frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 10 January 2023
DOI 10.3389/fnins.2022.1080521

OPEN ACCESS

EDITED BY
Yuqi Han,
Tsinghua University, China

REVIEWED BY
Tiancheng Dong,
Wuhan University, China
Wendong Zheng,
Tsinghua University, China

*CORRESPONDENCE
Liangyu Zhao
zhaoly@bit.edu.cn

SPECIALTY SECTION
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

RECEIVED 26 October 2022
ACCEPTED 22 November 2022
PUBLISHED 10 January 2023

CITATION
Cui J, Wu J and Zhao L (2023) Learning channel-selective and aberrance repressed correlation filter with memory model for unmanned aerial vehicle object tracking.
Front. Neurosci. 16:1080521.
doi: 10.3389/fnins.2022.1080521

Learning channel-selective and aberrance repressed correlation filter with memory model for unmanned aerial vehicle object tracking

Jianjie Cui[1], Jingwei Wu[2] and Liangyu Zhao[1]*

[1]School of Aerospace Engineering, Beijing Institute of Technology, Beijing, China, [2]The Second Academy of CASIC, Beijing, China

To ensure that computers can accomplish specific tasks intelligently and autonomously, it is common to introduce more knowledge into artificial intelligence (AI) technology as prior information, by imitating the structure and mindset of the human brain. Currently, unmanned aerial vehicle (UAV) tracking plays an important role in military and civilian fields. However, robust and accurate UAV tracking remains a demanding task, due to limited computing capability, unanticipated object appearance variations, and a volatile environment. In this paper, inspired by the memory mechanism and cognitive process in the human brain, and considering the computing resources of the platform, a novel tracking method based on Discriminative Correlation Filter (DCF) based trackers and memory model is proposed, by introducing dynamic feature-channel weight and aberrance repressed regularization into the loss function, and by adding an additional historical model retrieval module. Specifically, the feature-channel weight integrated into the spatial regularization (SR) enables the filter to select features. The aberrance repressed regularization provides potential interference information to the tracker and is advantageous in suppressing the aberrances caused by both background clutter and appearance changes of the target. By optimizing the aforementioned two jointly, the proposed tracker could restrain the potential distractors, and train a robust filter simultaneously by focusing on more reliable features. Furthermore, the overall loss function could be optimized with the Alternative Direction Method of Multipliers (ADMM) method, thereby improving the calculation efficiency of the algorithm. Meanwhile, with the historical model retrieval module, the tracker is encouraged to adopt some historical models of past video frames to update the tracker, and it is also incentivized to make full use of the historical information to construct a more reliable target appearance representation. By evaluating the method on two challenging UAV benchmarks, the results prove that this tracker shows superior performance compared with most other advanced tracking algorithms.

KEYWORDS

unmanned aerial vehicle, object tracking, discriminative correlation filter, channel regularization, aberrance repressed, historical memory

1. Introduction

The thinking ability endowed by the brain is fundamental. Due to its existence, human beings are more intelligent than animals. It is also the premise that humans have the capability to conduct scientific research (Kuroda et al., 2022). People rely on their brains to recognize the world, learn knowledge, and summarize rules. Their brains also allow them to use memory systems to store the information generated when experiencing different events (Atkinson and Shiffrin, 1968; Cornelio et al., 2022). In turn, the information serves as prior knowledge, helping people in dealing with similar problems better and adapting to new complex scenes faster. The core of artificial intelligence (AI) is to enable the machine to complete specific tasks independently through learning and using prior information (Connor et al., 2022; Foksinska et al., 2022; Nofallah et al., 2022; Pfeifer et al., 2022; Wang et al., 2022). The original scientific research mainly adopts the following two methods. The first is to mathematically represent the law (i.e. prior information) that people summarize when perceiving things, and then use mathematical expressions and logical frameworks to construct modules and methods for computers, just like teachers teach students what they know. The second is to build a variety of artificial neural networks based on the neural structure of the human brain and then use large-scale data to train and fit the network (Deng et al., 2022; Liu et al., 2022), aiming to enable the computer to automatically learn the characteristics of various things from the data itself, just like the students read books and learn by themselves. Although scientists have put a lot of effort into the research and utilization of the human brain, it is still a difficult task to determine how to endow computers with more and better prior knowledge through algorithms.

This paper mainly concentrates on visual object tracking on the UAV platform, which plays an important role in the field of computer vision, and is widely used in many tasks, such as collision avoidance (Baca et al., 2018), traffic monitoring (Elloumi et al., 2018), military surveillance (Shao et al., 2019), and aerial cinematography (Gschwindt et al., 2019). By adopting this technology, it aims to predict the precise status of the target in a video sequence captured by an onboard camera only with the information given in the first frame (Han et al., 2022). Over the past few years, a lot of effort has been put into the tracking field. However, it is still a challenging task to design a robust and efficient tracker, when considering the various complex UAV tracking scenarios, e.g., occlusion, change of viewpoint, and limited power capacity.

In the past decade, the research on visual object tracking mainly adopted the two methods below, namely the discriminative correlation filter (DCF)-based method and the Siamese-based method. The Siamese-based method (Bertinetto et al., 2016b; Li et al., 2018a; Wang et al., 2019; Voigtlaender et al., 2020; Javed et al., 2022) aims at the offline learning of the similarity measurement function between image patches, by maximizing the distance between the target and the background patches while minimizing the distance between the different image patches belonging to the same target. Such a method consists of two identical subchannels that are used to process the target template and the current frame search area, respectively. The target location is determined by computing the partial similarity between the target template and each location in the search area. Moreover, the Siamese-based method uses neural network architecture and numerous training data to obtain excellent feature extraction capability, so it needs to occupy a large number of computing resources in the tracking process. DCF-based methods are based on the correlation theory in the field of signal processing, and it computes the correlation between different image patches by convolution. Such a method usually adopts the hand-crafted features carefully designed with prior information and aims at training a correlation filter online in the region around the target by minimizing a least squares loss. Due to the convolution theorem, DCF-based methods can track objects at hundreds of frames per second (FPS) with only one CPU. Considering that the computing resources of the UAV platform are very limited, and the speed is a key issue in addition to the tracking performance, this paper mainly concentrates on target tracking based on DCF methods.

The development history of the DCF-based method is the process by which people integrated more and better prior information into the tracking framework. As people add their understanding of tracking tasks as regular constraints to the loss function (Mueller et al., 2017; Han et al., 2019b), the trained correlation filter becomes more and more discriminative and robust. Mosse (Bolme et al., 2010), as the originator of correlation filtering, deemed target tracking as a problem of binary classification, and trained the filter by randomly sampling a fixed number of background samples as negative samples. This greatly limits its discriminative power. To effectively increase the number of training samples, which was critical to the performance of the trained classifier, KCF (Henriques et al., 2014) introduced the circulant matrix into the tracking framework and obtained a large number of virtual negative samples by circularly shifting the target samples. The cyclic shifting greatly increased the training samples and caused boundary effects that seriously limited the improvement of tracking performance simultaneously. To mitigate the boundary effect, SRDCF (Danelljan et al., 2015) added the SR term into the loss function, aiming at penalizing the non-zero value near the template boundaries. BACF (Kiani Galoogahi et al., 2017) generated lots of real background samples, by expanding the search area and introducing a binary mask for middle elements cropping. To solve the scale change of the target, DSST (Danelljan et al., 2014a) introduced an independent scale filter, in addition to the classical correlation filter used for locating, as well as SAMF (Li and Zhu, 2014) sampled multiscale images, thereby building image pyramids. For the improvement of the feature representation, CN (Danelljan et al.,

2014b) brought in color features, while ECO (Danelljan et al., 2017) added depth features obtained from off-line training of the neural network. STRCF (Li et al., 2018b) brought in additional temporal constraints to the SRDCF to limit the variation of the filter in consecutive frames. This effectively reduced the risk of filter degradation in case of sudden large appearance variations. SAT (Han et al., 2019a) advocated a kurtosis-based updating scheme to guarantee a high-confidence template updating. ASRCF (Dai et al., 2019) realized the adaptive suppression of clutter in different regions by regarding the SR term, introduced in SRDCF, as a variable. MUSTer (Hong et al., 2015) built the short-term and long-term memory stores, thereby processing the target appearance memories. Autotrack (Li et al., 2020) reformulated the loss function by introducing the change of response maps into temporal regularization (TR) and SR terms, thereby realizing adaptive adjustment. Regardless of the great progress in DCF-based tracking methods, there are still some issues to solve. (1) Most original trackers treat the features of different dimensions equally. Features of different dimensions play different roles in tracking different scenarios and different kinds of targets. The tracker is easily biased by similar interference due to ignorance of the feature channel information. (2) Most original trackers have insufficient ability to suppress potential interference. Most of the original methods merely utilize the same and fixed bowl-shaped SR term centered on the target, aiming at giving more weight to the background area for suppression. Additional suppression is not applied to the potential interference according to the actual tracking situation, thus leading to limited anti-aberrance capability. (3) Most original trackers do not effectively use historical information. Most of the original methods updated the filter with a constant update rate, thereby causing the waste of historical information and the risk of filter degradation. Historical information is one of the most important factors in the tracking process and should be efficiently used to enhance the discriminant capacity of the tracker.

