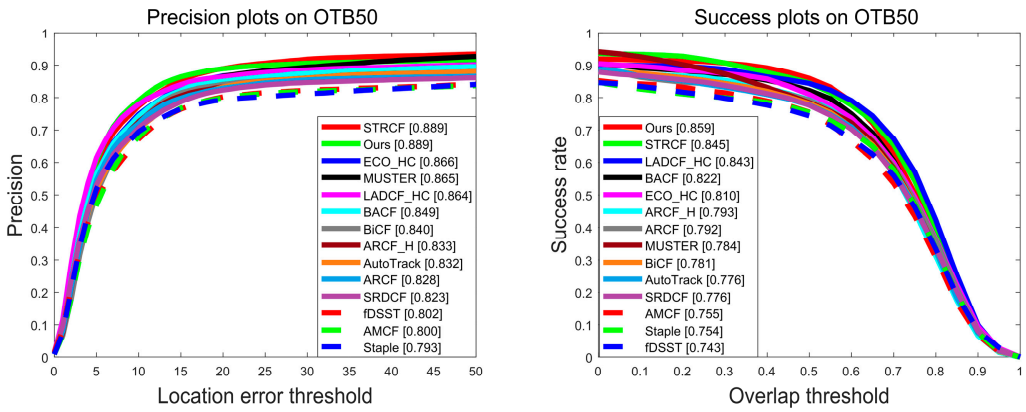

Figure 5. Evaluation results in terms of precision and success plots on OTB50 benchmark.

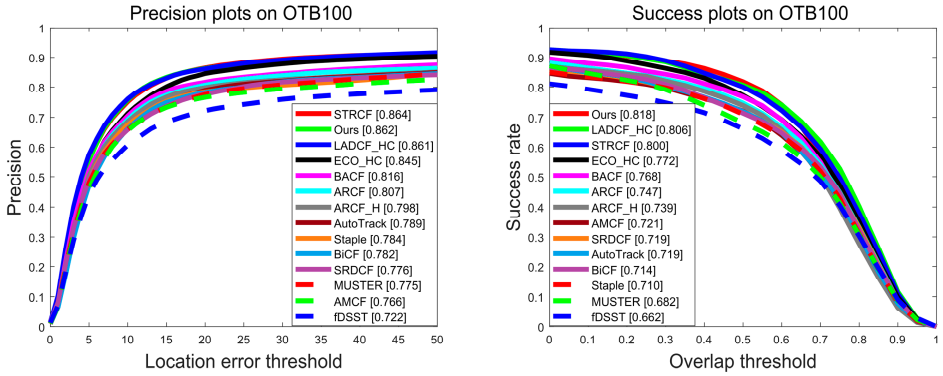

Figure 6. Evaluation results in terms of precision and success plots on OTB100 benchmark.

The performance of the mainstream trackers on the OTB100 benchmark in terms of accuracy and tracking speed in Table 1. The proposed AMRCF achieves excellent performance in terms of success rate and accuracy. Although the proposed algorithm greatly improves the accuracy of the basic tracker, it sacrifices the tracking speed due to the large number of additional operations added to the baseline BACF.

Table 1. Accuracy and tracking speed performance of the top-8 trackers on the OTB100 benchmark.

Trackers	STRCF	ECO_HC	LADCF_HC	BACF	ARCF	AMCF	SRDCF	Ours
DP (%)	86.4	84.5	86.1	81.6	80.7	76.6	77.6	86.2
SR (%)	80.0	77.2	80.6	76.8	74.7	72.1	71.9	81.8
FPS	25.57	51.06	20.23	39.49	17.07	34.44	7.22	12.56

Note: The top and second ranked outcomes are shown with red and green.

4.2.2. Attribute Evaluation

To further demonstrate the tracking performance of the proposed AMRCF in the complex real-world environment, Figure 7 shows the attribute-based experimental results on the OTB100 benchmark. It is clear that most of the attribute-based results of our proposed algorithm obtained the best performance compared to the SOTA trackers, especially in the background clutter, motion blur, fast motion, where the success rate and distance precision reached 85.7% and 85.9%, 80.9% and 82%, 77.5% and 82%, all ranking the best results, while in the out-of-view, fast motion, illumination variation, in-plane-rotation also obtained excellent performance.

