IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 8, 2021, accepted May 28, 2021, date of publication June 7, 2021, date of current version June 15, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3086821

Correlation Tracking via Spatial-Temporal Constraints and Structured Sparse Regularization

DAN TIAN, SHOUYU ZANG, AND BINBIN TU
School of Information Engineering, Shenyang University, Shenyang 110044, China

Corresponding author: Dan Tian (www.sltd2008@163.com)

This work was supported in part by the National Science Foundation of China under Grant 61703285, in part by the Liaoning Science Foundation of China under Grant 2019-MS-237, and in part by the Liaoning Ph.D. Research Startup Foundation of China under Grant 2020-BS-263.

ABSTRACT Discriminative correlation filter (DCF) has achieved promising performance in visual tracking for its high efficiency and high accuracy. However, DCF trackers usually suffer from some challenges, such as boundary effects and appearance changes. In this paper, we propose a novel correlation tracking method via spatial-temporal constraints and structured sparse regularization. Firstly, we introduce the background-aware selection strategy to extract real negative examples, and penalize the filter coefficients close to the boundary locations for spatial protection, both of which can alleviate the boundary effects. Secondly, we restrict the filters with structured sparse regularization to handle the local appearance changes, and exploit temporal consistent constraint on the filters to address the global appearance changes. Finally, we employ the alternative direction method of multipliers to optimize our correlation tracking model. In our optimization framework, we combine grayscale, color names, histogram of orientation gradient with deep features for appearance learning to improve the discrimination. Meanwhile, we penalize spatial constraint and structured sparse regularization alternatively based on occlusion detection to enhance processing efficiency. The qualitative and quantitative experiments are conducted on the OTB dataset. Experimental results demonstrate that the proposed tracker has better performance than other state-of-the-art trackers.

INDEX TERMS Object tracking, correlation filter, spatial-temporal constraints, structured sparse regularization, deep feature.

I. INTRODUCTION

Visual tracking is essentially a motion estimation problem. As a hot topic in the computer vision field, it plays an important role in many realistic applications, such as video surveillance, autonomous driving, and human-computer interaction. However, visual tracking is still a challenging task in some complicated scenarios, as the target undergoes deformation, occlusion, rotation, illumination variation, and background clutter.

Discriminative correlation filter (DCF) based trackers have received significant attention for the competitive performance in recent benchmarks. The core idea of DCF tracking is to construct cyclically shifted samples, so that fast Fourier transform in the frequency domain can be used to accelerate the tracking speed. Many advanced tracking methods based on DCF have been developed in recent years. Zheng *et al.* [1] develop a multi-task deep dual correlation filters based method for visual tracking, which takes full advantage of the multi-level features of deep networks. Li *et al.* [2] present an adaptive multiple contexts correlation tracking framework, which utilizes a sigmoid spatial weight map to control the influence of local contexts. Yuan *et al.* [3] introduce a metric learning model in the correlation tracking framework to solve the target scale problem. Wang *et al.* [4] jointly compress and transfer CNN models within a knowledge distillation framework for real-time correlation tracking. Yuan *et al.* [5] present a self-supervised learning based tracker in a deep correlation framework, which can improve the representational ability. Moorthy *et al.* [6] estimate the target location based on the distribution of correlation response. Fan *et al.* [7] learn the correlation filter on a larger search region for robust tracking without the distraction of the interference region. Yuan *et al.* [8] propose a particle filter redetection based correlation tracker, which can redetect the unreliable object location. Yuan *et al.* [9] learn temporal regularized

The associate editor coordinating the review of this manuscript and approving it for publication was Szidónia Lefkovits.

correlation filters to adapt to the change of the tracking scenes. Huang *et al.* [10] propose a constrained multi-kernel correlation tracking filter, which uses spatial constraints to address the unwanted boundary effects.

DCF trackers usually suffer from some undesired issues, including boundary effects, drastic appearance changes of the target object, to name a few [11]–[13]. One issue in DCF based trackers is the boundary effects caused by cyclically shifted sampling. The artificial negative examples can affect the discriminative ability of the learned model, and reduce the robustness of correlation tracking. To avoid the risk of tracking drift, related works have been focused on by researchers. Bolme *et al.* [14] weight the image by a cosine window, so that the pixel values near the edge gradually reduce to zero. Ji and Wang [15] exploit spatial prior information to penalize correlation filter coefficients, which can obtain a larger set of negative samples to learn a more discriminative model. Feng *et al.* [16] consider tracking reliability and object saliency in the energy function of DCF, which reflects the spatial-temporal information effectively. Galoogahi *et al.* [17] extract real negative examples from the background to learn filters, which demonstrates superior accuracy. Li *et al.* [18] incorporate temporal and spatial regularization into DCF tracking, and instead of the single sample with multiple training samples. Danelljan *et al.* [19] penalize correlation filters coefficients with a spatial regularization component to address the effects of the periodic assumption.