The brain can perceive the interference information in the background, independently select the optimal features to describe the target, and use historical memory to achieve an accurate target location in the current frame. When considering the above, a UAV tracking algorithm with repressed dynamic aberrance, a channel selective correlation filter, and a historical model retrieval module is proposed to solve the aforementioned problems. Moreover, by formulating the dynamic feature channel weight and the aberrance repressed regularization into the integral loss function, the tracking algorithm is built, thereby enabling the filter to highlight valuable features in the channel domain and using response maps to sense and suppress background interference in advance. Meanwhile, the model retrieval module, by imitating brain memory realizes the adaptive update of the tracker. This paper has the main contributions as follows.

i) A novel tracking method, that integrates the aberrance repressed regularization and dynamic feature channel weight into the loss function of the DCF framework, is proposed. For joint modeling of the two factors, the tracker obtains the ability to screen target features based on actual background interference and learns more differentiated target appearance representation. Thus, the loss function could be solved in very few iterations by employing an efficient ADMM algorithm.

ii) A model retrieval module is employed which can realize the adaptive update of the tracker by saving the history filters. This module can also enhance the tracker's learning of the appearance of the trusted targets with historical information and reduce the pollution of unreliable samples for the tracker.

iii) By giving the experimental validation conducted on two public UAV datasets, the effectiveness of this method is demonstrated.

2. Proposed methodologies

2.1. Revisted autotrack

In this section, the baseline Autotrack of this tracker shall be revised.

Most original trackers, based on the discriminative correlation filters (DCF), attempt to add a variety of regularization terms such as spatial regularization (SR) and temporal regularization (TR), thereby improving the discrimination ability to target and background. Such regularization terms are usually predefined fixed parameters, so flexibility and adaptability are lacking in cluttered and challenging scenarios. To realize automatic adjustment of the hyper-parameters of the SR and TR terms during tracking, Autotrack constructs them with the response maps obtained during detection. Specifically, Autotrack introduces the partial response variation Λ to the SR parameter $\tilde{\boldsymbol{u}}$, and the global response variation $\|\boldsymbol{\Lambda}\|_2$ to the reference value $\tilde{\theta}$ of the coefficient of the TR term. The partial response variation Λ is defined as the variation of response maps between two continuous frames, with the Equation as below.

$$\boldsymbol{\Lambda} = \frac{\boldsymbol{R}_t[\psi_\Delta] - \boldsymbol{R}_{t-1}}{\boldsymbol{R}_{t-1}} \tag{1}$$

Where, $\boldsymbol{R}$ refers to the response map calculated in the detection phase. $[\psi_\Delta]$ represents the shift operator which makes the response peaks in response maps of two continuous frames coincide with each other. As for Autotrack, the integral objective

loss function is shown below:

$$\begin{aligned} \boldsymbol{E}(\boldsymbol{H}_t, \theta_t) &= \frac{1}{2}\left\| \boldsymbol{y} - \sum_{k=1}^{K} \boldsymbol{x}_t^k \circledast \boldsymbol{h}_t^k \right\|_2^2 + \frac{1}{2}\sum_{k=1}^{K}\left\| \tilde{\boldsymbol{u}} \odot \boldsymbol{h}_t^k \right\|_2^2 \\ &+ \frac{\theta_t}{2}\sum_{k=1}^{K}\left\| \boldsymbol{h}_t^k - \boldsymbol{h}_{t-1}^k \right\|_2^2 + \frac{1}{2}\left\| \theta_t - \tilde{\boldsymbol{\theta}} \right\|_2^2 \\ \text{s.t.} \quad \tilde{\boldsymbol{u}} &= \boldsymbol{P}^{\top}\delta log(\boldsymbol{\Lambda}+1) + \boldsymbol{u} \\ \tilde{\boldsymbol{\theta}} &= \frac{\zeta}{1 + log(\nu \left\|\boldsymbol{\Lambda}\right\|_2 + 1)}, \left\|\boldsymbol{\Lambda}\right\|_2 \leq \phi \end{aligned} \tag{2}$$

Where, $\boldsymbol{X}_t = [\boldsymbol{x}_t^1, \boldsymbol{x}_t^2, ..., \boldsymbol{x}_t^K]$ and $\boldsymbol{H}_t = [\boldsymbol{h}_t^1, \boldsymbol{h}_t^2, ..., \boldsymbol{h}_t^K]$ represent the trained filter and the extracted target feature matrix at t frame, respectively. K is the total number of feature channels. $\boldsymbol{x}_t^k \in \boldsymbol{R}^{T\times T}$ indicates the sample feature vector with length T in frame t in k channel and $y \in \boldsymbol{R}^{T\times T}$ is the desired corresponding label set in the Gaussian shape. $\tilde{\boldsymbol{u}}$ and θ_t represent the coefficients of SR and TR, respectively. $\tilde{\theta}$ is the reference value of θ_t used for measuring the change in the tracking response map between two continuous frames. $\mathbf{P}^{\mathrm{T}} \in R^{T\times T}$ is a binary matrix, used in cropping the central elements of the training sample $\boldsymbol{X}_t$. δ is a constant that can be used in balancing the weights of partial response variations. $\boldsymbol{u}$ represents a fixed bowl-shaped matrix of SR which is identical to the STRCF tracker. $\circledast$ and $\odot$ represent the convolution operation and the elemental multiplication, respectively. $\|\|_2^2$ is the Euclidean norm.

SR and TR, constructed by response maps variation, enable the trained filter in Autotrack to adjust automatically while flying and be more adaptable to different scenarios. Although this method has achieved outstanding performance, it does have two limitations. a) This method uses the response map generated by the filter in the previous frame, rather than the learned filter in this frame, thus leading to insufficient suppression of interference. Sudden changes in response maps give important information regarding the similarity of the current object and the appearance model and reveal potential aberrances. The tracker should reduce the learning of irrelevant objects according to the changes during the training phase. b) The weight of each feature channel is equivalent. Different channels describe the objects in different dimensions. There may be many similar features between the target and the background, which are useless or even have a negative effect on the discriminatory ability of trackers. Thus, the filter selects partial distinctive features based on the actual situation for training and updating.