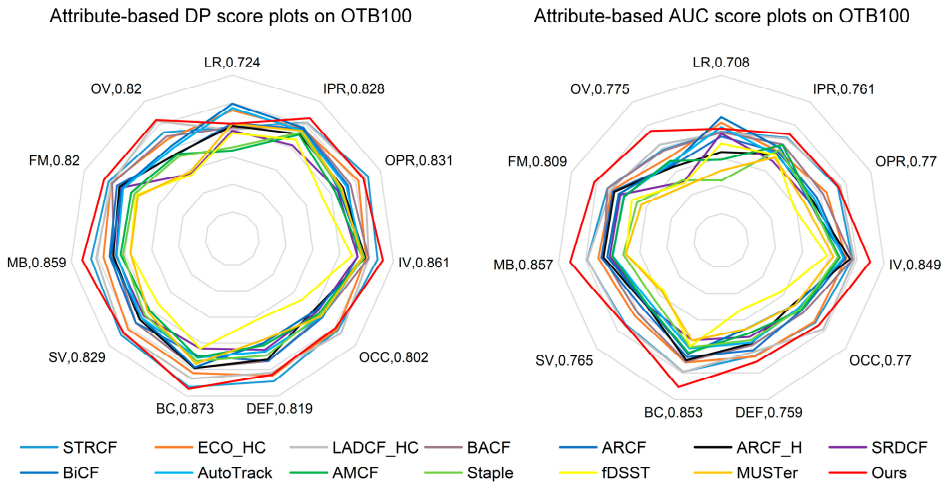

Figure 7. The attribute-based experimental results for different SOTA trackers on the OTB100 benchmark. The score adjoining the attribute-name indicates the AMRCF result.

4.2.3. Qualitative Evaluation

To further demonstrate the superiority of the proposed AMRCF, Figure 8 shows the results of a qualitative comparison between the AMRCF and seven SOTA trackers, including STRCF, LADCF_HC, ECO_HC, ARCF, SRDCF, BACF and AMCF. The selected test sequences are Soccer, Dragonbaby, Box, Bolt 2, Girl 2 and Basketball (from top to bottom), where each sequence contains at least multiple challenging attributes, including OCC (Soccer, Dragonbaby, Box, Girl 2 and Basketball), BC (Soccer, Box, Bolt 2 and Basketball), FM (Soccer and Dragonbaby), MB (Soccer, Dragonbaby, Box and Girl 2), DEF (Bolt 2, Girl 2 and Basketball), etc. The results show that the proposed AMRCF was able to perform the tracking task relatively well in a variety of scenarios. Significantly, the proposed tracker demonstrates its superior performance in the "Soccer" sequence with a cluttered background or in the "Girl2" and "Box" sequences which are partially or completely obscured. In the "Soccer" sequence, despite massive background interference, deformation and occlusion problems in frames 100–200, AMRCF was still able to accomplish the subsequent tracking task while most trackers lost the target. In frames 105–120 of the "Girl 2" sequence, the target is lost due to full occlusion. However, when the target reappears in the 120-frame view, AMRCF is still able to redetect and achieve stable tracking. Comparing the performance of the baseline BACF tracker, this further validates that the improvements in model structure and high-confidence updating strategy of the proposed tracker are not redundant.

Figure 8. Qualitative evaluation. Top performing trackers compared with the proposed AMRCF on the OTB100 benchmark (from top to bottom, these refer to Soccer, Dragonbaby, Box, Bolt 2, Girl 2 and Basketball, respectively).

4.3. Evaluation of the TC128 Benchmark

In order to approximate real tracking scenes, unlike the OTB dataset which includes 25% grey scale sequences, the Temple Color benchmark collects 128 colour videos containing 27 object categories. An overview of the experimental results for the TC128 benchmark is presented in Figure 9. The proposed AMRCF achieves excellent tracking performance with the success rate and DP of 69.7% and 75.8%, respectively. Comparing the experimental results of the baseline tracker BACF (whose success rate and DP are 61.3% and 65.3%, respectively), the proposed AMRCF outperforms the performance by 8.4% and 11.2%, respectively. In comparing with the SRDCF, ARCF and AMCF tracker, the performance of the proposed method gains 9.5%, 5.6% and 9.2% in DP score.

