mathematics

MDPI

Article

Second-Order Spatial-Temporal Correlation Filters for Visual Tracking

Yufeng Yu [1], Long Chen [1], Haoyang He [2], Jianhui Liu [3], Weipeng Zhang [4] and Guoxia Xu [5,*]

[1] Department of Computer and Information Science, University of Macau, Macau 999078, China; yuyufeng220@163.com (Y.Y.); longchen@um.edu.mo (L.C.)
[2] Department of Statistics, Guangzhou University, Guangzhou 510006, China; hoeyeungho@163.com
[3] Jiangsu Province Key Lab on Image Processing and Image Communication, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; 1864700023@e.gzhu.edu.cn
[4] PLA Strategic Support Force, Beijing 450001, China; guagua_mitnick@163.com
[5] Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjovik, Norway
* Correspondence: gxxu.re@gmail.com

Abstract: Discriminative correlation filters (DCFs) have been widely used in visual object tracking, but often suffer from two problems: the boundary effect and temporal filtering degradation. To deal with these issues, many DCF-based variants have been proposed and have improved the accuracy of visual object tracking. However, these trackers only adopt first-order data-fitting information and have difficulty maintaining robust tracking in unconstrained scenarios, especially in the case of complex appearance variations. In this paper, by introducing a second-order data-fitting term to the DCF, we propose a second-order spatial–temporal correlation filter (SSCF) learning model. To be specific, the SSCF tracker both incorporates the first-order and second-order data-fitting terms into the DCF framework and makes the learned correlation filter more discriminative. Meanwhile, the spatial–temporal regularization was integrated to develop a robust model in tracking with complex appearance variations. Extensive experiments were conducted on the benchmarking databases CVPR2013, OTB100, DTB70, UAV123, and UAVDT-M. The results demonstrated that our SSCF can achieve competitive performance compared to the state-of-the-art trackers. When penalty parameter λ was set to 10^{-5}, our SSCF gained DP scores of 0.882, 0.868, 0.706, 0.676, and 0.928 on the CVPR2013, OTB100, DTB70, UAV123, and UAVDT-M databases, respectively.

Keywords: correlation filters; second-order fitting; visual tracking

MSC: 68T45

Citation: 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. https://doi.org/10.3390/math10050684

Academic Editors: Jianping Gou, Weihua Ou, Shaoning Zeng and Lan Du

Received: 10 January 2022
Accepted: 17 February 2022
Published: 22 February 2022

1. Introduction

Visual object tracking is a fundamental problem in the field of computer vision, which has a wide range of applications in human–computer interaction, video surveillance, unmanned driving, and so on. The task of visual object tracking always suffers from the challenges of appearance variations, such as illumination variation, fast motion, out-of-plane rotation, and in-plane rotation. To deal with these challenges, various innovative trackers have been proposed and achieved significant progress in tracking performance and robustness. Among these tracking methods, discriminative-filter-based trackers [1–5] have received significant attention due to their competitive performance.

The standard discriminative-correlation-filter (DCF)-based tracker treats the filter learning as a ridge regression problem, and the objective function can be transferred to the frequency domain by the fast Fourier transform (FFT) for the solution. Bolme et al. [6] first learned the correlation filter to perform the target tracking task and proposed a minimum output sum of squared error (MOSSE) model. The MOSSE trains the filter

by calculating the minimum actual and expected mean-squared errors of sequence images. Inspired by the MOSSE, Henriques et al. [7] considered that cyclic displacement could be used to replace random sampling to achieve dense sampling and proposed a theoretical framework to explore the effect of dense sampling. The proposed framework formulates a kernelized correlation filter to improve the tracking performance. Zhang et al. [8] adopted the Bayesian principle to build a spatial–temporal context model for tracking. However, these CF-based trackers only utilize single-channel features, which is not robust in the tracking scenarios with complex appearance variations. To tackle this issue, some CF-based methods [9–19] extract multiple features to learn the filters. The commonly used handcrafted features include the histogram of oriented gradients (HOG), color names (CNs), the local binary pattern (LBP), and scale-invariant feature transform (SIFT). These features describe the shape and color information of the targets. Trackers using multiple features are more robust to the fast movement and deformation variation of targets. For instance, Galoogahi et al. [17] employed multi-channel HOG descriptors in the frequency domain to extract HOG features for filter learning and proposed a multi-channel CF tracker (MCCF). Huang et al. [14] used hybrid color features to learn filters in which the compressed CN features and the HOG features based on the opponent color space were extracted, and principal component analysis was used to reduce the computational cost. Li et al. [12] integrated the raw pixel, HOG, and color label features into the DCF framework and presented an adaptive multiple feature tracker. Kumar et al. [19] exploited the LBP, color histogram, and pyramid of the histogram of gradients to model the object's appearance and developed an adaptive multi-cue particle filter method for real-time visual tracking.

Even though these DCF-based trackers using multi-channel features succeed to some extent, some aspects such as the redundancy of multi-channel features, the boundary effect, and data fitting have not been fully explored. To tackle these issues, many structural regularized DCF methods [20–26] have been presented. Zhu et al. [2] proposed an adaptive attribute-aware strategy to distinguish the importance of different channel features. Jain et al. [20] presented a channel graph regularized CF model by introducing a channel weighing strategy in which a channel regularizer was integrated into the CF framework to learn the channel weights. Xu et al. [22] proposed a channel selection scheme for multi-channel feature representations and adopted a low-rank approximation to learn filters in a low-dimensional manifold. In addition, many trackers propose a variety of strategies to solve the boundary effect. The SRDCF [23] incorporates a spatial regularizer into the DCF to deal with the problem caused by the periodic assumption. Li et al. [24] supplemented the temporal regularization term into the SRDCF tracker [23] and proposed a spatial–temporal regularization CF framework. To be specific, the STRCF integrates both temporal regularization and spatial regularization into the standard DCF model and can perform model updating and DCF learning simultaneously. As a result, the STRCF could be regarded as an approximation of the SRDCF with multiple samples and achieves better tracking performance than the SRDCF. The BACF [25] utilizes a cropping matrix to extract patches densely from the background and expands the search area at a low computational cost. Xu et al. [26] combined temporal consistency constraints and spatial feature selection to propose an adaptive DCF model in which the multi-channel filters can be learned in a low-dimensional manifold space. However, the aforementioned trackers only employ the first-order data-fitting information of the feature maps. In other words, such methods do not consider high-order data-fitting information for tracking.

On the basis of the above-mentioned analysis, we propose a novel CF-based tracker, the second-order spatial–temporal correlation filter (SSCF) learning model. We formulated our tracking algorithm by incorporating a second-order data-fitting term into the DCF framework, which helps to take full advantage of target features against surrounding background clutter. The main contributions of the SSCF are summarized as follows:

- We propose a new discriminative correlation filter model for visual tracking with complex appearance variations, unlike prior DCF-based trackers in which the first-

order data-fitting information is only used. We incorporated the second-order data fitting and spatial–temporal regularization into the DCF framework and developed a more robust tracker;
- An effective alternating-direction method-of-multipliers (ADMM)-based algorithm was used to solve the proposed tracking model;
- Extensive experiments on the benchmarking databases demonstrated that our SSCF can achieve competitive performance compared to the state-of-the-art trackers.

The remainder of this paper is organized as follows. Section 2 introduces the related work. Section 3 describes the detailed mathematical formulation of the proposed model and introduces the optimization algorithm. Section 4 reports the experimental results and the corresponding analysis. Finally, Section 5 draws the conclusions.

2. Related Work

In this section, we review mainly three categories of tracking methods, including trackers based on target detection, trackers based on clustering, and channel-reliability learning trackers.

Since target detection techniques [27–29] have attracted wide attention in the computer vision field, many trackers based on target detection have been proposed. Guan et al. [30] proposed a joint detection and tracking framework for object tracking in which the detection threshold was adaptively modified according to the information fed back to the detector by the tracker. Zhang et al. [31] employed a faster recurrent convolutional neural network to extract the candidate detection areas and proposed a multi-target tracking algorithm. In [32], Liu et al. combined motion detection with correlation filtering and presented a new model for object tracking. The presented model determines the object position via the weighted outputs of motion detection and the tracker. Considering that the existing kernelized correlation filter tracking methods fail to identify occlusion, Min et al. [33] adopted a detector to assist the occlusion judgment and improve the tracking performance.