Drastic appearance changes of the target object are the main challenge for robust tracking. The abrupt appearance changes due to deformation, partial and full occlusion, motion blur, in-plane and out-of-plane rotation, scale variation, and illumination variation can affect the stability of object tracking. Great progress has been made in recent years in solving these problems. Xu *et al.* [20] present a temporal consistency preserving model to keep the global structure in the manifold space and preserve appearance diversity. Sun *et al.* [21] introduce the ROI pooled features into the correlation tracking, which can offer robust target representation. Sui *et al.* [22] leverage anisotropic filter response to replace Gaussian-shaped response in the tracking model, which can adapt the abrupt appearance changes. Huang *et al.* [23] restrict the rate of alteration in response maps to suppress the aberrances happening. Dai *et al.* [24] propose an adaptive spatial regularization scheme to respond to the appearance variations reliably.

The existing correlation trackers have achieved a lot in dealing with the challenges of boundary effects and appearance changes, but few trackers consider both global and local appearance changes of the target simultaneously. To solve this problem, we present a spatial-temporal constraints and structured sparse regularization based formulation for correlation tracking. Comparative evaluations are performed on the OTB benchmark. The experimental results validate the robustness and effectiveness of our tracker against some state-of-the-art DCF trackers.

The main contributions of this work include:

- To alleviate the boundary effects, we introduce the background-aware selection strategy to extract real negative samples, so that the tracking process can apply real foreground and background information. Furthermore, we penalize the filter coefficients close to the boundary locations for spatial protection.
- To learn the appearance changes of the target object, we restrict the filters with structured sparse regularization to cope with the local appearance changes (e.g., due to deformation, occlusion), and exploit temporal consistent constraint on the filters to tackle the global appearance changes (e.g., due to illumination variation, motion blur).
- Our correlation tracker can be optimized in the Fourier domain via the alternative direction method of multipliers (ADMM). To strengthen the discrimination ability, we combine grayscale, color names (CN), histogram of orientation gradient (HOG) with deep CNN features for tracking training. To enhance the processing efficiency, we penalize spatial constraint and structured sparse regularization alternatively based on occlusion detection.

The remainder of this paper is organized as follows: Section II reviews the related correlation trackers, and gives a detailed description of our proposed tracking model. In Section III, the implementation details are presented. The qualitative and quantitative results are depicted and analyzed in Section IV. Finally, the conclusion is addressed in section V.

II. SPATIAL-TEMPORAL CONSTRAINTS AND STRUCTURED SPARSE REGULARIZATION FOR DCF TRACKING

A. ORIGINAL MULTI-CHANNEL DCF TRACKING MODEL

The original multi-channel DCF tracker [25] can be obtained by optimizing the following formula,

$$\arg\min_{w} \frac{1}{2}\left\| y - \sum_{d=1}^{D} x^{d} * w^{d} \right\|_2^2 + \frac{\lambda}{2}\sum_{d=1}^{D}\left\| w^{d} \right\|_2^2, \qquad (1)$$

where $y \in \mathbb{R}^N$ is the desired Gaussian-shaped response, $x^d \in \mathbb{R}^N$ and $w^d \in \mathbb{R}^N$ denote the feature map and the filter in the d-th channel, D is the dimension of feature channels, λ is a regularization parameter, and $*$ denotes the spatial correlation operator.

The original multi-channel DCF tracker suffers from unexpected spatial boundary effects caused by cyclically shifted sampling. To solve this problem, the representative approaches are BACF [17] and STRCF [18].

B. REVISIT BACF

BACF learns the multi-channel DCF by minimizing the optimization problem,

$$\arg\min_{w} \frac{1}{2}\left\| y - \sum_{d=1}^{D} Bx^{d} * w^{d} \right\|_2^2 + \frac{\lambda}{2}\sum_{d=1}^{D}\left\| w^{d} \right\|_2^2, \qquad (2)$$

where $B \in \mathbb{R}^{M \times N}$ is a binary cropping operator used to select the mid M elements of the feature sample. Usually, $M \ll N$.

BACF applies the real foreground and background samples rather than cyclically shifted samples for DCF learning so as to resolve the boundary effects.

C. REVISIT STRCF

STRCF learns the optimal filters by minimizing the cost function,

$$\arg\min_{w} \frac{1}{2} \left\| \sum_{d=1}^{D} x_t^d * w^d - y \right\|_2^2 + \frac{1}{2} \sum_{d=1}^{D} \left\| f \bullet w^d \right\|_2^2 + \frac{\lambda}{2} \left\| w - w_{t-1} \right\|^2, \quad (3)$$

where f is the spatial regularization matrix, w_{t-1} denotes the correlation filters in the $(t-1)$-th frame. In Eq.3, the last two terms penalize the spatial regularization and the temporal regularization respectively. In addition, to inhibit the boundary effects, STRCF can also alleviate the impact of occlusion by keeping the DCFs close to the former ones.