2.2. Loss function construction

To solve the above problems and enhance the discrimination ability and anti-interference ability of the tracker, the weight of the feature channel and aberrance suppression are introduced together to restrain the filter. Specifically, feature channel weight, which is treated as an optimization variable, updates simultaneously with the filter. Also, the variation of two continuous response maps, as an aberrance suppression regularization, is integrated into the training process. The loss function is shown below.

$$\begin{aligned} \boldsymbol{E}(\boldsymbol{H}_t, \theta_t, \boldsymbol{v}_t) &= \frac{1}{2}\left\| \boldsymbol{y} - \sum_{k=1}^{K} \boldsymbol{x}_t^k \circledast \boldsymbol{h}_t^k \right\|_2^2 + \frac{1}{2}\sum_{k=1}^{K}\left\| \upsilon_t^k \tilde{\boldsymbol{u}} \odot \boldsymbol{h}_t^k \right\|_2^2 \\ &+ \frac{\lambda_1}{2}\sum_{k=1}^{K}\left\| \upsilon_t^k - \upsilon_0^k \right\|_2^2 \\ &+ \frac{\theta_t}{2}\sum_{k=1}^{K}\left\| \boldsymbol{h}_t^k - \boldsymbol{h}_{t-1}^k \right\|_2^2 + \frac{1}{2}\left\| \theta_t - \tilde{\boldsymbol{\theta}} \right\|_2^2 \\ &+ \frac{\lambda_2}{2}\left\| \boldsymbol{Q}_{t-1} - \sum_{k=1}^{K} \boldsymbol{x}_t^k \circledast \boldsymbol{h}_t^k \right\|_2^2 \end{aligned} \tag{3}$$

Where υ_t^k is the weight coefficient of feature channel k at t frame. It should be noted that υ_t^k is not a fixed parameter, but a variable that changes with the target appearance during the tracking. The constant υ_0^k is regarded as the reference of υ_t^k, which represent the advance distributions of targets in the different feature channels. υ_0^k is set to 1, thereby ensuring that each feature channel has the same weight in the initial state. $\boldsymbol{Q}_{t-1}$ refers to the response map generated from the previous frame, and is equivalent to $\sum_{k=1}^{K} \boldsymbol{x}_{t-1}^k \circledast \boldsymbol{h}_{t-1}^k$. Thus, it can be treated as a constant signal during the optimization stage. λ_1, and λ_2 are parameters that control model overfitting.

Equation 3 consists of six items that can be divided into four parts. The first part constitutes the first item, the regression term. The second part, including the second and third items, is the SR integrated with channel selection. The third part, consisting of the fourth and fifth items, is the TR borrowed from Autotrack. The fourth part, made up of the last item, is the regularization term, aiming at restricting and counteracting the aberrances created by the background information. For the introduction of channel weight υ_t^k, the feature sifting of the filter is realized in the channel domain by mitigating the impact of features having no relation to the targets and by excluding needless information. By introducing aberrance repressed regularization, which gives greater penalties for interference, the ability of the tracker to identify the aberrance in the background, and suppress the subsequent changes of response maps on the basis of the baseline, is further improved. The fusion of these two factors enables the filter to find the aberrance in time, and utilize the best features, thereby maximizing the differentiation between the target and background.

2.3. Optimization

As observed from Equation 3, the optimization of the overall loss function involves the complex correlation operation between matrices. Therefore, to reduce computational complexity, and reduce sufficient computing efficiency, the Parseval theorem is used to convert complex correlation operations into simple elemental multiplication operations and move the loss function from the time domain to the Fourier domain as $\boldsymbol{E}(\boldsymbol{H}_t, \hat{\boldsymbol{G}}_t, \theta_t, \boldsymbol{\upsilon}_t)$. Besides, the constraint parameter $\hat{\boldsymbol{g}}_t^k = \sqrt{T}\boldsymbol{F}\mathbf{P}^{\mathrm{T}}\boldsymbol{h}_t^k$ is used in constituting the Augmented Lagrangian function $\boldsymbol{L}(\boldsymbol{H}_t, \hat{\boldsymbol{G}}_t, \boldsymbol{\upsilon}, \theta_t, \hat{\boldsymbol{M}}_t)$ as follows:

$$\begin{aligned}
\boldsymbol{L}(\boldsymbol{H}_t, \hat{\boldsymbol{G}}_t, \boldsymbol{\upsilon}, \theta_t, \hat{\boldsymbol{M}}_t) &= \boldsymbol{E}(\boldsymbol{H}_t, \hat{\boldsymbol{G}}_t, \theta_t, \boldsymbol{\upsilon}_t) \\
&\quad + \sum_{k=1}^{K} (\hat{\boldsymbol{g}}_t^k - \sqrt{T}\boldsymbol{F}\mathbf{P}^{\mathrm{T}}\boldsymbol{h}_t^k)\hat{\boldsymbol{m}}_t^k \\
&\quad + \frac{\mu}{2}\sum_{k=1}^{K}\left\|\hat{\boldsymbol{g}}_t^k - \sqrt{T}\boldsymbol{F}\mathbf{P}^{\mathrm{T}}\boldsymbol{h}_t^k\right\|_2^2 \\
\boldsymbol{E}(\boldsymbol{H}_t, \hat{\boldsymbol{G}}_t, \theta_t, \boldsymbol{\upsilon}_t) &= \frac{1}{2}\left\|\hat{\boldsymbol{y}} - \sum_{k=1}^{K}\hat{\boldsymbol{x}}_t^k \odot \hat{\boldsymbol{g}}_t^k\right\|_2^2 \\
&\quad + \frac{1}{2}\sum_{k=1}^{K}\left\|\upsilon_t^k\tilde{\boldsymbol{u}} \odot \boldsymbol{h}_t^k\right\|_2^2 + \frac{\theta_t}{2}\sum_{k=1}^{K}\left\|\hat{\boldsymbol{g}}_t^k - \hat{\boldsymbol{g}}_{t-1}^k\right\|_2^2 \\
&\quad + \frac{1}{2}\left\|\theta_t - \tilde{\boldsymbol{\theta}}\right\|_2^2 + \frac{\lambda_1}{2}\sum_{k=1}^{K}\left\|\upsilon_t^k - \upsilon_0^k\right\|_2^2 \\
&\quad + \frac{\lambda_2}{2}\left\|\hat{\boldsymbol{Q}}_{t-1} - \sum_{k=1}^{K}\hat{\boldsymbol{x}}_t^k \odot \hat{\boldsymbol{g}}_t^k\right\|_2^2
\end{aligned} \tag{4}$$

Where symbol $\hat{}$ represents the discrete Fourier transformation (DFT), for example, $\hat{\boldsymbol{y}} = \sqrt{N}\boldsymbol{F}\boldsymbol{y}$ and $\boldsymbol{F}$ called the Fourier matrix is the orthonormal $N \times N$ matrix of complex basis vectors. $\boldsymbol{m}$ refers to the Lagrangian multiplier, and $\boldsymbol{\mu}$ represents the penalty parameter. For simplification, $\hat{\boldsymbol{G}}_t = [\hat{\boldsymbol{g}}_t^1, \hat{\boldsymbol{g}}_t^2, \hat{\boldsymbol{g}}_t^3, ..., \hat{\boldsymbol{g}}_t^K]$ and $\hat{\boldsymbol{M}}_t = [\hat{\boldsymbol{m}}_t^1, \hat{\boldsymbol{m}}_t^2, \hat{\boldsymbol{m}}_t^3, ..., \hat{\boldsymbol{m}}_t^K]$ are defined. By assigning $\hat{\boldsymbol{s}}_t^k = \frac{1}{\mu}\hat{\boldsymbol{m}}_t^k$ the optimization of Equation (4) is equivalent to solving equation (5).

$$\begin{aligned}
\boldsymbol{L}(\boldsymbol{H}_t, \hat{\boldsymbol{G}}_t, \boldsymbol{\upsilon}, \theta_t, \hat{\boldsymbol{S}}_t) &= \boldsymbol{E}(\boldsymbol{H}_t, \hat{\boldsymbol{G}}_t, \boldsymbol{\upsilon}, \theta_t) \\
&\quad + \frac{\mu}{2}\sum_{k=1}^{K}\left\|\hat{\boldsymbol{g}}_t^k - \sqrt{T}\boldsymbol{F}\mathbf{P}^{\mathrm{T}}\boldsymbol{h}_t^k + \hat{\boldsymbol{s}}_t^k\right\|_2^2
\end{aligned} \tag{5}$$

Considering the complexity of the above-mentioned function, the alternative direction method of multipliers (ADMM) (Lin et al., 2010) is applied to speed up the calculation. Specifically, the function of optimization can be divided into a few sub-problems to be solved iteratively. During the solution of every subproblem, only one variable is contained to be optimized, while the others are regarded as fixed constants temporarily. In this way, each subproblem and its relevant closed-form solution can be given in detail below.