Clustering-based algorithms [34,35] have been commonly used in pattern recognition and computer vision, such as image segmentation [36] and patten classification [37]. Inspired by this, many researchers use clustering algorithms to improve the performance of object tracking. For instance, Keuper et al. [38] combined motion segmentation with object tracking and presented a correlation co-clustering model to improve the performance. In [39], Li et al. developed an intuitionistic fuzzy clustering model for object tracking. Specifically, the local information of the targets is incorporated into the intuitionistic fuzzy clustering to improve the robustness. Considering that DBSCAN clustering does not require the number of clusters, He et al. [40] employed a DBSCAN clustering-based track-to-track fusion strategy for multi-target tracking.

Recently, the idea of different weights distinguishing the importance of different components has been widely used in pattern classification [41,42] and face recognition [43]. Similarly, some DCF-based channel-reliability learning trackers have been proposed to deal with the problem of model degradation. Du et al. [44] argued that different channels have different contributions in the tracking process and proposed a joint channel-reliability and correlation-filter learning model. The proposed tracker assigns each channel a weight to distinguish the different importance. To exploit the interaction between different channels, Jain et al. [20] assigned similar weights to similar channels to emphasize important channels and developed a channel attention model. Li et al. [45] argued that the existing trackers do not consider the complementary information of different channels and proposed a channel-feature integration method. All channels of each feature share an importance map to avoid overfitting. In [46], the authors introduced channel and spatial reliability to the DCF framework and employed the reliability scores to weight the per-channel filter responses. The experiments showed that the channel weights were able to improve the tracking performance. These methods principally focus on overcoming model degradation by incorporating channel reliability and enhance the discriminative performance to some extent.

3. The Proposed Model

3.1. Objective Function Construction

As mentioned above, the existing DCF-based methods only utilize first-order data-fitting information and ignore high-order data-fitting information for tracking, which cannot take full advantage of target features against surrounding background clutter and suffer from the stability–plasticity dilemma. To deal with these issues, we built a second-order spatial–temporal correlation-filter learning framework. Specifically, we incorporated a second-order data-fitting term and spatial–temporal regularization into the DCF framework and formulated a robust model. The objective function is able to be formulated as below.

We first denote the dataset $\mathbb{S} = \{\mathbb{X}_t\}_{t=1}^{T}$, and each frame $\mathbb{X}_t \in R^{M\times N\times K}$ contains K feature maps with a size of $M \times N$. $\mathbf{Y} \in R^{M\times N}$ is the Gaussian-shaped label. Our aim was to learn a multi-channel convolution filter $\mathbb{F} \in R^{M\times N\times K}$ by minimizing the following objective function:

$$\begin{aligned}\min_{\mathbb{F}} \frac{1}{2}\left\|\sum_{k=1}^{K}\mathbf{X}_t^k * \mathbf{F}^k - \mathbf{Y}\right\|_F^2 + \frac{1}{2}\sum_{k=1}^{K}\|\mathbf{W}\cdot\mathbf{F}^k\|_F^2 \\ +\frac{\lambda}{2}\left\|\sum_{k=1}^{K}\mathbf{X}_t^k * \mathbf{F}^k * \mathbf{X}_t^k - \mathbf{Y}\right\|_F^2 + \frac{\mu}{2}\|\mathbb{F}-\mathbb{F}_{t-1}\|_F^2\end{aligned} \tag{1}$$

where $*$ represents the convolution operator and $\cdot$ denotes the Hadamard product. $\mathbf{W}$ is the spatial regularization matrix, and $\mathbb{F}_{t-1}$ is the correlation filter used in the $t-1$-th frame. λ and μ are penalty parameters. The first term is the first-order data-fitting term, which is a generic formulation for learning the filter in DCF-based trackers. The second term is the spatial regularizer to solve the boundary effect. The third term is the second-order data-fitting term, which can be helpful to make full use of discriminative target features. The last term is the temporal regularizer to force the current frame filter close to the previous one, which helps to prevent the effect caused by the corrupted samples.

3.2. Optimization Algorithm

It can be noted that the objective function in Equation (1) is convex, and the minimization problem can be solved by the ADMM algorithm. To be specific, we introduced an auxiliary variable $\mathbb{G} \in R^{M\times N\times K}$ by restricting $\mathbb{F} = \mathbb{G}$ and constructed the augmented Lagrangian form of Equation (1) as:

$$\begin{aligned}L(\mathbb{F},\mathbb{G},\mathbb{S}) = &\frac{1}{2}\left\|\sum_{k=1}^{K}\mathbf{X}_t^k * \mathbf{F}^k - \mathbf{Y}\right\|_F^2 + \frac{1}{2}\sum_{k=1}^{K}\|\mathbf{W}\cdot\mathbf{G}^k\|_F^2 \\ &+\frac{\lambda}{2}\left\|\sum_{k=1}^{K}\mathbf{X}_t^k * \mathbf{F}^k * \mathbf{X}_t^k - \mathbf{Y}\right\|_F^2 + \frac{\mu}{2}\|\mathbb{F}-\mathbb{F}_{t-1}\|_F^2 \\ &+\frac{\gamma}{2}\sum_{k=1}^{K}\|\mathbf{F}^k-\mathbf{G}^k\|_F^2 + \sum_{k=1}^{K}Tr((\mathbf{F}^k-\mathbf{G}^k)^T\mathbf{S}^k)\end{aligned} \tag{2}$$

where $\mathbb{S} = [\mathbf{S}^1, \mathbf{S}^2, \cdots, \mathbf{S}^K] \in R^{M\times N\times K}$ is the Lagrange multiplier and γ is the stepsize. Assuming $\mathbb{H} = \frac{1}{\gamma}\mathbb{S}$, Equation (2) can be written as:

$$L(\mathbb{F},\mathbb{G},\mathbb{H}) = \frac{1}{2}\left\|\sum_{k=1}^{K}\mathbf{X}_t^k * \mathbf{F}^k - \mathbf{Y}\right\|_F^2 + \frac{1}{2}\sum_{k=1}^{K}\|\mathbf{W}\cdot\mathbf{G}^k\|_F^2 + \frac{\lambda}{2}\left\|\sum_{k=1}^{K}\mathbf{X}_t^k * \mathbf{F}^k * \mathbf{X}_t^k - \mathbf{Y}\right\|_F^2 + \frac{\mu}{2}\|\mathbb{F}-\mathbb{F}_{t-1}\|_F^2 + \frac{\gamma}{2}\sum_{k=1}^{K}\left\|\mathbf{F}^k - \mathbf{G}^k + \mathbf{H}^k\right\|_F^2 \tag{3}$$

The optimization problem can be divided into several subproblems as follows.

$$\mathbb{F}^{(l+1)} = \arg\min_{\mathbb{F}}\left\|\sum_{k=1}^{K}\mathbf{X}_t^k * \mathbf{F}^k - \mathbf{Y}\right\|_F^2 + \left\|\sum_{k=1}^{K}\mathbf{X}_t^k * \mathbf{F}^k * \mathbf{X}_t^k - \mathbf{Y}\right\|_F^2 + \gamma\sum_{k=1}^{K}\left\|\mathbf{F}^k - \mathbf{G}^k + \mathbf{H}^k\right\|_F^2 + \mu\|\mathbb{F}-\mathbb{F}_{t-1}\|_F^2 \tag{4}$$

$$\mathbb{G}^{(l+1)} = \arg\min_{\mathbb{G}}\sum_{k=1}^{K}\|\mathbf{W}\cdot\mathbf{G}^k\|_F^2 + \gamma\sum_{k=1}^{K}\left\|\mathbf{F}^k - \mathbf{G}^k + \mathbf{H}^k\right\|_F^2 \tag{5}$$

$$\mathbb{H}^{(l+1)} = \mathbb{H}^{(l)} + \mathbb{F}^{(l+1)} - \mathbb{G}^{(l+1)} \tag{6}$$

Then, we can alternatively solve each subproblem as follows:

Solving $\mathbb{F}$: According to Parseval's theorem, the subproblem in Equation (4) can be formulated in the Fourier domain as:

$$\arg\min_{\hat{\mathbb{F}}}\left\|\sum_{k=1}^{K}\hat{\mathbf{X}}_t^k\cdot\hat{\mathbf{F}}^k - \hat{\mathbf{Y}}\right\|_F^2 + \lambda\left\|\sum_{k=1}^{K}\hat{\mathbf{X}}_t^k\cdot\hat{\mathbf{F}}^k\cdot\hat{\mathbf{X}}_t^k - \hat{\mathbf{Y}}\right\|_F^2 + \gamma\sum_{k=1}^{K}\left\|\hat{\mathbf{F}}^k - \hat{\mathbf{G}}^k + \hat{\mathbf{H}}^k\right\|_F^2 + \mu\|\hat{\mathbb{F}}-\hat{\mathbb{F}}_{t-1}\|_F^2 \tag{7}$$

Here, $\hat{\mathbb{F}}$ represents the discrete Fourier transform (DFT) of $\mathbb{F}$. From Equation (7), it can be noted that the i-th row and the j-th element of $\hat{\mathbf{Y}}$ only depend on the i-th row and the j-th element of $\hat{\mathbb{F}}$ and $\hat{\mathbb{X}}_t$ across all K channels. Assume $v_{ij}(\mathbb{F})$ is a K-dimensional vector that contains the i-th row and the j-th elements of $\mathbb{F}$ along all K channels. Optimizing the problem in Equation (7) is equivalent to solving the following MN subproblems:

$$\arg\min_{v_{ij}(\hat{\mathbb{F}})} \parallel v_{ij}(\hat{\mathbb{X}}_t)^T v_{ij}(\hat{\mathbb{F}}) - \hat{y}_{ij} \parallel_2^2 + \mu \parallel v_{ij}(\hat{\mathbb{F}}) - v_{ij}(\hat{\mathbb{F}}_{t-1}) \parallel_2^2 + \lambda \parallel (v_{ij}(\hat{\mathbb{X}}_t)\cdot v_{ij}(\hat{\mathbb{X}}_t))^T v_{ij}(\hat{\mathbb{F}}) - \hat{y}_{ij} \parallel_2^2 + \gamma \parallel v_{ij}(\hat{\mathbb{F}}) - v_{ij}(\hat{\mathbb{G}}) + v_{ij}(\hat{\mathbb{H}}) \parallel_2^2 \tag{8}$$

where $i = 1,\cdots,M$ and $j = 1,\cdots,N$.

Taking the derivative of Equation (8) with respect to $v_{ij}(\hat{\mathbb{F}})$ as zero, we have:

$$v_{ij}(\hat{\mathbb{F}}) = (\mathbf{Q} + (\gamma+\mu)\mathbf{I})^{-1}\mathbf{z} \tag{9}$$

Here, $\mathbf{Q} = v_{ij}(\hat{\mathbb{X}}_t)v_{ij}(\hat{\mathbb{X}}_t)^T + \lambda(v_{ij}(\hat{\mathbb{X}}_t)\cdot v_{ij}(\hat{\mathbb{X}}_t))(v_{ij}(\hat{\mathbb{X}}_t)\cdot v_{ij}(\hat{\mathbb{X}}_t))^T$ and $\mathbf{z} = v_{ij}(\hat{\mathbb{X}}_t)\hat{y}_{ij} + \mu v_{ij}(\hat{\mathbb{F}}_{t-1}) + \lambda(v_{ij}(\hat{\mathbb{X}}_t)\cdot v_{ij}(\hat{\mathbb{X}}_t)) + \gamma v_{ij}(\hat{\mathbb{G}}) - \gamma v_{ij}(\hat{\mathbb{H}})$.

Solving $\mathbb{G}$: From Equation (5), each element of $\mathbb{G}$ is able to be updated independently, and we adopted the same strategy as solving $\mathbb{F}$. Assume $v_{ij}(\mathbb{G})$ is a K-dimensional vector

that contains the i-th row and the j-th elements of $\mathbb{G}$ along all K channels. Optimizing the problem in Equation (5) is equivalent to solving the following MN subproblems:

$$\arg\min_{v_{ij}(\mathbb{G})} w_{ij}^2 \parallel v_{ij}(\mathbb{G}) \parallel_2^2 + \gamma \parallel v_{ij}(\mathbb{F}) - v_{ij}(\mathbb{G}) + v_{ij}(\mathbb{H}) \parallel_2^2 \tag{10}$$

Taking the derivative of Equation (10) with respect to $v_{ij}(\mathbb{G})$ as zero, we have:

$$v_{ij}(\mathbb{G}) = (\mathbf{P}^T\mathbf{P} + \gamma\mathbf{I})^{-1}(\gamma v_{ij}(\mathbb{F}) + \gamma v_{ij}(\mathbb{H})) \tag{11}$$

where $\mathbf{P}$ is a diagonal matrix and each diagonal element is w_{ij}.

Updating $\mathbb{H}$: Let $v_{ij}(\mathbb{H})$ be a K-dimensional vector that contains the i-th row and the j-th elements of $\mathbb{G}$ along all K channels. In the $l+1$-th iteration of the ADMM, the Lagrange multiplier vector $v_{ij}(\mathbb{H})$ can be updated as follows:

$$v_{ij}(\mathbb{H})^{(l+1)} = v_{ij}(\mathbb{H})^{(l)} + v_{ij}(\mathbb{F})^{(l+1)} - v_{ij}(\mathbb{G})^{(l+1)} \tag{12}$$

The details of the optimization procedure can be seen in Algorithm 1.

Algorithm 1 SSCF algorithm

Input: Feature maps $\mathbb{X}_t$, Gaussian-shaped label $\mathbf{Y}$, previous correlation filters $\mathbb{F}_{t-1}$, spatial regularization matrix $\mathbf{W}$, initial values $\mathbb{G}^{(0)}$ and $\mathbb{H}^{(0)}$.
Output: Estimated correlation filters $\mathbb{F}$.

1: **repeat** Step 2–Step 5
2: Update $v_{ij}(\hat{\mathbb{F}})^{(l+1)}$ via Equation (9);
3: Update $v_{ij}(\mathbb{G})^{(l+1)}$ via Equation (11);
4: Update $v_{ij}(\mathbb{H})^{(l+1)}$ via Equation (12);
5: $l = l + 1$;
6: **Until** $v_{ij}(\hat{\mathbb{F}})$, $v_{ij}(\mathbb{G})$, $v_{ij}(\mathbb{H})$ have converged;
7: Obtain correlation filters $\mathbb{F}$ by applying the inverse DFT.

3.3. Computational Complexity

In this subsection, we discuss the computational complexity of the presented SSCF. As shown in Section 3.2, we divided the optimization problem into several subproblems. According to the Parseval theorem and the ADMM algorithm, the complexity of solving $\mathbf{F}$ is $O(KMN)$ in each iteration. Taking the DFT and inverse DFT into account, the computational complexity of solving $\mathbf{F}$ is $O(KMN\log(MN))$. Moreover, the complexity of subproblems $\mathbf{H}$ and $\mathbf{G}$ is $O(KMN)$. Suppose the number of iteration is T: the whole computational complexity of the proposed SSCF is $O(TKMN(\log(MN)+1))$. In view of this, the speed of our tracker is not fast.

4. Experiment Results and Analysis

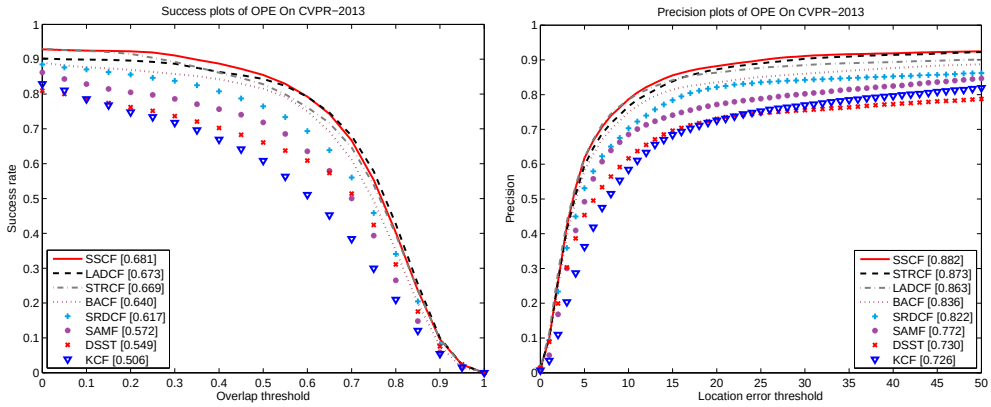
This section provides the experiments to validate the superiority of the presented SSCF in target tracking. To evaluate the performance of the proposed model, we compared it with the state-of-the-art trackers, including spatially regularized discriminative correlation filters (SRDCFs) [23], kernelized correlation filters (KCFs) [47], spatial–temporal regularized correlation filters (STRCFs) [24], background-aware correlation filters (BACFs) [25], learning adaptive discriminative correlation filters (LADCFs) [26], discriminative scale space tracking (DSST) [48], the scale-adaptive with multiple features tracker (SAMF) [12], ECOHC [49], ARCF-HC [50], the MSCF [51], and AutoTrack [52]. These experiments were

conducted on the CVPR2013 [53], OTB50 [54], OTB100 [54], DTB70 [55], UAV123 [56], and UAVDT-M databases [57].