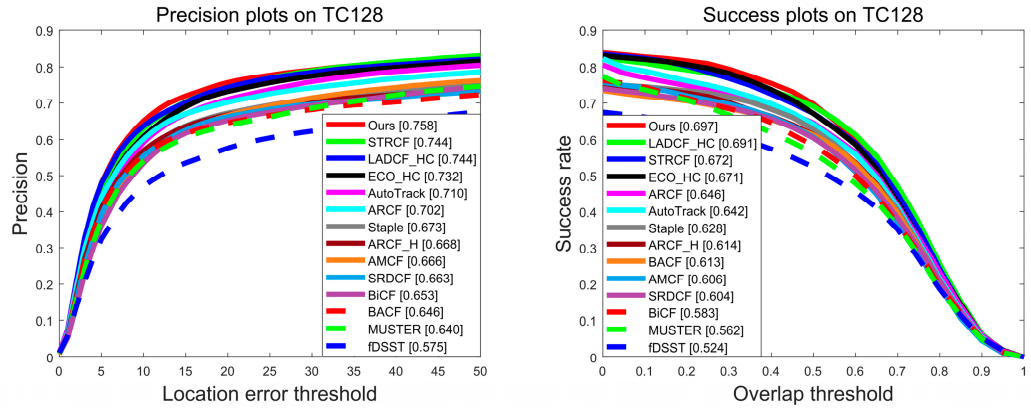

Figure 9. Evaluation results in terms of precision and success plots on TC128 benchmark.

4.4. Evaluation of the UAV123 Benchmark

The UAV123 benchmark consists of 91 image sequences captured by low-altitude UAVs. In contrast with the OTB100, the UAV123 has the challenges of a wide variation of shooting angles, very long video sequences, small targets and targets that disappear completely from the view for an extended time, which poses an even greater challenge for the tracker. On the one hand, the tracker model needs to be robust and able to follow the target again after complete occlusion, and on the other hand, the model needs to be quickly updated to accommodate rapid changes in shooting angles. An overview of the experimental results for the UAV123 benchmark is presented in Figure 10. As it can be seen, the proposed AMRCF is ranked first in the success rate (58.4%) and third in DP (69%), respectively. Compared to the baseline tracker, the proposed tracker is outperformed by 2.9% and 3% in terms of success rate and DP.

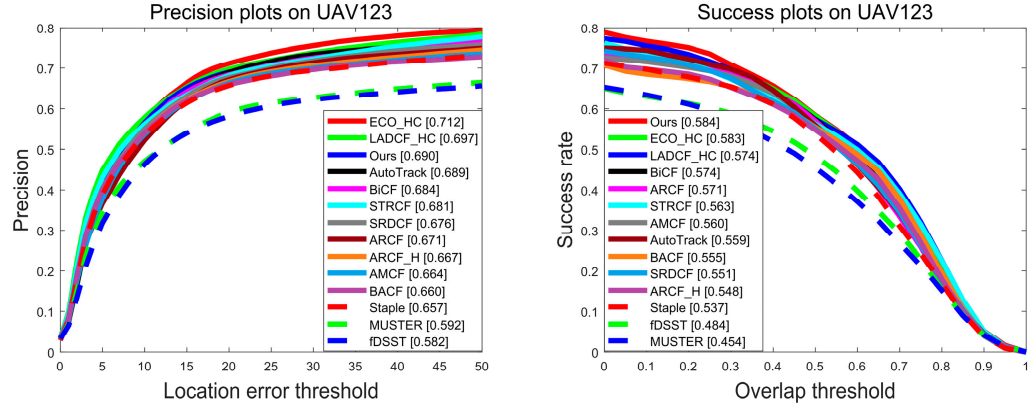

Figure 10. Evaluation results in terms of precision and success plots on the UAV123 benchmark.

Figure 11 shows the success plots for nine representative attributes on UAV123 benchmark, including Camera Motion (CM), Scale Variation (SV), Illumination Variation (IV),

Aspect Ratio Change (ARC), Fast Motion (FM), Viewpoint Change (VC), Similar Object (SO), Partial Occlusion (PO) and Out-of-View (OV). Compared with other SOTA trackers, the proposed AMRCF performs remarkably well in these challenging attributes. Four of the nine attribute results ranked first, including CM (58.1%), SV (54%), ARC (48.2%), FM (43.9%), and the rest maintained the top three positions, slightly worse than the first, which proved that AMRCF maintained a high tracking performance for dealing with various tracking scenarios. This can also be attributed to the synergy of multiple regularization constraints and the dynamic model update strategy, which greatly mitigates the boundary effects, response distortions and filter degradation while providing a robust appearance model.