Subproblem for $\hat{\boldsymbol{G}}_t$: By giving $\boldsymbol{H}_t$, $\boldsymbol{\upsilon}$, θ_t, $\hat{\boldsymbol{S}}_t$, the optimal $\hat{\boldsymbol{G}}_t^*$ could be obtained by solving the optimization problem:

$$\begin{aligned}
\hat{\boldsymbol{G}}_t^* &= \arg\min_{\hat{\boldsymbol{G}}_t^*}\{\frac{1}{2}\left\|\hat{\boldsymbol{y}} - \sum_{k=1}^{K}\hat{\boldsymbol{x}}_t^k \odot \hat{\boldsymbol{g}}_t^k\right\|_2^2 + \frac{\theta_t}{2}\sum_{k=1}^{K}\left\|\hat{\boldsymbol{g}}_t^k - \hat{\boldsymbol{g}}_{t-1}^k\right\|_2^2 \\
&\quad + \frac{\lambda_2}{2}\left\|\hat{\boldsymbol{Q}}_{t-1} - \sum_{k=1}^{K}\hat{\boldsymbol{x}}_t^k \odot \hat{\boldsymbol{g}}_t^k\right\|_2^2 \\
&\quad + \frac{\mu}{2}\sum_{k=1}^{K}\left\|\hat{\boldsymbol{g}}_t^k - \sqrt{T}\boldsymbol{F}\mathbf{P}^{\mathrm{T}}\boldsymbol{h}_t^k + \hat{\boldsymbol{s}}_t^k\right\|_2^2\}
\end{aligned} \tag{6}$$

However, it is still very difficult to solve Equation 6 directly, because this subproblem containing $\hat{\boldsymbol{X}}_k\hat{\boldsymbol{g}}_k$ shows a high computation complexity and needs multiple iterations in ADMM. Fortunately, $\hat{\boldsymbol{X}}_k$ is sparse, which means that each element of $\hat{\boldsymbol{y}}(\hat{\boldsymbol{y}}(n), n = 1, 2, \ldots, N)$ is merely related to $\hat{\boldsymbol{x}}_k(n) = [\hat{\boldsymbol{x}}_k(n)^1, \hat{\boldsymbol{x}}_k(n)^2, \ldots, \hat{\boldsymbol{x}}_k(n)^D]$ and $\hat{\boldsymbol{g}}_k(n) = [conj(\hat{\boldsymbol{g}}_k(n)^1), conj(\hat{\boldsymbol{g}}_k(n)^2), ..., conj(\hat{\boldsymbol{g}}_k(n)^D)]$, where $conj()$ refers to the complex conjugate operation. Thus, this subproblem can be divided into N simpler problems across K channels as follows.

$$\begin{aligned}
\Gamma_j^*(\hat{\boldsymbol{G}}_t) &= \arg\min_{\Gamma_j(\hat{\boldsymbol{G}}_t)}\{\left\|\hat{\boldsymbol{y}}_j - \Gamma_j(\hat{\boldsymbol{X}}_\mathbf{t})^\top\Gamma_j(\hat{\boldsymbol{G}}_t)\right\|_2^2 \\
&\quad + \mu\left\|\Gamma_j(\hat{\boldsymbol{G}}_t) + \Gamma_j(\hat{\boldsymbol{s}}_t) - \Gamma_j(\sqrt{T}\mathbf{F}\mathbf{P}^\top\boldsymbol{H}_t)\right\|_2^2 \\
&\quad + \theta_t\left\|\Gamma_j(\hat{\boldsymbol{G}}_t) - \Gamma_j(\hat{\boldsymbol{G}}_{t-1})\right\|_2^2 \\
&\quad + \frac{\lambda_2}{2}\left\|\hat{\boldsymbol{Q}}_{t-1} - \Gamma_j(\hat{\boldsymbol{X}}_\mathbf{t})^\top\Gamma_j(\hat{\boldsymbol{G}}_t)\right\|_2^2\}
\end{aligned} \tag{7}$$

Where, $\Gamma_j(\hat{\boldsymbol{G}}_t) \in C^{(}K \times 1)$ indicates the vector including all K channel value of $\hat{\boldsymbol{G}}_t$ on pixel $j(j = 1, 2, \ldots, N)$. By introducing the Sherman-Morrison formula $(uv^H + A)^{-1} = A^{-1} - \frac{A^{-1}uv^HA^{-1}}{v^HA^{-1}u+1}$, the inverse operation in the derivation can be further simplified and accelerated.Then, the closed-form solution of this subproblem can be obtained as follows.

$$\Gamma_j^*(\hat{\boldsymbol{G}}_t) = \frac{1}{\mu + \theta_t}(\mathbf{I} - \frac{(1+\lambda_2)\Gamma_j(\hat{\mathbf{X}}_t)\Gamma_j(\hat{\mathbf{X}}_t)^\top}{\theta_t + \mu + (1+\lambda_2)\Gamma_j(\hat{\mathbf{X}}_t)^\top\Gamma_j(\hat{\mathbf{X}}_t)})\boldsymbol{\rho} \tag{8}$$

Where $\boldsymbol{\rho}$ is merely an intermediate variable for simple representation and $\boldsymbol{\rho} = \Gamma_j(\hat{\mathbf{X}}_t)\hat{\mathbf{y}}_j + \theta_t\Gamma_j(\hat{\boldsymbol{G}}_{t-1}) - \mu\Gamma_j(\hat{\mathbf{S}}_t) + \mu\Gamma_j(\sqrt{T}\mathbf{F}\mathbf{P}^\top\boldsymbol{H}_t) + \lambda_1\Gamma_j(\hat{\mathbf{X}}_t)\hat{\mathbf{Q}}_{t-1}$

Subproblem for $\boldsymbol{H}_t$: By fixing $\hat{\boldsymbol{G}}_t$,$\boldsymbol{\upsilon}$,θ_t,$\hat{\boldsymbol{S}}_t$, $\boldsymbol{H}_t$ can be solved with the equation below:

$$\begin{aligned}
\boldsymbol{h}_t^{k*} &= \arg\min_{\boldsymbol{h}_t^k}\{\frac{1}{2}\left\|\upsilon_t^k\tilde{\boldsymbol{u}} \odot \boldsymbol{h}_t^k\right\|_2^2 + \frac{\mu}{2}\left\|\hat{\boldsymbol{g}}_t^k - \sqrt{T}\boldsymbol{F}\mathbf{P}^{\mathrm{T}}\boldsymbol{h}_t^k + \hat{\boldsymbol{s}}_t^k\right\|_2^2\} \\
&= \frac{\mu T\mathbf{p} \odot (\boldsymbol{s}_t^k + \boldsymbol{g}_t^k)}{\lambda_1(\upsilon_t^k\tilde{\boldsymbol{u}} \odot \upsilon_t^k\tilde{\boldsymbol{u}}) + \mu T\mathbf{p}}
\end{aligned} \tag{9}$$

Where, $\mathbf{p} = [\mathbf{P}_{11}, \mathbf{P}_{22}, \ldots, \mathbf{P}_{TT}]^\top$ represents the column vector, that composed of the diagonal elements of $\mathbf{P}$. As observed in Equation 9, the computational cost on $\boldsymbol{h}_t^{k*}$ solution is very low, because it only involves the element-wise operation and an inverse fast Fourier transform.