In the experiments, our tracker was implemented using MATLAB R2017a on a computer with an i7-8700K processor (3.7GHz) with 48GB RAM. λ was set to 10^{-5}, and other parameters were set to the same values as the STRCF. The histogram of oriented gradients (HOG) features were used to conduct the comparative experiments. In addition, we followed the one-pass evaluation (OPE) protocol [53] to evaluate the performance of different trackers. The success and precision plots are reported based on the bounding box overlap and center location error. The AUC is the area under the curve of the success plot, and the distance precision (DP) is the percentage of the location errors within 20 px.

4.1. Results on the CVPR2013 Database

The CVPR2013 database contains 50 fully annotated video sequences with 11 different attributes, such as background clutter, low resolution, occlusion, and out of view. The overall performance, which is summarized by the success and precision plots, is listed in Figure 1. It can be observed that the proposed SSCF achieved the top-ranking results. The area under the curve (AUC) and distance precision (DP) scores were 0.681 and 0.882, respectively. Specifically, the AUC and DP scores of SSCF were higher by 1.2% and 0.9% than the STRCF. This indicates that incorporating the second-order data-fitting term is effective at improving the tracking performance.



(**a**) (**b**)

Figure 1. Success plots (**a**) and precision plots (**b**) of the proposed SSCF and other trackers on the CVPR2013 database.

To evaluate the robustness of the proposed SSCF on different attributes, we constructed subsets with different dominant attributes for the experiments. The 11 challenging factors were background clutter (BC), low resolution (LR), illumination variation (IV), motion blur (MB), out of view (OV), fast motion (FM), deformation (DEF), occlusion (OCC), out-of-plane rotation (OPR), scale variation (SV), and in-plane rotation (IPR). Table 1 shows the AUC and DP scores of the proposed SSCF and the other trackers on the 11 attributes on the CVPR2013 database. Despite not all scores of the proposed SSCF being the highest, our method achieved the best robustness. Especially for the AUC scores on the different attributes, our SSCF outperformed the other trackers, except LADCF.

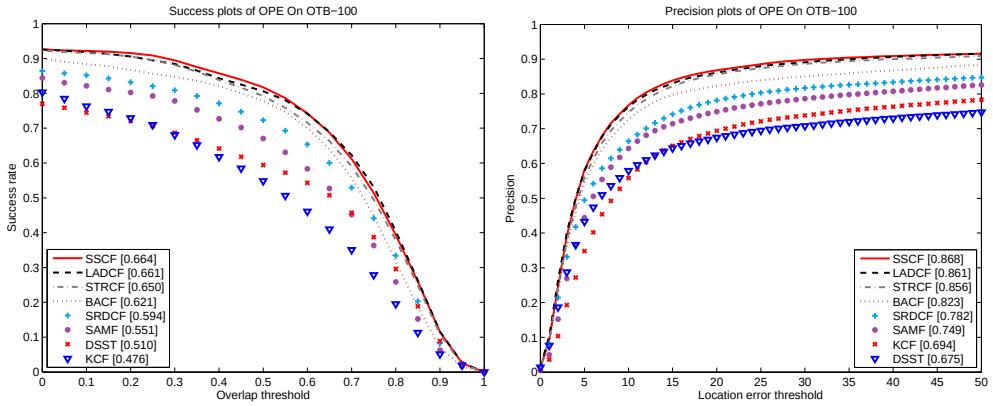
4.2. Results on the OTB100 Database

OTB100 is a database containing 100 challenging video sequences, and these sequences consist of more than 28,000 fully annotated frames. The results of the success and precision plots for all trackers are shown in Figure 2. From the figure, the proposed SSCF outperformed all the competing trackers in its overall performance. Our tracker achieved 0.664 and 0.868 in terms of the AUC and DP scores, respectively.

We also provide the attribute-based evaluation to validate the robustness of our SSCF. The AUC and DP scores of all trackers on the 11 different attributes are reported in Table 2. From the DP scores listed in the table, the proposed SSCF outperformed all competing trackers on eight attributes. In terms of the AUC scores, our tracker performed better than the other trackers on seven attributes. On other attributes, the SSCF was among the top-three trackers. These results demonstrate that our SSCF is more robust than the other trackers.

Table 1. The area under the curve (AUC) and distance precision (DP) scores of the proposed SSCF and the other trackers on different attributes on the CVPR2013 database. The top-three methods on each attribute are denoted by different colors: red, blue, and green. That is, red represents the best performance, blue represents the second best, and green represents the third best (AUC/DP).

Attributes	DSST [48]	KCF [47]	SAMF[12]	SRDCF [23]	BACF [25]	STRCF [24]	LADCF [26]	SSCF
FM	0.413/0.485	0.435/0.559	0.460/0.568	0.541/0.691	0.583/0.766	0.572/0.697	0.591/0.728	0.604/0.754
BC	0.517/0.694	0.535/0.753	0.520/0.676	0.587/0.803	0.631/0.833	0.625/0.850	0.592/0.783	0.641/0.840
DEF	0.492/0.633	0.512/0.702	0.604/0.775	0.609/0.811	0.644/0.832	0.639/0.854	0.657/0.852	0.680/0.885
IPR	0.555/0.753	0.484/0.702	0.512/0.692	0.550/0.739	0.622/0.824	0.621/0.802	0.612/0.785	0.633/0.826
IV	0.551/0.711	0.477/0.699	0.498/0.655	0.557/0.727	0.600/0.788	0.599/0.779	0.599/0.752	0.630/0.799
LR	0.378/0.682	0.272/0.629	0.376/0.709	0.471/0.767	0.406/0.659	0.540/0.777	0.580/0.776	0.510/0.744
MB	0.433/0.504	0.462/0.589	0.428/0.507	0.560/0.719	0.609/0.790	0.566/0.681	0.579/0.702	0.626/0.778
OCC	0.523/0.690	0.499/0.724	0.598/0.816	0.610/0.815	0.612/0.797	0.646/0.854	0.673/0.869	0.673/0.872
OPR	0.529/0.723	0.485/0.710	0.549/0.749	0.586/0.796	0.620/0.822	0.651/0.863	0.657/0.850	0.667/0.875
OV	0.462/0.511	0.550/0.650	0.555/0.636	0.555/0.680	0.553/0.706	0.632/0.728	0.633/0.720	0.652/0.748
SV	0.546/0.738	0.427/0.679	0.507/0.723	0.587/0.778	0.584/0.765	0.647/0.836	0.649/0.821	0.639/0.823



(**a**) (**b**)

Figure 2. Success plots (**a**) and precision plots (**b**) of the proposed SSCF and the other trackers on the OTB100 database.

Table 2. The area under the curve (AUC) and distance precision (DP) scores of the proposed SSCF and the other trackers on different attributes on the OTB100 database. The top-three methods on each attribute are denoted by different colors: red, blue, and green. That is, red represents the best performance, blue represents the second best, and green represents the third best (AUC/DP).