D. OUR CORRELATION TRACKING MODEL

Motivated by BACF and STRCF, we present a spatial-temporal constraints and structured sparse regularization based correlation tracking model as follows:

$$\arg\min_{w} \frac{1}{2} \left\| \sum_{d=1}^{D} Bx_t^d * w^d - y \right\|_2^2 + \frac{\eta}{2} \sum_{d=1}^{D} \left\| f \cdot w^d \right\|_2^2 + \frac{\lambda}{2} \left\| w - w_{t-1} \right\|^2 + (1-\eta) \sum_{d=1}^{D} \left\| w^d \right\|_{2,1}, \quad (4)$$

where $\|\cdot\|_{2,1}$ denotes the $L_{2,1}$ norm, which carries out the L_2-norm on the filters firstly, then calculates the L_1-norm to realize structured joint sparse. η is a binary parameter,

$$\eta = \begin{cases} 0, & \mathit{if}\ PSR \le \sigma \\ 1, & \mathit{otherwise}. \end{cases} \quad (5)$$

The value η is set based on the Peak to Sidelobe Ratio (PSR) of the response map, σ is an experiential threshold.

In model (4), the first three terms are based on BACF and STRCF, which depict convolution error, spatial constraint, and temporal consistency, respectively. The last term is our novel idea which represents structured sparse regularization.

In this optimization, the first term describes the squared error between the actual output and the desired Gaussian-shaped response. We utilize the BACF method to extract real negative examples densely from the background, and then the tracking process can apply the true foreground and background information. This can help to acquire more reliable filter coefficients and alleviate the boundary effects.

The second term employs the STRCF method to restrain the filters with spatial regularization weights, which makes the filters to learn more weights for the reliable features of the central regions, and penalize the unreliable features close to the boundary regions. This constraint strategy can relieve the boundary effects issue effectively via spatial regularization.

The third term introduces the temporal consistency constraint on the filters. Based on the STRCF method, we consider the historical information of the target appearance, and force the filters to change smoothly across successive frames. This penalty item profits to reduce redundant uncorrelated information, thereby can adapt the global appearance changes (e.g., due to illumination variation, motion blur) of the target object.

The last term restricts the filters with structured sparse regularization. The standard L_1 sparse regularization of the filters can manage the major local appearance changes of the target object. However, variable selection based on L_1-norm can only be carried out for a single variable, because the correlation between continuous variables cannot be considered. Therefore, we utilize $L_{2,1}$-norm to impose a structured sparse regularization on the filters, which can exploit the relationship between the filters jointly through all the feature channels. This joint regularization can avoid small and dense errors, which is more effective to cope with the local appearance changes (e.g., due to deformation, occlusion) of the target object.

To enhance the tracking efficiency, we penalize spatial constraint (the second term) and structured sparse regularization (the last term) alternatively based on occlusion detection. Here, we utilize the PSR value of the response map to check for occlusion.

III. OPTIMIZATION

The model in Eq. 4 is a convex optimization problem. The optimal global solution can be found via the ADMM. We introduce a slack variable g, and then the model can be expressed equivalently in an augmented Lagrange formulation,

$$\begin{aligned} \mathcal{L}(f, g, s) = {} & \frac{1}{2} \left\| \sum_{d=1}^{D} Bx_t^d * w^d - y \right\|_2^2 + \frac{\eta}{2} \sum_{d=1}^{D} \left\| f \cdot g^d \right\|_2^2 \\ & + (1-\eta) \sum_{d=1}^{D} \left\| g^d \right\|_{2,1} + \sum_{d=1}^{D} \left(w^d - g^d \right)^T s^d \\ & + \frac{\gamma}{2} \sum_{d=1}^{D} \left\| w^d - g^d \right\|^2 + \frac{\lambda}{2} \left\| w - w_{t-1} \right\|^2, \\ s.t.\ & w = g \end{aligned} \quad (6)$$

where s, γ are the Lagrange vector and penalty factor. To integrate the equality constraint into the formulation, we introduce $h = s/\gamma$. Then, the objective function can be reformulated as follows:

$$\begin{aligned} \mathcal{L}(f, g, h) = {} & \frac{1}{2} \left\| \sum_{d=1}^{D} Bx_t^d * w^d - y \right\|_2^2 + \frac{\eta}{2} \sum_{d=1}^{D} \left\| f \cdot g^d \right\|_2^2 \\ & + (1-\eta) \sum_{d=1}^{D} \left\| g^d \right\|_{2,1} \end{aligned}$$

$$+\frac{\gamma}{2}\sum_{d=1}^{D}\left\|w^d - g^d + h^d\right\|^2 + \frac{\lambda}{2}\left\|w - w_{t-1}\right\|^2. \tag{7}$$