Subproblem for θ_t: By treating $\hat{\boldsymbol{G}}_t$,$\boldsymbol{\upsilon}$,$\boldsymbol{H}_t$,$\hat{\boldsymbol{S}}_t$ as constants, the optimal θ_t can be obtained by solving the problem of optimization below:

$$\begin{aligned}\theta_t^* &= \arg\min_{\theta_t} \{\frac{\theta_t}{2}\sum_{k=1}^{K}\left\|\hat{\boldsymbol{g}}_t^k - \hat{\boldsymbol{g}}_{t-1}^k\right\|_2^2 + \frac{1}{2}\left\|\theta_t - \tilde{\theta}\right\|_2^2\} \\ &= \tilde{\theta} - \frac{\sum_{k=1}^{K}\left\|\hat{\boldsymbol{g}}_t^k - \hat{\boldsymbol{g}}_{t-1}^k\right\|_2^2}{2}\end{aligned} \tag{10}$$

Subproblem for $\boldsymbol{\upsilon_t}^*$ Given $\hat{\boldsymbol{G}}_t$,θ_t,$\boldsymbol{H}_t$,$\hat{\boldsymbol{S}}_t$, $\boldsymbol{\upsilon_t^k}$ can be optimized with the following equation.

$$\begin{aligned}\upsilon_t^{k*} &= \arg\min_{\upsilon_t^k} \frac{1}{2}\sum_{k=1}^{K}\left\|\upsilon_t^k\tilde{\boldsymbol{u}} \odot \boldsymbol{h}_t^k\right\|_2^2 + \frac{\lambda_1}{2}\sum_{k=1}^{K}\left\|\upsilon_t^k - \upsilon_0^k\right\|_2^2 \\ &= \frac{\lambda_1\upsilon_0^k}{(\tilde{\boldsymbol{u}} \odot h_t^k)^\top(\tilde{\boldsymbol{u}} \odot h_t^k) + \lambda_1}\end{aligned} \tag{11}$$

Lagrangian multiplier updating:

$$\hat{\boldsymbol{S}}_t^{i+1} = \hat{\boldsymbol{S}}_t^{\,i} + \mu^i(\hat{\boldsymbol{G}}_t^{\,i+1} - \hat{\boldsymbol{H}}_t^{\,i+1}) \tag{12}$$

Where, i and $i+1$ represent the previous and current iterations. The new $\hat{\boldsymbol{G}}_t$ $\hat{\boldsymbol{H}}$ obtained from the above optimization solution is used to update the Lagrangian multiplier. The regularization constant observes the updating laws of $\mu^{i+1} = min(\mu_{max}, \beta\mu^i)$, thereby ensuring the convergence of the integral model according to ADMM.

2.4. Historical model retrieval module

Most original tracking methods use linear interpolation with a constant learning rate β, like Equation 13, to update the filter. However, such an updating method not only causes the tracker to indiscriminately treat all the historical information but also results in filter pollution and degradation. The tracking result is poor when faced with complex scenes, such as partial occlusion, and camera defocus. Too high a learning rate causes the tracker to easily overfit and then neglect historical information, while too low a learning rate disenables the tracker from effectively learning the change of targets. Considering that the human brain can recall historical memory to make the best choice when identifying targets and HMTS tracker, the history filter, namely the historical model retrieval module is retrieved, and the best filter of the current frame is obtained by selecting and linear interpolating several effective filters. Specifically, historical filters are saved first, and a filter library is built. After the training phase of each frame, the correlation between each template and the current sample image is calculated. Several historical templates with the highest scores are selected and the scores are used as weights to linear interpolate them, thereby obtaining a tracking template for the next frame object location. This module is described below in detail with mathematical symbols.

$$\boldsymbol{h}_t = \beta\boldsymbol{h} + (1-\beta)\boldsymbol{h}_{t-1} \tag{13}$$

Similiar to HMTS tracker (Chen et al., 2022), this method retains the filter for each frame as the historical model $\boldsymbol{H}_{hist}$. However, the HMTS tracker builds the filters library with all historical filters, which causes much computing burden and redundancy. For example, when tracking to the end of a long video, there are numerous historical filters, and there is great similarity in target appearance between the current filter and the front filter. Therefore, the size is fixed to ϕ_{hist} and the filters library is constructed as $\boldsymbol{H}_{hist} = \{(\boldsymbol{h}_i, s_i)\}_{i=1}^{\phi_{hist}}$. s_i refers to the score of each historical model.

As expressed by the regression term in the loss function Equation 3, the convolution results of the trained filter and sample should ideally present a Gaussian shape centered on the target, namely the label $\boldsymbol{y}$. The basis of correlation filtering theory is as below: the more similar the two signals are, the greater the correlation between them is. Thus, like the HMTS tracker, the s_i is defined as the correlation between the label $\boldsymbol{y}$ and the convolution results $\boldsymbol{R}_i$ of different historical filters $\boldsymbol{H}_i, i \in [1, \phi_{hist}]$ and the current frame target samples $\boldsymbol{X}_t$. The equation of s_i is as follows:

$$\begin{aligned}s_i &= max(\mathscr{F}^{-1}(\boldsymbol{y}^H\boldsymbol{R}_i)) \\ \boldsymbol{R}_i &= \left\|\sum_{k=1}^{K}\boldsymbol{x}_t^k \circledast \boldsymbol{h}_i^k\right\|_2^2\end{aligned} \tag{14}$$

Where $\mathscr{F}^{-1}$ represents the inverse Fourier transform,^{H} indicates the conjugate transpose, and $max(\cdot)$ refers to the maximum of the vector.

After the tracker training phase in accordance with Section 2.3, Equation (14) is adopted to calculate the scores of the trained filter in the current frame and historical filters in the filters library. Next, the historical model with the lowest score in the filter library is replaced by the filter trained from the current frame, thereby ensuring no change in the number of filters in the library. It needs to be noted that, since the first frame is the most accurate manually labeled target information, the filter of the first frame shall always remain in the filter library. The filter $\boldsymbol{h}_t$ used for object detection in the next frame can be obtained by a linear weighting of the filters with the top ϕ_{scores} scores.

$$\begin{aligned}\boldsymbol{h}_t &= \sum_i s_i\boldsymbol{h}_i \\ s.t.Rank(s_i) &\geq \phi_{scores}\end{aligned} \tag{15}$$

Where, $Rank(s_i)$ represents the index of s_i in the set $\{s_i\}_{i=1}^{\phi_{hist}}$, which is ranked in descending and $i \in [1, \phi_{hist}]$. It needs to be noted that the filter trained in the first frame always participates in the calculation of Equation 15 and it is given the lowest weight in ϕ_{scores} filters if $Rank(s_1) \leq \phi_{scores}$.

3. Experiments

In this section, the tracking performance of the proposed tracker is evaluated against the nine state-of-the-art trackers, namely AutoTrack, ASRCF, ECO-HC, STRCF, SRDCF, BACF, LADCF (Xu et al., 2019), MCCT-H (Wang et al., 2018) and Staple (Bertinetto et al., 2016a) on two difficult UAV benchmarks (UAV123 Mueller et al., 2016 and VisDrone2018-test-dev Zhu et al., 2018). For the measurement of the performance of the aforementioned trackers, the employed evaluation metric named one-pass evaluation(OPE) contained two indicators, namely Precision Rate and Success Rate. It needs to be noted that the precision plot threshold is set to 10 pixels in UAV123 and to 21 pixels in VisDrone2018-test-dev, when considering the different target sizes from different UAV datasets.

3.1. Implementation details

Our tracker was used in MATLAB-2017a with an Intel i7-9750H CPU, and 16GB of RAM, and runs at a 25 FPS average with hand-crafted characteristics for target representation. The common hyper-parameters are kept to the same values as the baseline Autotrack, namely $\delta = 0.2$, $\nu = 2 \times 10^{-5}$, and $\zeta = 13$. The SR constraint coefficient λ_1 and the response aberrance regularization constraint coefficient λ_2 which are unique to the proposed tracker, are set as 0.71 and 0.001, respectively. In the historical model retrieval module, $\phi_{hist} = 30$ and $\phi_{scores} = 20$ are determined. As for the ADMM algorithm, the number of iterations is set as 4, $\beta = 10$, and $\mu_{max} = 10^4$, which also shares the same parameters as in Autotrack.

3.2. Quantitative evaluation

UAV123 is the most commonly used dataset in UAV object tracking, with 123 videos with more than 110K frames composed. In these sequences, 12 of the challenging attributes involved, such as background clutter, aspect ratio change, and similar object, required a more accurate and stable tracking algorithm. The quantitative comparison of different trackers is shown in Figure 1, and it can be observed that our tracker shows the best precision with the second success rate, slightly lower than ECO-HC. However, the proposed method achieves a remarkable advantage of 2.6% in precision and 1.5% in success rate, compared with the baseline tracker Autotrack.