Attributes	DSST [48]	KCF [47]	SAMF [12]	SRDCF [23]	BACF [25]	STRCF [24]	LADCF [26]	SSCF
FM	0.439/0.540	0.457/0.617	0.502/0.649	0.586/0.749	0.600/0.791	0.617/0.780	0.625/0.790	0.635/0.803
BC	0.521/0.703	0.509/0.731	0.532/0.705	0.584/0.777	0.643/0.861	0.648/0.872	0.637/0.830	0.679/0.884
DEF	0.414/0.532	0.427/0.600	0.500/0.671	0.533/0.715	0.599/0.802	0.596/0.825	0.595/0.812	0.613/0.835
IPR	0.496/0.681	0.468/0.698	0.515/0.717	0.535/0.729	0.583/0.787	0.593/0.794	0.601/0.810	0.602/0.817
IV	0.551/0.709	0.468/0.699	0.524/0.697	0.600/0.770	0.632/0.821	0.640/0.819	0.649/0.808	0.666/0.833
LR	0.370/0.649	0.290/0.671	0.425/0.766	0.514/0.765	0.516/0.797	0.579/0.843	0.614/0.850	0.576/0.834
MB	0.458/0.551	0.456/0.594	0.519/0.648	0.580/0.739	0.590/0.762	0.637/0.797	0.646/0.807	0.672/0.845
OCC	0.447/0.587	0.442/0.626	0.536/0.722	0.551/0.719	0.576/0.743	0.606/0.797	0.644/0.830	0.638/0.827
OPR	0.466/0.637	0.447/0.665	0.530/0.728	0.542/0.729	0.584/0.785	0.619/0.836	0.632/0.838	0.632/0.850
OV	0.383/0.481	0.418/0.540	0.495/0.662	0.464/0.601	0.521/0.721	0.585/0.766	0.613/0.815	0.600/0.777
SV	0.468/0.638	0.400/0.642	0.498/0.713	0.562/0.746	0.571/0.769	0.632/0.842	0.636/0.836	0.634/0.843

4.3. Results on the OTB50 Database

Figure 3 lists the success plots comparing the presented method on OTB50 with the existing trackers. The overall performance is summarized in Figure 3a. It can be seen that the proposed SSCF had the best success rates. The success plots of all trackers on the 11 different attributes are shown in Figure 3b–l. The proposed SSCF outperformed the existing trackers on eight attributes, i.e., fast motion, background clutter, motion blur, illumination variation, in-plane rotation, occlusion, out-of-plane rotation, and out of view. Our SSCF incorporates the second-order data fitting and spatial–temporal regularization into the DCF framework to develop a robust tracking pattern. The tracking results of the SSCF on the other three attributes were among the top two. This also demonstrates the effectiveness and robustness of our tracker.

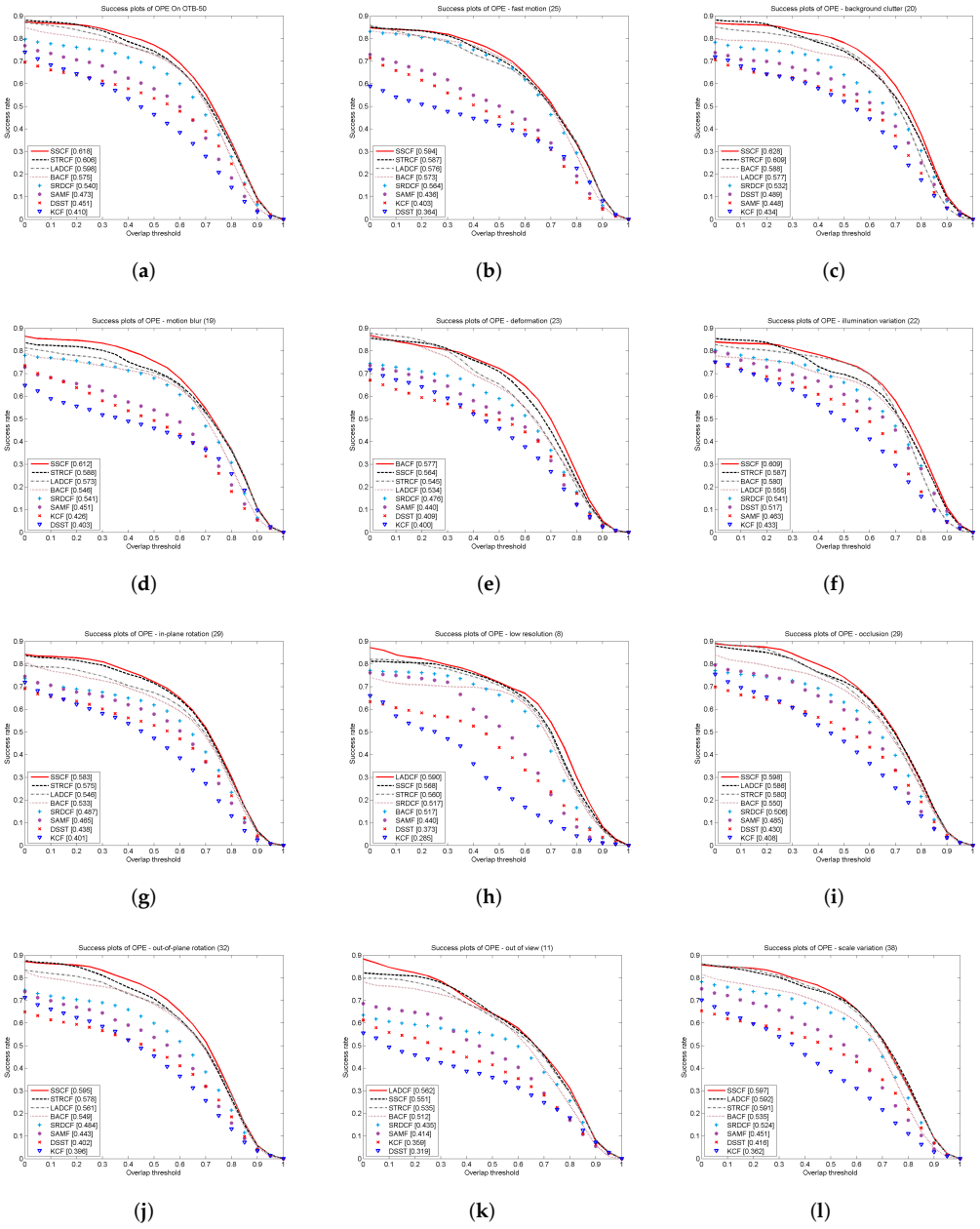


Figure 3. Success plots of the proposed SSCF and the other trackers on the OTB50 database. (**a**) Overall performance; (**b–l**) success plots on the 11 different attributes.

4.4. Results on the DTB70 Database

Figures 4 and 5 show the success plots and precision plots comparing the presented method on the DTB70 database with the existing trackers. The overall performance is summarized in Figures 4a and 5a. It is observed that our SSCF achieved the best results in the overall performance. The success plots and precision plots of all trackers on the 11 different attributes are shown in Figures 4b–l and 5b–l. Our SSCF outperformed the existing trackers on nine attributes except motion blur and low resolution.