This formulation can be decomposed into some subproblems. Then, the ADMM can solve these subproblems alternately as follows,

$$\begin{cases} w^{(i+1)} = \arg\min_{w} \left\|\sum_{d=1}^{D} Bx_t^d * w^d - y\right\|_2^2 + \gamma\left\|w - g + h\right\|^2 \\ \qquad + \lambda\left\|w - w_{t-1}\right\|^2 \\ if\,\eta = 1 \\ g^{(i+1)} = \arg\min_{g} \sum_{d=1}^{D}\left\|f \cdot g^d\right\|_2^2 + \gamma\left\|w - g + h\right\|^2 \\ elseif\;\; \eta = 0 \\ g^{(i+1)} = \arg\min_{g} \sum_{d=1}^{D}\left\|g^d\right\|_{2,1} + \frac{\gamma}{2}\left\|w - g + h\right\|^2 \\ end \\ h^{(i+1)} = h^{(i)} + w^{(i+1)} - g^{(i+1)}. \end{cases} \tag{8}$$

Next, we introduce the solution of each subproblem in detail.

Subproblem w: To enhance computing efficiency, we learn the correlation filters in the frequency domain. Therefore, the subproblem w need to be rewritten based on Parseval's theorem as follows,

$$\hat{w}^d = \arg\min_{\hat{w}^d} \left\|B\hat{x}_t^d \odot \hat{w}^d - \hat{y}\right\|_2^2 + \gamma\left\|\hat{w} - \hat{g} + \hat{h}\right\|^2 + \lambda\left\|\hat{w} - \hat{w}_{t-1}\right\|^2. \tag{9}$$

The superscript ^ represents the discrete Fourier transformation (DFT), the operator $\odot$ denotes element-wise multiplication. This minimization problem requires to be solved for each channel respectively. The closed-form solution can be obtained by setting the derivative of Eq. 9 to zero, then

$$\hat{w}^d = \frac{\hat{x}_t^d \odot \hat{y}^* + \lambda\hat{w}_{t-1} + \gamma\hat{g} - \hat{h}}{\hat{x}_t^d \odot \hat{x}_t^{d*} + \lambda + \gamma}, \tag{10}$$

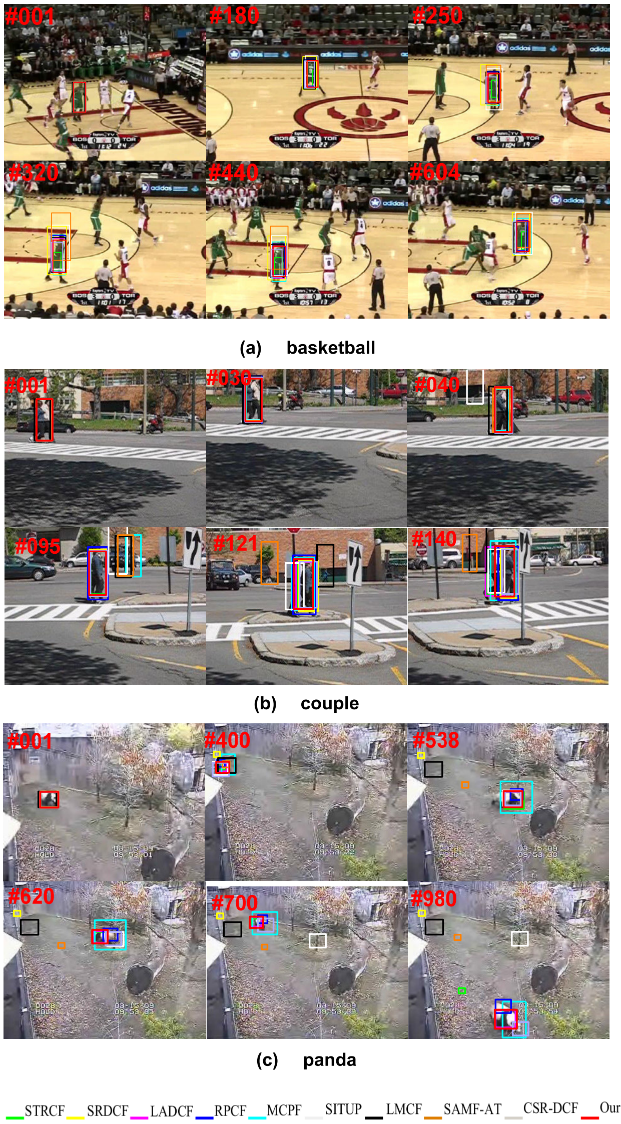
where superscript $*$ denotes conjugate transpose operation. Finally, w can be obtained by the inverse DFT.

Subproblem g: To update the slack variable g, we firstly perform occlusion detection based on the *PSR* value of the response map.

If there's no occlusion, we set the value $\eta = 1$ to enhance the spatial constraint. That is to solve the second sub-equation in Eq.(8). The solution g can be calculated as,

$$g = (F^T F + \gamma I)^{-1}(\gamma w + \gamma h), \tag{11}$$

where F is constituted by jointing D diagonal matrices $Diag(f)$.