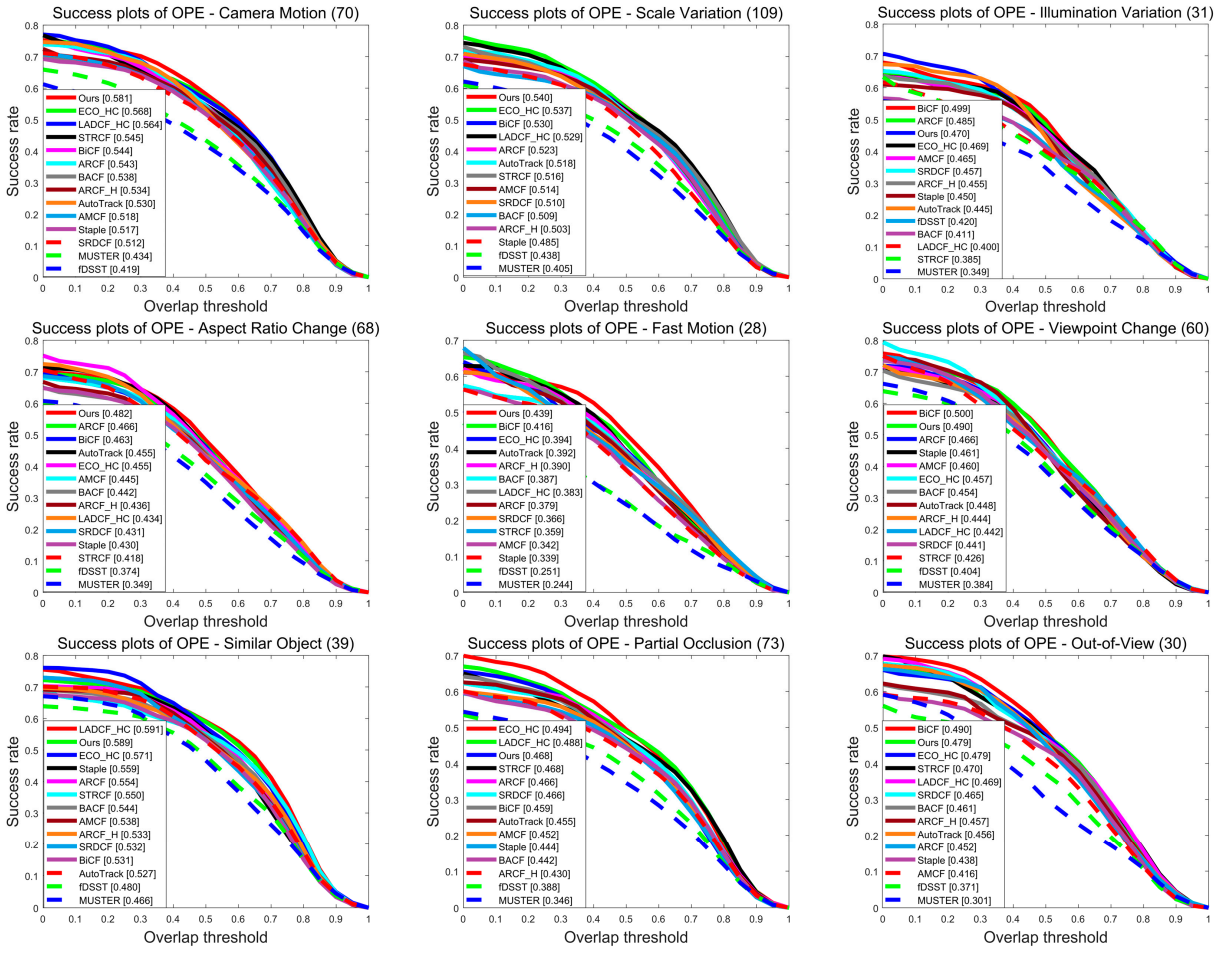

Figure 11. Attribute-based evaluation results in terms of success plots on UAV123 benchmark.