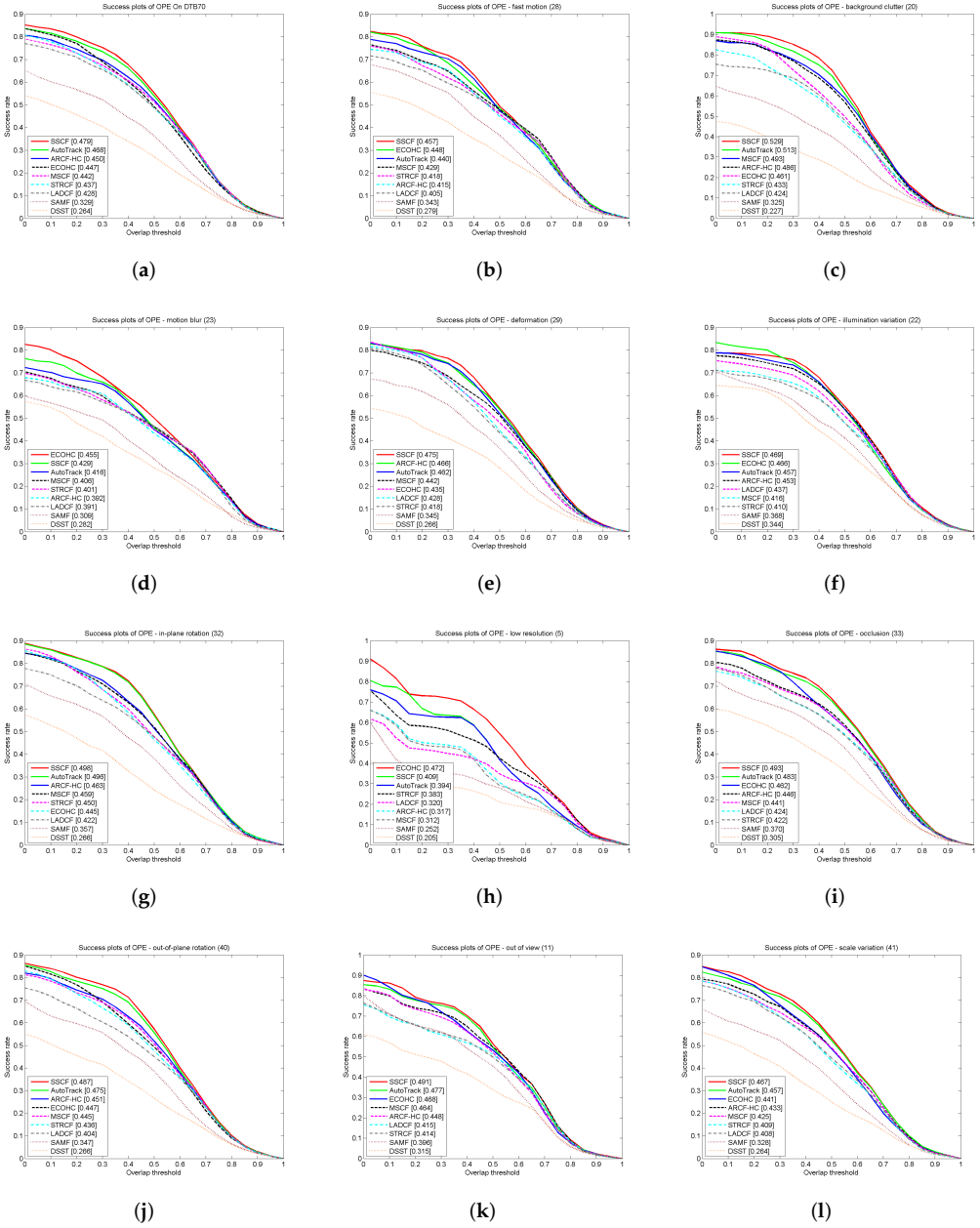


Figure 4. Success plots of the proposed SSCF and the other trackers on the DTB70 database. (**a**) Overall performance; (**b–l**) success plots on the 11 different attributes.

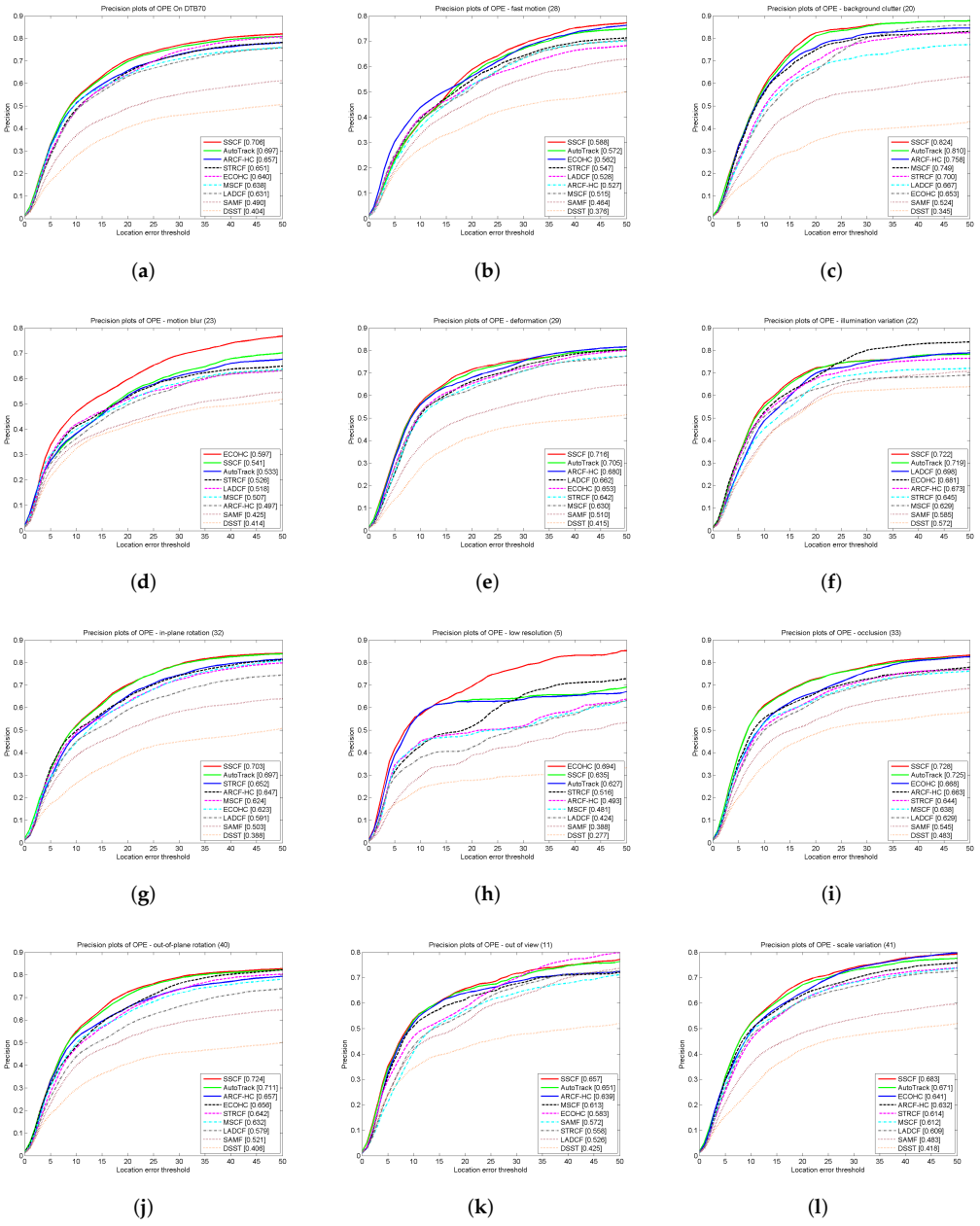


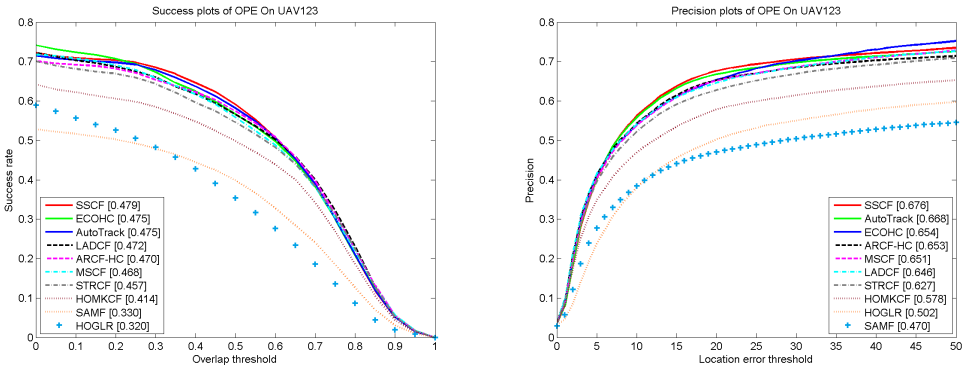
Figure 5. Precision plots of the proposed SSCF and the other trackers on the DTB70 database. (**a**) Overall performance; (**b**–**l**) precision plots on the 11 different attributes.

4.5. Results on the UAV123 Database

The UAV123 dataset contains 123 video sequences, which is the most commonly used and most comprehensive dataset for UAV tracking. The overall performance, which is summarized by success and precision plots, is listed in Figure 6. It can be observed that the proposed SSCF achieved the top-ranking results. The area under the curve (AUC) and distance precision (DP) scores were 0.479 and 0.676, respectively.

In order to visually show the performance of the proposed SSCF in the tracking process, we selected three different types of video sequences, namely person, boat, and car sequences, to conduct the experiments. As shown in Figure 7, each column corresponds to three frames of the images, and the images were randomly selected from the video sequences. The comparative methods were five trackers, including our SSCF, AutoTrack, the MSCF, the STRCF, and the LADCF, marked in green, red, blue, yellow, and orange,

respectively. It can be seen that our SSCF always tracked the correct target and had the best performance. The STRCF and LADCF were not robust in tracking the small targets.



(**a**) (**b**)

Figure 6. Success plots (**a**) and precision plots (**b**) of the proposed SSCF and the other trackers on the UAV123 database.