(a) basketball

(b) couple

(c) panda

FIGURE 1. Qualitative results on three sequences with deformation.

If there is occlusion, we set the value $\eta = 0$ to perform the structured sparse regularization as the third sub-equation in Eq. (8), which can be separated for each spatial feature,

$$g_j = \arg\min_{g_j} \left\|g_j\right\|_{2,1} + \frac{\gamma}{2}\left\|w_j - g_j + h_j\right\|^2 \tag{12}$$

Then, the solution g can be computed by

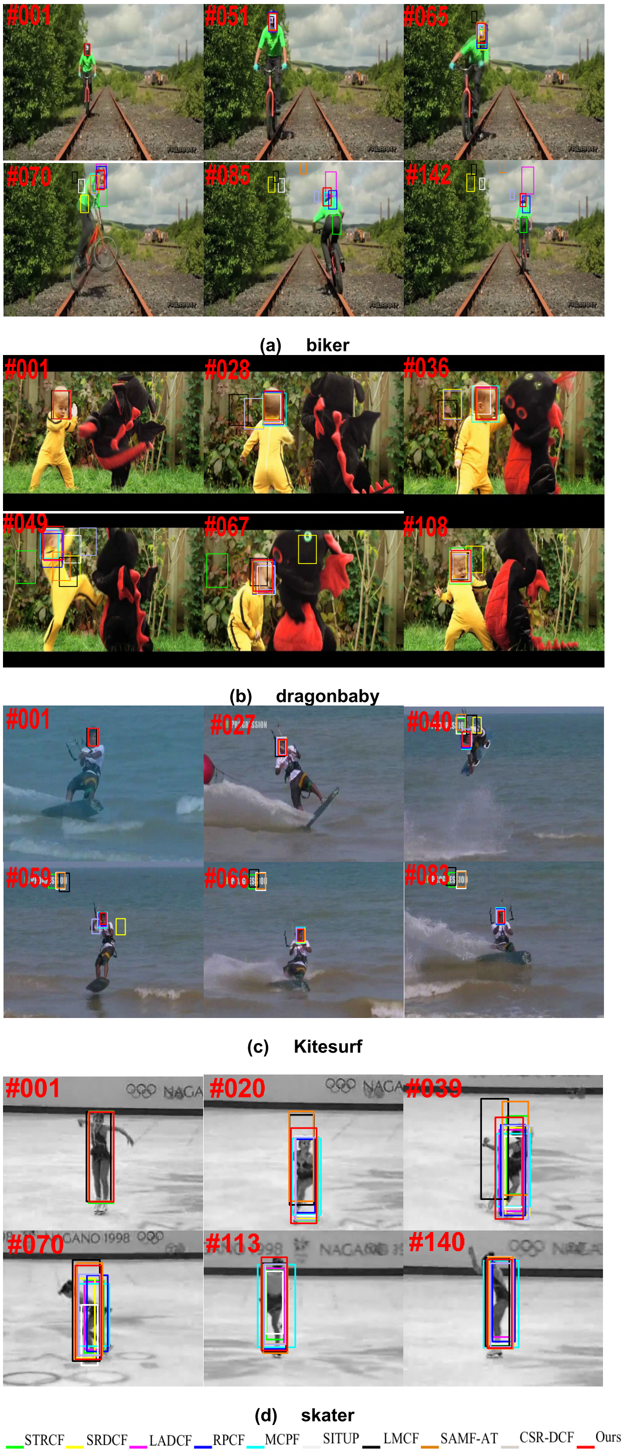
$$g = \max\left(0, 1 - \frac{1}{\gamma\left\|f + h\right\|_2}\right)(f + h). \tag{13}$$

Updating γ: The penalty parameter γ is updated by

$$\gamma^{(i+1)} = \min(\gamma^{\max}, \rho\gamma^{(i)}), \tag{14}$$

where $\gamma^{\max}$ is the maximum value of γ, ρ denotes the scale factor. $\rho > 1$ can accelerate the convergence.

Complexity Analysis: The subproblem w requires performing DFT and inverse DFT operation in $O(DMN\log(MN))$ complexity. For the subproblem g, the computational cost for