4.5. Compared with Deep Feature-Based Trackers

In this section, we discuss the effectiveness of the proposed AMRCF by comparing the deep feature-based tracking algorithms, including LADCF [35], DeepSTRCF [16], SRDCFdecon [48], SiamFC3s [49], SiamFC [49], DCFNet [50] and CFNet [51]. From Table 2, we can see that our algorithm ranks third in terms of success rate and first in terms of precision on the OTB50 benchmark. In terms of tracking speed, however, we have a relatively obvious advantage over deep feature-based algorithms. The attribute-based experimental results, presented in Table 3, show that the proposed AMRCF achieves the excellent performance in precision with other deep feature-based trackers under IV, OPR, DEF, MB, FM, IPR, OV and BC challenge attributes, which further demonstrate the advantages of our tracker architecture.

Table 2. Accuracy and tracking speed performance of the deep learning trackers on the OTB50 benchmark.

Trackers	LADCF	DeepSTRCF	SRDCFdecon	SiamFC3s	SiamFC	DCFNet	CFNet	Ours
DP (%)	87.2	86.6	81.4	77.9	72.0	84.2	72.4	85.9
SR (%)	88.4	87.3	87.0	80.9	77.2	87.7	78.1	88.9
FPS	5.93	3.57	2.41	9.82	12.82	99.37	5.7	13.13

Note: The first, second and third ranked outcomes are shown in red, green and blue.

Table 3. Precision scores of AMRCF and deep feature-based trackers on the OTB50 benchmark.

	IV	OPR	SV	OCC	DEF	MB	FM	IPR	OV	BC	LR
LADCF	79.0	87.6	84.9	**89.8**	87.3	74.8	79.2	81.5	79.4	81.4	55.4
DeepSTRCF	78.0	86.3	82.7	88.4	86.9	74.6	75.9	79.8	81.1	79.9	56.9
SRDCFdecon	81.8	85.8	83.3	86.1	84.2	79.0	78.3	80.6	78.5	84.6	52.7
SiamFC3s	70.9	78.8	79.6	80.2	74.3	69.8	72.3	74.3	78.0	73.2	**65.9**
SiamFC	67.6	74.7	77.3	68.0	70.1	52.5	66.9	71.7	64.3	72.3	48.9
DCFNet	82.7	86.7	**88.3**	87.0	83.4	74.3	77.4	83.9	80.3	82.8	64.2
CFNet	70.3	79.3	78.6	77.4	73.2	56.7	56.8	73.5	57.1	70.9	57.9
Ours	**84.5**	**88.5**	84.4	88.0	**88.6**	**81.4**	**79.3**	**85.6**	**81.6**	**86.6**	56.3

Note: The best results are shown in bold.

4.6. Ablation Studies and Effectiveness Discussion

In this section, we discuss the effectiveness of the different modules in the proposed the AMRCF and compare them with the baseline BACF tracker. Moreover, AMRCF-AS adds the adaptive spatial regularization module (AS) to the baseline BACF tracker. However, AMRCF-AM, AMRCF-AR and AMRCF-CL add the augmented memory module (AM), aberrance repression module (AR) and channel weight with learning rate update module (CL) to the BACF, respectively. AMRCF-AS-AM and AMRCF-AS-AM-CL are obtained by adding AM module and CL module to AMRCF-AS in turn. The AMRCF method is our complete tracking framework, which incorporates all modules.

The overall ablation experimental results are presented in Table 4. It can be seen that the original baseline BACF tracker achieves scores of 66.2% and 73.8% in the average success and precision. Benefiting from the AS module, AMRCF-AS outperforms the baseline in terms of the average accuracy and precision by 2.1% and 1.9%, respectively. Meanwhile, AMRCF-AM also outperforms the baseline in terms of the average accuracy and precision by 2.6% and 2.4%, respectively, which indicates that the selective recollection of history views is effective for the overall tracker performance. For the average accuracy and precision, AMRCF-AR increased by 3.2% and 3.1% while AMRCF-CL increased by 3.5% and 4.1%. The above experimental results all indicate the effectiveness of each module for the tracking framework. Furthermore, the performance of the baseline BACF improves with the introduction of key modules. Ultimately, the average success and precision scores of the proposed AMRCF, which incorporates all modules, exceeded the baseline tracker by 3.9% and 3.8%, respectively.