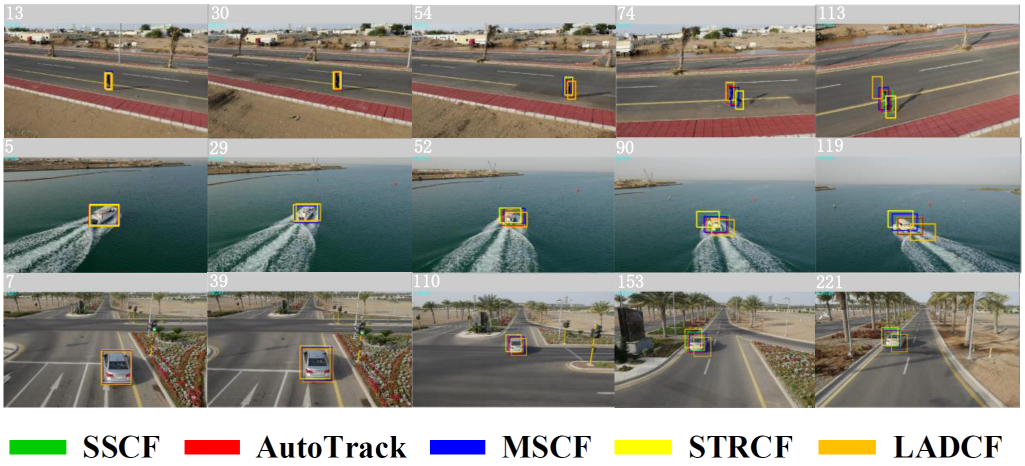


Figure 7. The qualitative analysis of different trackers on three video sequences.

4.6. Results on the UAVDT-M Database

In this section, we compare our SSCF with the existing methods on the UAVDT-M database. We also report the running speed of these methods. The running speed was measured in frames per second (FPS). Table 3 shows the comparison results. It can be observed that our SSCF achieved better performance than the existing trackers. The area under the curve (AUC) and distance precision (DP) scores were 0.667 and 0.928, respectively. However, It should be pointed out that the performance improvement of our tracker came at the expense of speed reduction.

Table 3. The area under the curve (AUC), distance precision (DP) scores, and FPS of the proposed SSCF and other trackers on the UAVDT-M database.

Methods	SSCF	AutoTrack [52]	MSCF [51]	STRCF [24]	ARCF-HC [50]	LADCF [26]	ECOHC [49]	DSST [48]
DP	0.928	0.917	0.913	0.904	0.902	0.895	0.891	0.878
AUC	0.667	0.655	0.642	0.625	0.636	0.614	0.602	0.530
FPS	3.8	65.4	37.6	9.3	15.3	18.2	15.9	100.7

5. Conclusions

In this paper, we proposed a new model called the second-order spatial–temporal correlation filter (SSCF) for visual object tracking. The SSCF is a DCF framework of

combining the second-order data-fitting term and spatial–temporal regularization. To solve the proposed model, we divided the optimization problem into several subproblems and adopted the ADMM algorithm to solve each subproblem. By taking full advantage of the second-order data-fitting information, the SSCF becomes more discriminative and robust in addressing complex tracking situations. Extensive experiments on the benchmarking databases demonstrated that our SSCF can achieve competitive performance compared to the state-of-the-art trackers.

It can be noted that the presented SSCF achieved better tracking results than the existing trackers on most of the attributes, but it was not robust on a few attributes, such as low resolution and occlusion. Recently, occlusion-processing methods have been presented in face recognition such as occlusion dictionary learning [58,59] and the occlusion-invariant model [60]. Can these occlusion processing methods be used for object tracking with occlusion? If the answer is yes, how can we design a new model to enhance the performance? It also should be pointed out that the performance improvement of our tracker came at the expense of speed reduction. How to improve the running speed of our SSCF is an important problem. In addition, although the proposed SSCF achieved better results than the existing methods, the accuracy was not high when tracking small targets. Self-paced learning has been widely used in computer vision and machine learning [61]. Combining self-paced learning and filter learning could potentially yield better performance in tracking small targets. In future work, we will focus on these topics.

Author Contributions: Conceptualization, Y.Y. and G.X.; methodology, Y.Y. and G.X.; software, H.H. and J.L.; validation, L.C., H.H., W.Z. and G.X.; writing—original draft preparation, Y.Y.; writing—review and editing, Y.Y., L.C., J.L., W.Z. and G.X.; supervision, L.C. and G.X.; funding acquisition, L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Science and Technology Development Fund, Macau SAR (File no. 0119/2018/A3), in part by the National Natural Science Foundation of China under Grant 62006056, in part by the Natural Science Foundation of Guangdong Province under Grant 2019A1515011266, in part by National Statistical Science Research Project of China under Grant 2020LY090, and in part by Science and Technology Planning Project of Guangzhou under Grant 202102020699.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: We greatly thank the Reviewers and Editors for the insightful comments and suggestions.

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

References

1. Yang, J.; Tang, W.; Ding, Z. Long-Term Target Tracking of UAVs Based on Kernelized Correlation Filter. *Mathematics* **2021**, *9*, 3006. [CrossRef]
2. Zhu, X.-F.; Wu, X.-J.; Xu, T.; Feng, Z.; Kittler, J. Robust visual object tracking via adaptive attribute-aware discriminative correlation filters. *IEEE Trans. Multimed.* **2021**, *24*, 1–13. [CrossRef]
3. Deng, C.; He, S.; Han, Y.; Zhao, B. Learning dynamic spatial–temporal regularization for uav object tracking. *IEEE Signal Process. Lett.* **2021**, *28*, 1230–1234. [CrossRef]
4. Yang, H.; Wang, J.; Miao, Y.; Yang, Y.; Zhao, Z.; Wang, Z.; Sun, Q.; Wu, D.O. Combining Spatio-Temporal Context and Kalman Filtering for Visual Tracking. *Mathematics* **2019**, *7*, 1059. [CrossRef]
5. Fang, S.; Ma, Y.; Li, Z.; Zhang, B. A visual tracking algorithm via confidence-based multi-feature correlation filtering. *Multimed. Tools Appl.* **2021**, *80*, 23963–23982. [CrossRef]
6. 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.
7. 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.