VisDrone2018-test-dev is a dataset that is especially proposed for aerial object tracking competition. It consists of 35 videos captured from 14 different cities and covers various aspects including such as shooting position, tracking scene, target type, and object density. Different scenarios, weather conditions, and illumination changes are primarily addressed in this dataset. As shown in Figure 2, the proposed tracker is superior to all other evaluated trackers, and it can achieve 81.1% and 60.7% in the distance precision (DP) and the area under the curve (AUC), respectively. By comparing with the baseline tracker, Autotrack, our tracker accomplishes 2.3% and 3.4% of performance gains in precision and success rate, respectively.

3.3. Parametric sensitivity

As presented in Section 3.1, some hyper-parameters of the proposed tracker need to be set, namely the spatial-channel regularization constraint coefficient λ_1 and the response aberrance regularization constraint coefficient λ_2 in the loss function. In this section, the influence of different configurations on tracking results is investigated. When evaluating each hyper-parameter for a fair comparison, the common parameters are maintained at the same value as in Autotrack and all other parameters are fixed. Considering the operation speed, ϕ_{hist} is set as a constant of 30 and $\phi_{scores} = 20$ is set as a constant of 20 to ensure the efficient use of historical information and the effective reduction of redundancy. Table 1 exhibits the tracking results under different λ_1,λ_2 in VisDrone2018-test-dev, where ϕ_{scores} is fixed to 20. It can be observed that this tracker yields the best performance with $\lambda_1 = 0.001$ and $\lambda_2 = 0.71$.

3.4. Ablation experiments

As described in Section 2, in our method loss function is reconstructed by introducing the feature channel weight and aberrance repressed regularization, and an additional historical memory model is added to the baseline Autotrack. To prove the effectiveness of each module, ablation experiments were conducted. The results are shown in Table 2. AutoTrack_csar only reconstructs the loss function, while AutoTrack_hist only adds the historical memory model. As observed, by adding the two modules separately, the performance of the baseline tracker can be improved effectively. Moreover, by joining these two modules simultaneously, our method can achieve excellent performance against the baseline. This is mainly because the fusion of the two enables the tracker to effectively use historical information to prevent background clutter during tracking while establishing a more robust target appearance representation.

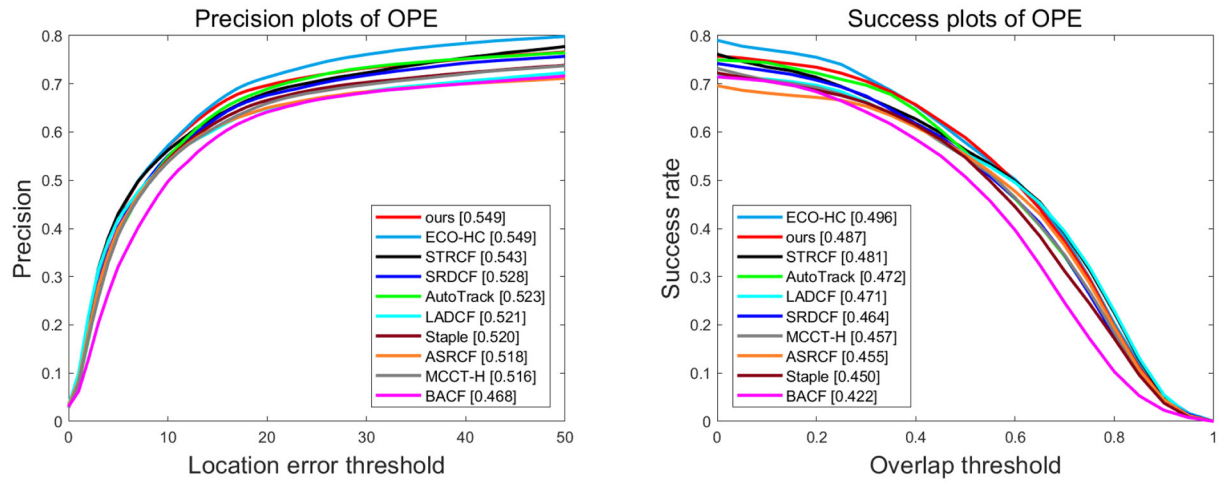

FIGURE 1
Overall performance of ten trackers on UAV123. The legend of success plot represents the area under the curve score of each tracker. The legend of the precision plot refers to the threshold scores at 10 pixels.

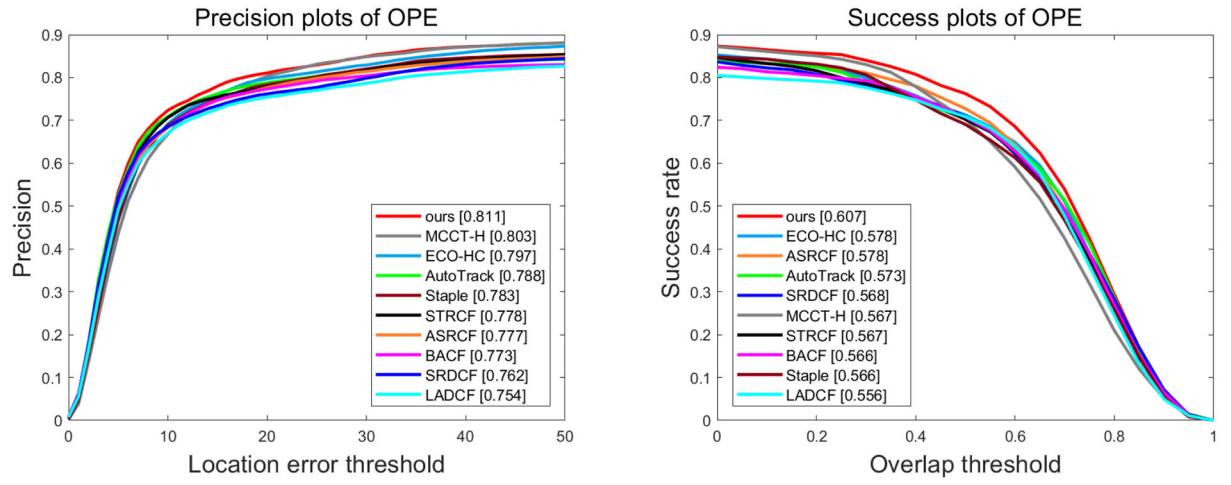

FIGURE 2
Overall performance of ten trackers on Visdrone2018-test-dev. The legend of success plot represents the area under the curve score of each tracker. The legend of the precision plot refers to threshold scores at 21 pixels.

TABLE 1 The success rate and precision rate (percentage) related to the varying number of regularization constraint coefficients on VisDrone2018-test-dev.

Parameter	λ_1				λ_2			
Value	**0.001**	0.1	0.5	1	**0.71**	0.01	0.1	1
Success Rate	**60.7**	58.4	59.2	59.0	**60.7**	58.5	59.1	59.0
Precision Rate	**81.1**	78.4	79.5	80.2	**81.1**	79.1	80.1	79.5

In historical models, $\phi_{scores} = 20$ and $\phi_{hist} = 30$. The threshold of precision rate is set to 21 pixels.
Bold values refer to first place in the experiments.

TABLE 2 The success rate and precision rate (percentage) of ablation experiments on UAV123.

Tracker	AutoTrack	Two regularization	Historical memory	Precision rate	Success rate
AutoTrack	✓			52.3	47.2
AutoTrack_csar	✓	✓		53.7	47.4
AutoTrack_hist	✓		✓	54.2	47.5
Ours	✓	✓	✓	**54.9**	**48.7**

The precision rate threshold is set as 10 pixels.
Bold values refer to first place in the experiments.