(a) biker

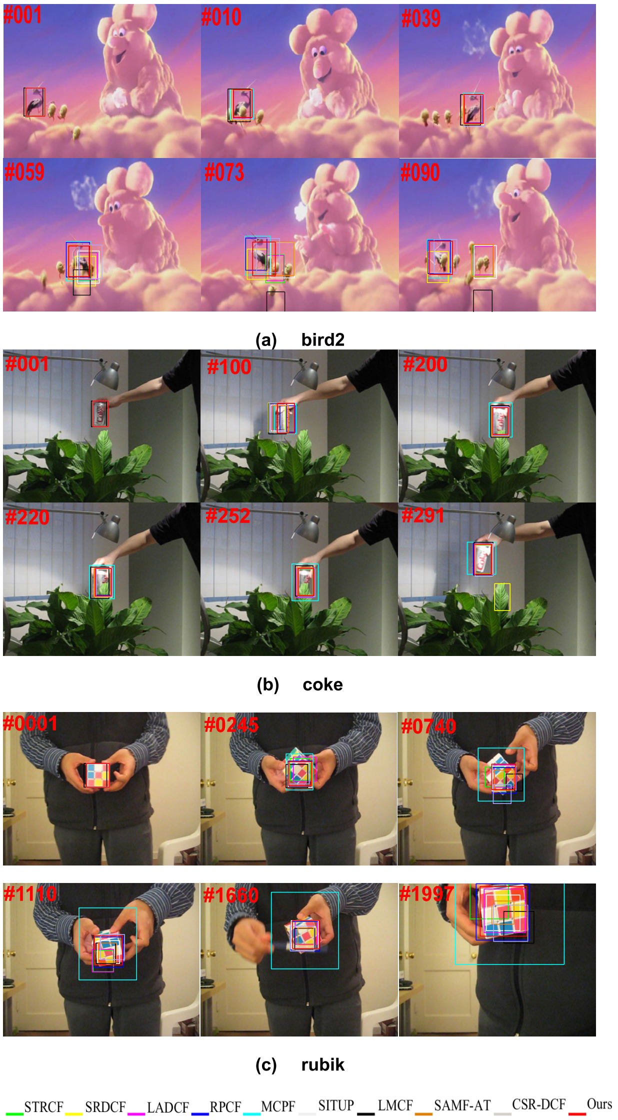
(b) dragonbaby

(c) Kitesurf

(d) skater

FIGURE 2. Qualitative results on four sequences with rotation.

the spatial constraint is $O(DMN)$ when there's no occlusion, the computational cost for the structured sparse regularization by element-wise operation is $O(1)$ when there's occlusion. The overall complexity of our optimization framework is $O(KDMN\log(MN))$ where K represents the maximum number of iterations.