Table 4. Overall evaluation performance on the OTB100 and UAV123 benchmarks with the progressive addition of different modules on the baseline BACF tracker.

Trackers	OTB100 Benchmark		UAV123 Benchmark		Average	
	Success (%)	Precision (%)	Success (%)	Precision (%)	Success (%)	Precision (%)
BACF (Baseline)	76.8	81.6	55.5	66.0	66.2	73.8
AMRCF-AS	79.3	83.4	57.3	68.0	68.3	75.7
AMRCF-AM	79.4	84.0	58.1	68.3	68.8	76.2

Table 4. *Cont.*

Trackers	OTB100 Benchmark		UAV123 Benchmark		Average	
	Success (%)	Precision (%)	Success (%)	Precision (%)	Success (%)	Precision (%)
AMRCF-CL	81.4	85.7	58.0	**70.1**	69.7	**77.9**
AMRCF-AR	81.5	85.4	57.3	68.3	69.4	76.9
AMRCF-AS-AM	80.7	85.5	58.3	68.6	69.5	77.1
AMRCF-AS-AM-CL	81.6	85.9	58.2	69.3	69.9	77.6
AMRCF	**81.8**	**86.2**	**58.4**	69.0	**70.1**	77.6

Note: The best results are shown in bold.

5. Conclusions

In this paper, based on the background-aware correlation filter, we proposed a novel augmented memory joint aberrance repressed correlation filter (AMRCF) for visual tracking. By introducing different regularizers (adaptive spatial regularization, augmented memory regularization and aberrance suppression regularization) and combining with a high-confidence updating strategy, the adverse effects caused by boundary effects, model degradation and response anomalies are effectively mitigated, making the trained model adaptable to changing tracking scenarios. The ADMM algorithm is employed to reduce the computational complexity during model optimization. In addition, extensive comparative experiments on four well-known benchmarks indicate that the proposed AMRCF tracker achieved a tracking performance comparable to 14 SOTA trackers, especially in environments such as background clutter, motion blur, out-of-view and fast motion. However, as the proposed tracker cannot achieve real time in terms of tracking speed, subsequent work will be carried out on increasing the speed of the tracker.

Author Contributions: Conceptualization, Y.J. and J.H.; methodology, X.S.; software, J.H.; validation, Y.J., J.H. and Y.B.; formal analysis, X.S.; investigation, Y.B.; resources, Y.B.; data curation, Z.W.; writing—original draft preparation, J.H.; writing—review and editing, K.H.b.G.; visualization, K.H.b.G.; supervision, X.S.; project administration, X.S.; funding acquisition, Y.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under 62061010 and 62161007, in part by the Guangxi Science and Technology Department Project under AA19254029, AA20302022 and AB21196041, in part by the Guangxi Natural Science Foundation of China under 2019GXNSFBA245072, in part by Nanning City Qingxiu District Science and Technology Major Special Project under 2018001 and in part by Innovation Project of GUET Graduate Education under 2022YCXS037.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bolme, D.S.; Beveridge, J.R.; Draper, B.A.; Lui, Y.M. Visual object tracking using adaptive correlation filters. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2544–2550.
2. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. Exploiting the circulant structure of tracking-by-detection with kernels. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; pp. 702–715.
3. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 583–596. [CrossRef] [PubMed]
4. Li, Y.; Zhu, J.; Hoi, S.C.; Song, W.; Wang, Z.; Liu, H. Robust estimation of similarity transformation for visual object tracking. *Proc. AAAI Conf. Artif. Intell.* **2019**, *33*, 8666–8673. [CrossRef]
5. Huang, D.; Luo, L.; Wen, M.; Chen, Z.; Zhang, C. Enable scale and aspect ratio adaptability in visual tracking with detection proposals. In Proceedings of the British Machine Vision Conference, Swansea, UK, 7–10 September 2015.