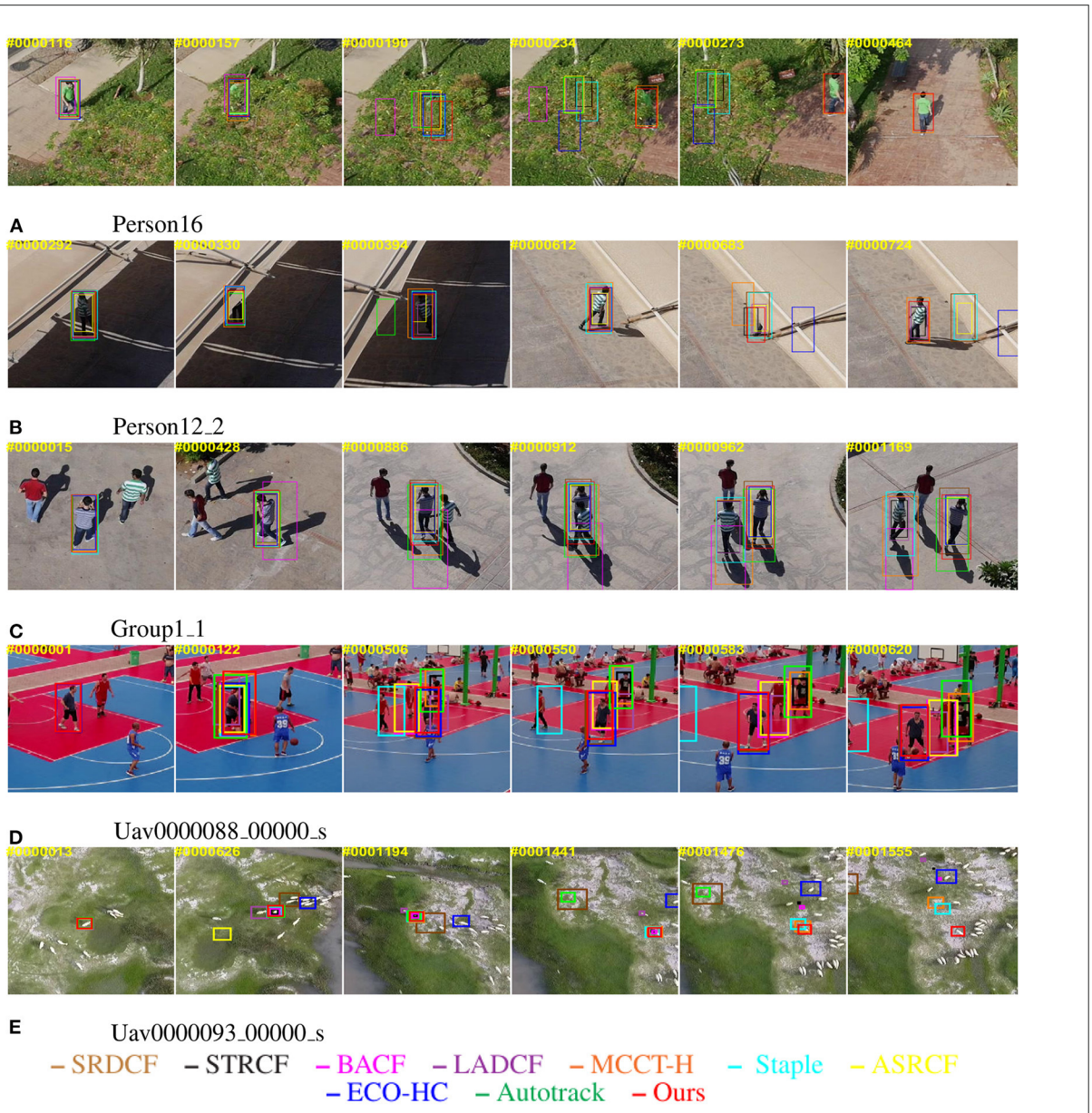

FIGURE 3
Qualitative performance evaluation of the proposed tracker and the other nine most advanced trackers on the typical UAV videos. The number in the upper left corner refers to the frame number. The tracking boxes in different colors represent the tracking results of different trackers in the frame. **(A)** Person16. **(B)** Person12_2. **(C)** Group1_1. **(D)** Uav0000088_00000_s. **(E)** Uav0000093_00000_s. The photos appearing in this figure have been reused from: 'A benchmark and simulator for uav tracking' and 'Vision meets drones: a challenge'. The corresponding website are 'https://cemse.kaust.edu.sa/ivul/uav123' and 'http://aiskyeye.com'.

3.5. Qualitative evaluation

In this subsection, the qualitative comparison is given to the proposed method and the aforementioned 9 state-of-the-art algorithms to better demonstrate the performance of each tracker in Figure 3. The above image sequences (containing person16, person12_2, group1_1 in UAV123 and Uav0000088_00000_s, and Uav0000093_00000_s in VisDrone2018-test-dev) mainly include three challenging attributes, namely similar object (SOB), background clutters (BC), and occlusion (OC). It can be observed that our tracker is effective in solving these difficult issues, and can locate the targets accurately.

When facing a similar object and background clutter, aberrance repressed regularization can help the tracker in accurately perceiving and fully restraining the interference regions in advance. Simultaneously, dynamic feature channel weight realizes the independent filtering of different dimensional features, thereby encouraging the filters to focus on more reliable and discriminative features between the target and a cluttered background. By jointly modeling the above two constraints, the tracker can learn the robust features of the target according to the environment and the interference from a cluttered background.

When there is an occlusion, the trackers can learn the features of the block and lose the target information, thus leading to model drift and a failure of tracking. With the introduction of a historical model retrieval module in our method, the tracker has a memory function similar to the human brain by saving a history template. The method of dynamic updating of the template encourages the tracker to reduce the learning rate when the training sample is abnormal, thereby effectively reducing the probability of template pollution. The memory function of the tracker also guarantees that the method can accurately lock the target again after the disappearance of the occlusion.

In summary, when challenging attributes occur during tracking, the addition of the two constraints endows the tracker with the ability to select the most distinguishing feature for sensing and suppressing the interference around the target, while the historical model retrieval module effectively reduces the pollution of interference and noise to the tracker. However, when meeting viewpoint change and rotation, the performance of our tracker is reduced because of rapid changes in the appearance of the target. In the future, we will explore how to refine tracking results to solve such problems.

4. Conclusion

Based on the idea that the brain can perceive interference information in the background, select the optimal features independently to describe the target, and use historical memory to achieve accurate target location in the current frame, in this paper, we propose a UAV tracking algorithm on the basis of repressed dynamic aberrance and a channel selective correlation filter with a historical model retrieval module combined. By jointly modeling feature channel weight and the aberrance repressed regularization, our tracker could restrain the potential distractors, and highlight the valuable features in the channel domain, thereby constructing a robust target appearance. With a historical model retrieval module, our tracker can make full use of the historical information to update the tracker, while effectively avoiding tracking drift. The experimental results on the two public UAV benchmarks demonstrate that the proposed method achieves better tracking results than the other advanced algorithms.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

JC and JW proposed the basic idea of this method and wrote the code together. JC completed theoretical modeling. JC and LZ performed the experiments and data analysis. JC wrote the first draft of the manuscript. JW and LZ revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Natural Science Foundation of China (Grant No. 12072027).

Acknowledgments

The authors thank Małgorzata Siemiątkowska, Anna Olichwer, Magdalena Żukowska, Maksymilan Koc, and Karolina Adaszewska, for their significant help with the organization of the study, and the psychophysiological data collection. We also would like to thank Marta Grześ – for their help with the participants' recruitment and contact. We also thank Piotr Kałowski, Dominika Pruszczak, Małgorzata Hanć, and Michał Lewandowski for their help with the misophonia interviews.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Atkinson, R. C., and Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. *Psychol. Learn. Motivat.* 2, 89–195. doi: 10.1016/S0079-7421(08)60422-3

Baca, T., Hert, D., Loianno, G., Saska, M., and Kumar, V. (2018). "Model predictive trajectory tracking and collision avoidance for reliable outdoor deployment of unmanned aerial vehicles," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Madrid: IEEE), 6753–6760.

Bertinetto, L., Valmadre, J., Golodetz, S., Miksik, O., and Torr, P. H. (2016a). "Staple: complementary learners for real-time tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV: IEEE), 1401–1409.

Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., and Torr, P. H. (2016b). "Fully-convolutional siamese networks for object tracking," in *European Conference on Computer Vision* (Cham: Springer), 850–865.

Bolme, D. S., Beveridge, J. R., Draper, B. A., and Lui, Y. M. (2010). "Visual object tracking using adaptive correlation filters," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (San Francisco, CA: IEEE), 2544–2550.