(a) bird2

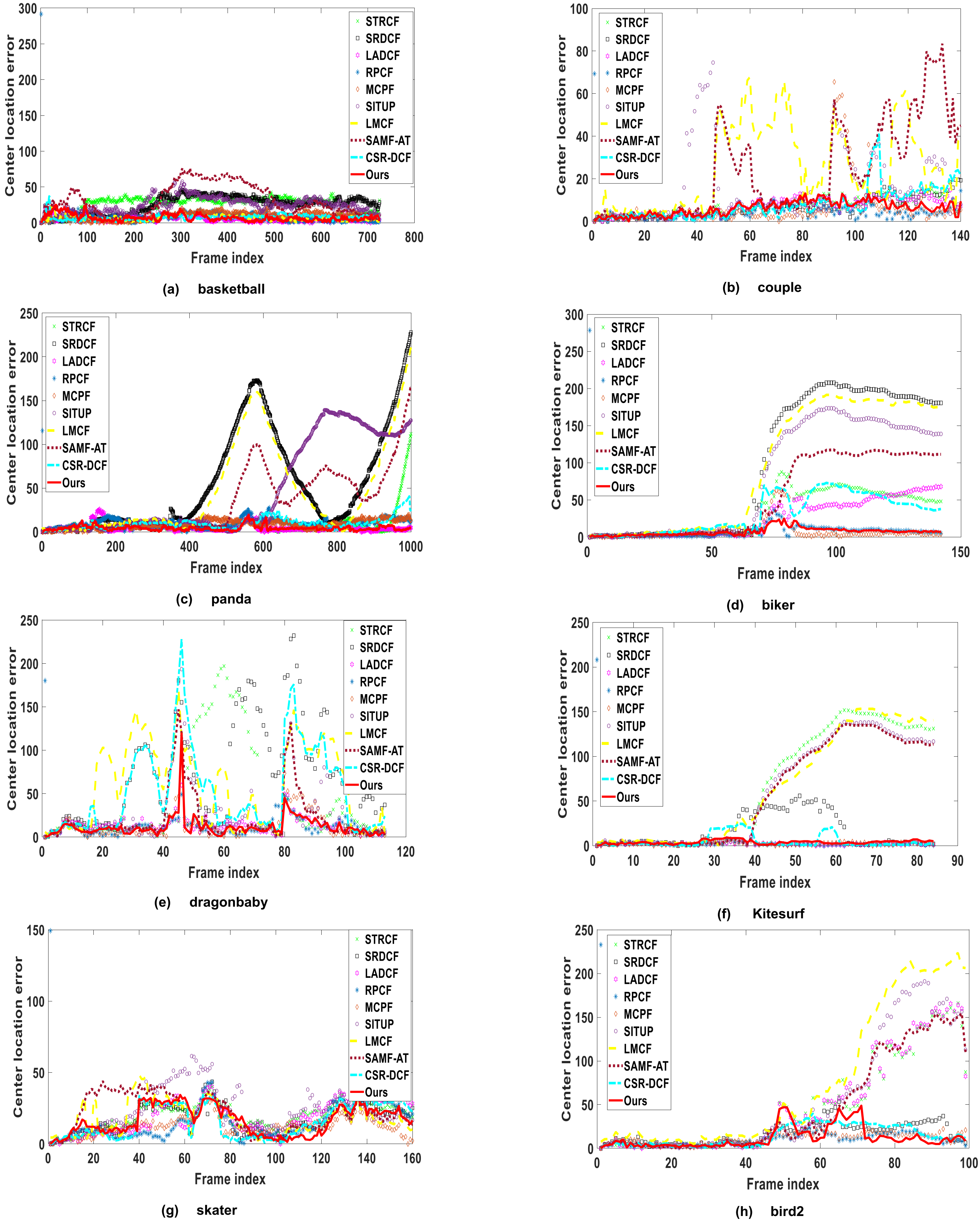
(b) coke

(c) rubik

FIGURE 3. Qualitative results on three sequences with occlusion.

IV. EXPERIMENTS

On the OTB benchmark, our tracker is compared with 9 state-of-the-art trackers including STRCF [18], SRDCF [19], LADCF [20], RPCF [21], MCPF [26], SITUP [27], LMCF [28], SAMF-AT [29], and CSR-DCF [30]. The source codes and the results of these trackers are provided publicly. The experiments are conducted on some challenging videos (basketball, couple, panda, biker, dragonbaby, kitesurf, skater, bird2, coke, and rubik). These videos are classified according to their main challenging factors including deformation, rotation, and occlusion, which can lead to global or local appearance changes of the tracked object respectively. We test our Matlab implementation on a PC machine with an Intel Core i5 running at 2.40 GHz. The tracking speed of our tracker is listed in Table 1. The qualitative and quantitative results show the superior effectiveness and robustness of our tracker.



(a) basketball

(b) couple

(c) panda

(d) biker

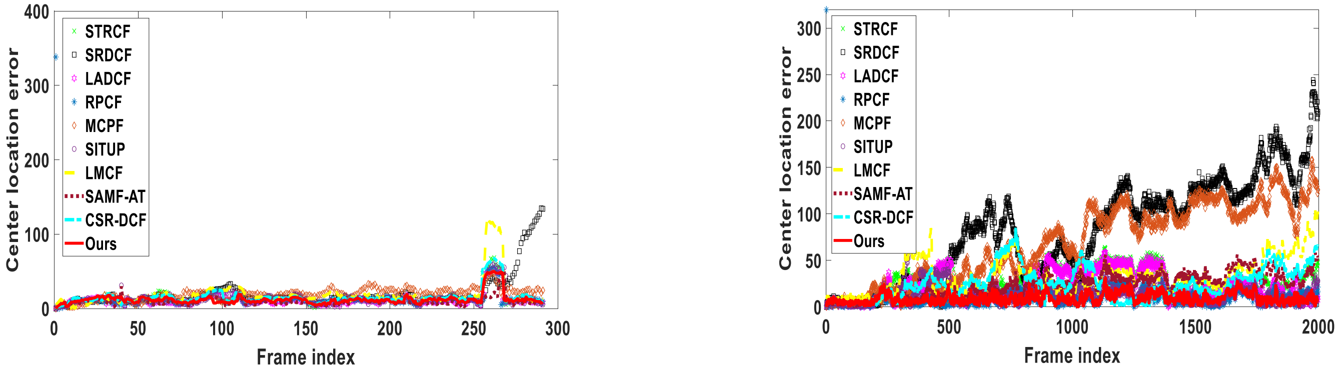
(e) dragonbaby

(f) Kitesurf

(g) skater

(h) bird2

FIGURE 4. Performance in precision over sequences with deformation, rotation and occlusion.



(i) coke

(j) rubik

FIGURE 4. ***(Continued.)*** **Performance in precision over sequences with deformation, rotation and occlusion.**

TABLE 1. **Tracking speed of our tracker.**

	basketball	couple	panda	biker	dragonbaby	kitesurf	skater	bird2	coke	rubik
fps	1.7462	2.1520	3.4218	3.3782	1.8237	2.9934	2.1453	1.8852	1.8941	1.8064

A. IMPLEMENTATION DETAILS

1) FEATURE EXTRACTION

To enhance the discrimination of our correlation tracker, we combine hand-crafted features (grayscale features, 31-channel HOG features, 10-channel CN features) with deep CNN features for object representation. Then, the multi-channel features are weighted by a cosine window to eliminate the discontinuity from the cyclic shifts.