6. Danelljan, M.; Häger, G.; Khan, F.S.; Felsberg, M. Discriminative scale space tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1561–1575. [CrossRef] [PubMed]
7. Danelljan, M.; Shahbaz Khan, F.; Felsberg, M.; Van de Weijer, J. Adaptive color attributes for real-time visual tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1090–1097.
8. Ma, C.; Huang, J.; Yang, X.; Yang, M. Hierarchical convolutional features for visual tracking. In Proceedings of the IEEE International Conference on Computer Vision, Washington, DC, USA, 7–13 December 2015; pp. 3074–3082.
9. Bertinetto, L.; Valmadre, J.; Golodetz, S.; Miksik, O.; Torr, P.H. Staple: Complementary learners for real-time tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1401–1409.
10. Lukežič, A.; Zajc, L.Č.; Kristan, M. Deformable parts correlation filters for robust visual tracking. *IEEE Trans. Cybern.* **2017**, *48*, 1849–1861. [CrossRef]
11. Sun, X.; Cheung, N.; Yao, H.; Guo, Y. Non-rigid object tracking via deformable patches using shape-preserved KCF and level sets. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5495–5503.
12. Liu, T.; Wang, G.; Yang, Q. Real-time part-based visual tracking via adaptive correlation filters. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4902–4912.
13. Danelljan, M.; Hager, G.; Shahbaz Khan, F.; Felsberg, M. Learning spatially regularized correlation filters for visual tracking. In Proceedings of the IEEE international Conference on Computer Vision, Washington, DC, USA, 7–13 December 2015; pp. 4310–4318.
14. Lukezic, A.; Vojir, T.; Cehovin Zajc, L.; Matas, J.; Kristan, M. Discriminative correlation filter with channel and spatial reliability. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6309–6318.
15. Kiani Galoogahi, H.; Fagg, A.; Lucey, S. Learning background-aware correlation filters for visual tracking. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1135–1143.
16. Li, F.; Tian, C.; Zuo, W.; Zhang, L.; Yang, M. Learning spatial-temporal regularized correlation filters for visual tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4904–4913.
17. Gu, P.; Liu, P.; Deng, J.; Chen, Z. Learning Spatial–Temporal Background-Aware Based Tracking. *Appl. Sci.* **2021**, *11*, 8427. [CrossRef]
18. Zhang, W.; Kang, B.; Zhang, S. Spatial-temporal aware long-term object tracking. *IEEE Access* **2020**, *8*, 71662–71684. [CrossRef]
19. Huang, Z.; Fu, C.; Li, Y.; Lin, F.; Lu, P. Learning aberrance repressed correlation filters for real-time UAV tracking. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 2891–2900.
20. Fu, C.; Huang, Z.; Li, Y.; Duan, R.; Lu, P. Boundary effect-aware visual tracking for UAV with online enhanced background learning and multi-frame consensus verification. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; pp. 4415–4422.
21. Gan, L.; Ma, Y. Long-term correlation filter tracking algorithm based on adaptive feature fusion. In Proceedings of the Thirteenth International Conference on Graphics and Image Processing (ICGIP 2021), Kunming, China, 18–20 August 2021; Volume 12083, pp. 253–263.
22. Hong, Z.; Chen, Z.; Wang, C.; Mei, X.; Prokhorov, D.; Tao, D. Multi-store tracker (muster): A cognitive psychology inspired approach to object tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 749–758.
23. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **2011**, *3*, 1–122. [CrossRef]
24. Danelljan, M.; Robinson, A.; Shahbaz Khan, F.; Felsberg, M. Beyond correlation filters: Learning continuous convolution operators for visual tracking. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 472–488.
25. Danelljan, M.; Bhat, G.; Shahbaz Khan, F.; Felsberg, M. Eco: Efficient convolution operators for tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6638–6646.
26. Bhat, G.; Johnander, J.; Danelljan, M.; Khan, F.S.; Felsberg, M. Unveiling the power of deep tracking. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 483–498.
27. Xu, L.; Gao, M.; Li, X.; Zhai, W.; Yu, M.; Li, Z. Joint spatiotemporal regularization and scale-aware correlation filters for visual tracking. *J. Electron. Imaging* **2021**, *30*, 043011. [CrossRef]
28. Li, T.; Ding, F.; Yang, W. UAV object tracking by background cues and aberrances response suppression mechanism. *Neural Comput. Appl.* **2021**, *33*, 3347–3361. [CrossRef]
29. Huang, B.; Xu, T.; Shen, Z.; Jiang, S.; Li, J. BSCF: Learning background suppressed correlation filter tracker for wireless multimedia sensor networks. *Ad Hoc Netw.* **2021**, *111*, 102340. [CrossRef]
30. Elayaperumal, D.; Joo, Y.H. Aberrance suppressed spatio-temporal correlation filters for visual object tracking. *Pattern Recognit.* **2021**, *115*, 107922. [CrossRef]