8. Zhang, K.; Zhang, L.; Liu, Q.; Zhang, D.; Yang, M.-H. Fast visual tracking via dense spatio-temporal context learning. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 127–141.
9. Wang, Y.; Luo, X.; Ding, L.; Wu, J.; Fu, S. Robust visual tracking via a hybrid correlation filter. *Multimed. Tools Appl.* **2019**, *78*, 31633–31648. [CrossRef]
10. 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.
11. Zhu, H.; Han, Y.; Wang, Y.; Yuan, G. Hybrid cascade filter with complementary features for visual tracking. *IEEE Signal Process. Lett.* **2021**, *28*, 86–90. [CrossRef]
12. Li, Y.; Zhu, J. A scale adaptive kernel correlation filter tracker with feature integration. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Cham, Switzerland, 2014; pp. 254–265.
13. Javed, S.; Mahmood, A.; Dias, J.; Seneviratne, L.; Werghi, N. Hierarchical spatiotemporal graph regularized discriminative correlation filter for visual object tracking. *IEEE Trans. Cybern.* **2021**. [CrossRef]
14. Huang, Y.; Zhao, Z.; Wu, B.; Mei, Z.; Gao, G. Visual object tracking with discriminative correlation filtering and hybrid color feature. *Multimedia Tools Appl.* **2019**, *78*, 34725–34744. [CrossRef]
15. 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, Amsterdam, The Netherlands, 8–16 October 2016; pp. 1430–1438.
16. Zhu, H.; Peng, H.; Xu, G.; Deng, L.; Cheng, Y.; Song, A. Bilateral weighted regression ranking model with spatial–temporal correlation filter for visual tracking. *IEEE Trans. Multimed.* **2021**. [CrossRef]
17. Galoogahi, H.K.; Sim, T.; Lucey, S. Multi-channel correlation filters. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 3072–3079.
18. Han, Y.; Deng, C.; Zhao, B.; Zhao, B. Spatial-temporal context-aware tracking. *IEEE Signal Process. Lett.* **2019**, *26*, 500–504. [CrossRef]
19. Kumar, A.; Walia, G.S.; Sharma, K. Real-time visual tracking via multi-cue based adaptive particle filter framework. *Multimed. Tools Appl.* **2020**, *79*, 20639–20663. [CrossRef]
20. Jain, M.; Tyagi, A.; Subramanyam, A.V.; Denman, S.; Sridharan, S.; Fookes, C. Channel graph regularized correlation filters for visual object tracking. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 715–729. [CrossRef]
21. Fu, C.; Xu, J.; Lin, F.; Guo, F.; Zhang, Z. Object saliency-aware dual regularized correlation filter for real-time aerial tracking. *IEEE Trans. Geosci. Remote. Sens.* **2020**, *58*, 8940–8951. [CrossRef]
22. Xu, T.; Feng, Z.-H.; Wu, X.-J.; Kittler, J. Joint group feature selection and discriminative filter learning for robust visual object tracking. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 7950–7960.
23. 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, Santiago, Chile, 7–13 December 2015; pp. 4310–4318.
24. Li, F.; Tian, C.; Zuo, W.; Zhang, L.; Yang, M.-H. 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; 4904–4913.
25. 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.
26. Xu, T.; Feng, Z.-H.; Wu, X.-J.; Kittler, J. Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking. *IEEE Trans. Image Process.* **2019**, *28*, 5596–5609. [CrossRef]
27. Deng, L.; Zhang, J.; Xu, G.; Zhu, H. Infrared small target detection via adaptive m-estimator ring top-hat transformation. *Pattern Recognit.* **2021**, *112*, 1–9. [CrossRef]
28. You, X.; Li, Q.; Tao, D.; Ou, W.; Gong, M. Local metric learning for exemplar-based object detection. *IEEE Trans. Circuits And Systems Video Technol.* **2014**, *24*, 1265–1276.
29. Zhu, H.; Ni, H.; Liu, S.; Xu, G.; Deng, L. Tnlrs: Target-aware non-local low-rank modeling with saliency filtering regularization for infrared small target detection. *IEEE Trans. Image Process.* **2020**, *29*, 9546–9558. [CrossRef]
30. Guan, Y.; Wang, Y. Joint detection and tracking scheme for target tracking in moving platform. In Proceedings of the IEEE Radar Conference (RadarConf20), Florence, Italy, 21–25 September 2020; pp. 1–4.
31. Zhang, L.; Fang, Q. Multi-target tracking based on target detection and mutual information. In Proceedings of the Chinese Control and Decision Conference (CCDC), Kunming, China, 22–24 May 2020; pp. 1242–1245.
32. Liu, C.; Gong, J.; Zhu, J.; Zhang, J.; Yan, Y. Correlation filter with motion detection for robust tracking of shape-deformed targets. *IEEE Access* **2020**, *8*, 89161–89170. [CrossRef]
33. Min, Y.; Wei, Z.; Tan, K. A detection aided multi-filter target tracking algorithm. *IEEE Access* **2019**, *7*, 71616–71626. [CrossRef]
34. Ou, W.; Yu, S.; Li, G.; Lu, J.; Zhang, K.; Xie, G. Multi-view non-negative matrix factorization by patch alignment framework with view consistency. *Neurocomputing* **2016**, *204*, 116–124. [CrossRef]
35. Long, Z.Z.; Xu, G.; Du, J.; Zhu, H.; Yu, Y.F. Flexible subspace clustering: A joint feature selection and k-means clustering framework. *Big Data Res.* **2021**, *23*, 1–9. [CrossRef]

36. Mishro, P.K.; Agrawal, S.; Panda, R.; Abraham, A. A novel type-2 fuzzy c-means clustering for brain mr image segmentation. *IEEE Trans. Cybern.* **2021**, *51*, 3901–3912. [CrossRef] [PubMed]
37. Ayo, F.E.; Folorunso, O.; Ibharalu, F.T.; Osinuga, I.A.; Abayomi-Alli, A. A probabilistic clustering model for hate speech classification in twitter. *Expert Syst. Appl.* **2021**, *173*, 1–21. [CrossRef]
38. Keuper, M.; Tang, S.; Andres, B.; Brox, T.; Schiele, B. Motion segmentation amp; multiple object tracking by correlation co-clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 140–153. [CrossRef]
39. Li, L.-Q.; Wang, X.-L.; Liu, Z.-X.; Xie, W.-X. A novel intuitionistic fuzzy clustering algorithm based on feature selection for multiple object tracking. *Int. J. Fuzzy Syst.* **2019**, *21*, 1613–1628. [CrossRef]
40. He, S.; Shin, H.-S.; Tsourdos, A. Multi-sensor multi-target tracking using domain knowledge and clustering. *IEEE Sens. J.* **2018**, *18*, 8074–8084. [CrossRef]
41. Gou, J.; Qiu, W.; Yi, Z.; Shen, X.; Zhan, Y.; Ou, W. Locality constrained representation-based k-nearest neighbor classification. *Knowl.-Based Syst.* **2019**, *167*, 38–52. [CrossRef]
42. Gou, J.; Ma, H.; Ou, W.; Zeng, S.; Rao, Y.; Yang, H. A generalized mean distance-based k-nearest neighbor classifier. *Expert Syst. Appl.* **2019**, *115*, 356–372. [CrossRef]
43. Yu, Y.-F.; Dai, D.-Q.; Ren, C.-X.; Huang, K.-K. Discriminative multi-layer illumination-robust feature extraction for face recognition. *Pattern Recognit.* **2017**, *67*, 201–212. [CrossRef]
44. Du, F.; Liu, P.; Zhao, W.; Tang, X. Joint channel reliability and correlation filters learning for visual tracking. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 1625–1638. [CrossRef]
45. Li, A.; Yang, M.; Yang, W. Feature integration with adaptive importance maps for visual tracking. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 779–785. [CrossRef]
46. Lukezic, A.; Vojir, T.; Ehovinzajc, L.; Matas, J.; Kristan, M. Discriminative correlation filter with channel and spatial reliability. *Int. Comput. Vis.* **2018**, *126*, 671–688. [CrossRef]
47. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 583–596. [CrossRef] [PubMed]
48. Danelljan, M.; Hager, G.; Khan, F.S.; Felsberg, M. Discriminative scale space tracking. *IEEE Trans. Pattern Anal. Mach.* **2016**, *39*, 1561–1575. [CrossRef] [PubMed]
49. Danelljan, M.; Bhat, G.; Khan, F.S.; 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.
50. 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.
51. Zheng, G.; Fu, C.; Ye, J.; Lin, F.; Ding, F. Mutation Sensitive Correlation Filter for Real-Time UAV Tracking with Adaptive Hybrid Label. In Proceedings of the IEEE International Conference on Robotics and Automation, Xi'an, China, 30 May–5 June 2021; pp. 503–509.
52. 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 (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 11920–11929.
53. Wu, Y.; Lim, J.; Yang, M.-H. Online object tracking: A benchmark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013; pp. 2411–2418.
54. Wu, Y.; Lim, J.; Yang, M.-H. Object tracking benchmark. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1834–1848. [CrossRef] [PubMed]
55. Li, S.; Yeung, D. Visual object tracking for unmanned aerial vehicles: A benchmark and new motion models. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4140–4146.
56. 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.
57. Du, D.; Qi, Y.; Yu, H.; Yang, Y.; Duan, K.; Li, G.; Zhang, W.; Huang, Q.; Tian, Q. The unmanned aerial vehicle benchmark: object detection and tracking. In Proceedings of the ECCV, Munich, Germany, 8–14 September 2018; pp. 370–386.
58. Ou, W.; You, X.; Tao, D.; Zhang, P.; Tang, Y.; Zhu, Z. Robust face recognition via occlusion dictionary learning. *Pattern Recognit.* **2014**, *47*, 1559–1572. [CrossRef]
59. Ou, W.; Luan, X.; Gou, J.; Zhou, Q.; Xiao, W.; Xiong, X.; Zeng, W. Robust discriminative nonnegative dictionary learning for occluded face recognition. *Pattern Recognit. Lett.* **2018**, *107*, 41–49. [CrossRef]
60. Sharma, S.; Kumar, V. Voxel-based 3d occlusion-invariant face recognition using game theory and simulated annealing. *Multimed. Tools Appl.* **2020**, *79*, 26517–26547. [CrossRef]
61. Zhu, H.; Qiao, Y.; Xu, G.; Deng, L.; Yu, Y.-F. Dspnet: A lightweight dilated convolution neural networks for spectral deconvolution with selfpaced learning. *IEEE Trans. Ind. Inform.* **2020**, *16*, 7392–7401. [CrossRef]