Chen, S., Wang, T., Wang, H., Wang, Y., Hong, J., Dong, T., et al. (2022). Vehicle tracking on satellite video based on historical model. *IEEE J. Select. Top. Appl. Earth Observat. Remote Sens.* 15, 7784–7796. doi: 10.1109/JSTARS.2022.3195522

Connor, S., Li, T., Roberts, R., Thakkar, S., Liu, Z., and Tong, W. (2022). The adaptability of using ai for safety evaluation in regulatory science: a case study of assessing drug-induced liver injury (dili). *Front. Artif. Intel.* 5, 1034631. doi: 10.3389/frai.2022.1034631

Cornelio, P., Haggard, P., Hornbaek, K., Georgiou, O., Bergström, J., Obrist, M., et al. (2022). The sense of agency in emerging technologies for human-computer integration: a review. *Front. Neurosci.* 16, 949138. doi: 10.3389/fnins.2022.949138

Dai, K., Wang, D., Lu, H., Sun, C., and Li, J. (2019). "Visual tracking via adaptive spatially-regularized correlation filters," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Long Beach, CA: IEEE), 4670–4679.

Danelljan, M., Bhat, G., Shahbaz Khan, F., and Felsberg, M. (2017). "ECO: efficient convolution operators for tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI: IEEE), 6638–6646.

Danelljan, M., Häger, G., Khan, F., and Felsberg, M. (2014a). "Accurate scale estimation for robust visual tracking," in *British Machine Vision Conference, Nottingham* (Nottingham: Bmva Press), 1–5.

Danelljan, M., Hager, G., Shahbaz Khan, F., and Felsberg, M. (2015). "Learning spatially regularized correlation filters for visual tracking," in *Proceedings of the IEEE International Conference on Computer Vision* (Santiago: IEEE), 4310–4318.

Danelljan, M., Shahbaz Khan, F., Felsberg, M., and Van de Weijer, J. (2014b). "Adaptive color attributes for real-time visual tracking," *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Columbus, OH: IEEE), 1090–1097.

Deng, C., Jing, D., Han, Y., Wang, S., and Wang, H. (2022). Far-net: fast anchor refining for arbitrary-oriented object detection. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi: 10.1109/LGRS.2022.3144513

Elloumi, M., Dhaou, R., Escrig, B., Idoudi, H., and Saidane, L. A. (2018). "Monitoring road traffic with a uav-based system," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)* (Barcelona: IEEE), 1–6.

Foksinska, A., Crowder, C., Crouse, A., Henrikson, J., Byrd, W., Rosenblatt, G., et al. (2022). The precision medicine process for treating rare disease using the artificial intelligence tool medikanren. *Front. Artif. Intell.* 5, 910216. doi: 10.3389/frai.2022.910216

Gschwindt, M., Camci, E., Bonatti, R., Wang, W., Kayacan, E., and Scherer, S. (2019). "Can a robot become a movie director? learning artistic principles for aerial cinematography," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Macau: IEEE), 1107–1114.

Han, Y., Deng, C., Zhao, B., and Tao, D. (2019a). State-aware anti-drift object tracking. *IEEE Trans. Image Process.* 28, 4075–4086. doi: 10.1109/TIP.2019.2905984

Han, Y., Deng, C., Zhao, B., and Zhao, B. (2019b). Spatial-temporal context-aware tracking. *IEEE Signal Process. Lett.* 26, 500–504. doi: 10.1109/LSP.2019.2895962

Han, Y., Liu, H., Wang, Y., and Liu, C. (2022). A comprehensive review for typical applications based upon unmanned aerial vehicle platform. *IEEE J. Select. Top. Appl. Earth Observat. Remote Sens.* 15, 9654–9666. doi: 10.1109/JSTARS.2022.3216564

Henriques, J. F., Caseiro, R., Martins, P., and Batista, J. (2014). High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 583–596. doi: 10.1109/TPAMI.2014.2345390

Hong, Z., Chen, Z., Wang, C., Mei, X., Prokhorov, D., and Tao, D. (2015). "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: IEEE), 749–758.

Javed, S., Danelljan, M., Khan, F. S., Khan, M. H., Felsberg, M., and Matas, J. (2022). Visual object tracking with discriminative filters and siamese networks: a survey and outlook. *IEEE Trans. Pattern Anal. Mach. Intell.* 1–20. doi: 10.1109/TPAMI.2022.3212594

Kiani Galoogahi, H., Fagg, A., and Lucey, S. (2017). "Learning background-aware correlation filters for visual tracking," in *Proceedings of the IEEE International Conference on Computer Vision* (Venice: IEEE), 1135–1143.

Kuroda, N., Ikeda, K., and Teramoto, W. (2022). Visual self-motion information contributes to passable width perception during a bike riding situation. *Front. Neurosci.* 16, 938446. doi: 10.3389/fnins.2022.938446

Li, B., Yan, J., Wu, W., Zhu, Z., and Hu, X. (2018a). "High performance visual tracking with siamese region proposal network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE), 8971–8980.

Li, F., Tian, C., Zuo, W., Zhang, L., and Yang, M.-H. (2018b). "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: IEEE), 4904–4913.

Li, Y., Fu, C., Ding, F., Huang, Z., and Lu, G. (2020). "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, 11923–11932.

Li, Y., and Zhu, J. (2014). "A scale adaptive kernel correlation filter tracker with feature integration," in *European Conference on Computer Vision* (Cham: Springer), 254–265.

Lin, Z., Chen, M., and Ma, Y. (2010). The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*. doi: 10.48550/arXiv.1009.5055

Liu, C., Ding, W., Chen, P., Zhuang, B., Wang, Y., Zhao, Y., et al. (2022). Rb-net: training highly accurate and efficient binary neural networks with reshaped point-wise convolution and balanced activation. *IEEE Trans. Circ. Syst. Video Technol.* 32, 6414–6424. doi: 10.1109/TCSVT.2022.3166803

Mueller, M., Smith, N., and Ghanem, B. (2016). "A benchmark and simulator for uav tracking," in *European Conference on Computer Vision* (Cham: Springer), 445–461.

Mueller, M., Smith, N., and Ghanem, B. (2017). "Context-aware correlation filter tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, H: IEEE), 1396–1404.

Nofallah, S., Wu, W., Liu, K., Ghezloo, F., Elmore, J. G., and Shapiro, L. G. (2022). Automated analysis of whole slide digital skin biopsy images. *Front. Artif. Intell.* 5, 1005086. doi: 10.3389/frai.2022.1005086

Pfeifer, L. D., Patabandige, M. W., and Desaire, H. (2022). Leveraging r (levr) for fast processing of mass spectrometry data and machine learning: applications analyzing fingerprints and glycopeptides. *Front. Anal. Sci.* 2, 961592. doi: 10.3389/frans.2022.961592

Shao, J., Du, B., Wu, C., and Zhang, L. (2019). Tracking objects from satellite videos: a velocity feature based correlation filter. *IEEE Trans. Geosci. Remote Sens.* 57, 7860–7871. doi: 10.1109/TGRS.2019.2916953

Voigtlaender, P., Luiten, J., Torr, P. H., and Leibe, B. (2020). "Siam r-CNN: visual tracking by re-detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 6578–6588.

Wang, N., Zhou, W., Tian, Q., Hong, R., Wang, M., and Li, H. (2018). "Multi-cue correlation filters for robust visual tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE), 4844–4853.

Wang, Q., Zhang, L., Bertinetto, L., Hu, W., and Torr, P. H. (2019). "Fast online object tracking and segmentation: a unifying approach," in *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition* (Long Beach, CA: IEEE), 1328–1338.

Wang, W., Han, Y., Deng, C., and Li, Z. (2022). Hyperspectral image classification via deep structure dictionary learning. *Remote Sens.* 14, 2266. doi: 10.3390/rs14092266

Xu, T., Feng, Z.-H., Wu, X.-J., and Kittler, J. (2019). Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking. *IEEE Trans. Image Process.* 28, 5596–5609. doi: 10.1109/TIP.2019.2919201

Zhu, P., Wen, L., Bian, X., Ling, H., and Hu, Q. (2018). Vision meets drones: a challenge. *arXiv preprint arXiv:1804.07437*. doi: 10.48550/arXiv.1804.07437