31. Dai, K.; Wang, D.; Lu, H.; Sun, C.; Li, J. Visual tracking via adaptive spatially-regularized correlation filters. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4670–4679.
32. Mueller, M.; Smith, N.; Ghanem, B. Context-aware correlation filter tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1396–1404.
33. Zha, Y.; Zhang, P.; Pu, L.; Zhang, L. Semantic-aware spatial regularization correlation filter for visual tracking. *IET Comput. Vis.* **2022**, *16*, 317–332. [CrossRef]
34. Li, Y.; Fu, C.; Ding, F.; Huang, Z.; Pan, J. Augmented memory for correlation filters in real-time UAV tracking. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; pp. 1559–1566.
35. Xu, T.; Feng, Z.; Wu, X.; Kittler, J. Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking. *IEEE Trans. Image Processing* **2019**, *28*, 5596–5609. [CrossRef] [PubMed]
36. Yu, Y.; Chen, L.; He, H.; Liu, J.; Zhang, W.; Xu, G. Second-Order Spatial-Temporal Correlation Filters for Visual Tracking. *Mathematics* **2022**, *10*, 684. [CrossRef]
37. Hu, Z.; Zou, M.; Chen, C.; Wu, Q. Tracking via context-aware regression correlation filter with a spatial–temporal regularization. *J. Electron. Imaging* **2020**, *29*, 023029. [CrossRef]
38. Xu, L.; Kim, P.; Wang, M.; Pan, J.; Yang, X.; Gao, M. Spatio-temporal joint aberrance suppressed correlation filter for visual tracking. *Complex Intell. Syst.* **2021**, 1–13. [CrossRef]
39. Wang, M.; Liu, Y.; Huang, Z. Large margin object tracking with circulant feature maps. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4021–4029.
40. Lin, F.; Fu, C.; He, Y.; Guo, F.; Tang, Q. BiCF: Learning bidirectional incongruity-aware correlation filter for efficient UAV object tracking. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 2365–2371.
41. Liu, L.; Feng, T.; Fu, Y. Learning Multifeature Correlation Filter and Saliency Redetection for Long-Term Object Tracking. *Symmetry* **2022**, *14*, 911. [CrossRef]
42. Kozat, S.S.; Venkatesan, R.; Mihçak, M.K. Robust perceptual image hashing via matrix invariants. In Proceedings of the 2004 International Conference on Image Processing, 2004. ICIP'04., Singapore, 24–27 October 2004; pp. 3443–3446.
43. Wu, Y.; Lim, J.; Yang, M. Online object tracking: A benchmark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; Volume 5, pp. 2411–2418.
44. Wu, Y.; Lim, J.; Yang, M.H. Object Tracking Benchmark. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1834–1848. [CrossRef] [PubMed]
45. Liang, P.; Blasch, E.; Ling, H. Encoding color information for visual tracking: Algorithms and benchmark. *IEEE Trans. Image Processing* **2015**, *24*, 5630–5644. [CrossRef] [PubMed]
46. Mueller, M.; Smith, N.; Ghanem, B. A benchmark and simulator for uav tracking. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 445–461.
47. Li, Y.; Fu, C.; Ding, F.; Huang, Z.; Lu, G. 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, Seattle, WA, USA, 13–19 June 2020; pp. 11923–11932.
48. Danelljan, M.; Hager, G.; Shahbaz Khan, F.; Felsberg, M. Adaptive decontamination of the training set: A unified formulation for discriminative visual tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1430–1438.
49. Bertinetto, L.; Valmadre, J.; Henriques, J.F.; Vedaldi, A.; Torr, P.H. Fully-convolutional siamese networks for object tracking. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 850–865.
50. Wang, Q.; Gao, J.; Xing, J.; Zhang, M.; Hu, W. Dcfnet: Discriminant correlation filters network for visual tracking. *arXiv* **2017**, arXiv:1704.04057.
51. Valmadre, J.; Bertinetto, L.; Henriques, J.; Vedaldi, A.; Torr, P.H. End-to-end representation learning for correlation filter based tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2805–2813.