2) PARAMETER SELECTION

We set the regularization parameter as $\lambda = 15$, experiential threshold as $\sigma = 20$. As for the ADMM, the maximum number of iterations is set to 2, the number of scales and scale step is set to 5 and 1.0. The initial penalty factor $\gamma^{(0)}$, maximum penalty factor $\gamma^{\max}$, and scale factor ρ are set to 1, 0.1 and 10 respectively.

B. QUALITATIVE RESULTS

Figures 1-3 present the qualitative tracking results of the compared trackers on 10 benchmark videos. We sort the results under the main challenging factors in the videos.

1) DEFORMATION

In the basketball sequence, the man undergoes deformation, background clutters, out-of-plane rotation, and illumination variation. Figure 1(a) depicts some representative results. The target scales obtained by STRCF, SRDCF, SITUP and SAMF-AT trackers have different degrees of deviation (e.g., #320 and #604). Furthermore, the target bounding boxes estimated by STRCF and SAMF-AT trackers deviate from the target to some extent. The couple sequence contains scenes with deformation, background clutters, and fast motion. Figure 1(b) shows that the results of SITUP, LMCF, MCPF and SAMF-AT trackers drift away from the couple in some moments, such as around #95. The panda sequence suffers from deformation, scale variation, and low resolution. Figure 1(c) reports that the results of STRCF, SRDCF, SITUP, LMCF and SAMF-AT trackers drift to the other areas around #700 and #980. All targets in these three sequences experience global appearance changes. For these challenges, our tracker exploits temporal consistent constraint on the filters to address the problem. We can complete the tracking successfully, and obviously outperform most of the other trackers.

2) ROTATION

In the biker sequence, the target face undergoes out-of-plane rotation, motion blur, and out-of-view simultaneously. Figure 2(a) presents some representative results. Most trackers cannot locate the object effectively when there is both out-of-plane rotation and motion blur (e.g., #70), and also cannot accomplish the tracking task successfully when the target face is out of view (e.g., #85 and #142) except for RPCF, MCPF, and our trackers. The dragonbaby sequence involves in-plane rotation, out-of-plane rotation, out-of-view, and motion blur. In Figure 2(b), we can see that the target baby has strong appearance variations when he moves. SRDCF, LMCF and CSR-DCF trackers cannot overcome the challenge for out-of-view around #28. In addition, STRCF and SITUP trackers cannot handle out-of-plane rotation and motion blur simultaneously (e.g., #49). In the kitesurf sequence, the sportsman suffers from out-of-plane rotation, in-plane rotation, and illumination variation. As shown in Figure 2(c), STRCF, SITUP, LMCF, and SAMF-AT trackers lose the object under drastic out-of-plane rotation (e.g., #59, #66, and #83). In the skater sequence, the skater does frequent in-plane rotation and out-of-plane rotation as shown

TABLE 2. CLE of the trackers.

	STRCF	SRDCF	LADCF	RPCF	MCPF	SITUP	LMCF	SAMF-AT	CSR-DCF	Ours
basketball	26.0	25.1	**5.7**	7.6	9.3	24.7	**6.9**	31.2	7.5	7.1
couple	7.5	6.6	7.0	**4.4**	7.2	15.4	24.5	23.0	7.7	**6.2**
panda	8.1	53.9	**4.9**	9.7	8.6	46.1	45.5	34.4	9.8	**4.3**
biker	32.7	97.9	26.1	8.9	**7.7**	79.7	94.4	54.2	31.5	**7.3**
dragonbaby	43.1	71.5	13.2	13.0	**12.9**	30.5	58.8	21.3	52.5	**11.9**
kitesurf	66.4	16.0	**2.1**	4.5	**2.2**	60.2	65.2	58.7	5.0	4.0
skater	19.3	17.0	18.7	**16.1**	**12.5**	24.2	17.6	21.9	**16.1**	16.6
bird2	47.7	17.0	48.80	12.1	**11.1**	54.7	75.4	46.4	14.5	**12.0**
coke	14.2	20.0	**11.4**	13.2	19.0	13.3	18.3	**9.3**	15.1	12.4
rubik	26.0	81.0	22.5	**9.8**	68.6	13.4	33.2	23.8	24.9	**9.3**
average	29.1	40.6	16.0	**9.9**	15.9	36.2	44.0	32.4	18.5	**9.1**

TABLE 3. OS rate of the trackers (%).

	STRCF	SRDCF	LADCF	RPCF	MCPF	SITUP	LMCF	SAMF-AT	CSR-DCF	Ours
basketball	37.0	52.7	**75.3**	**73.7**	67.9	55.8	73.2	49.4	65.4	72.8
couple	67.9	69.5	71.9	**76.4**	69.4	49.3	32.9	41.2	65.0	**73.3**
panda	39.4	12.0	45.5	26.9	**48.8**	32.4	25.1	23.2	43.5	**56.9**
biker	37.7	35.3	45.2	53.8	**58.2**	29.5	22.1	34.5	23.5	**66.1**
dragonbaby	50.6	25.2	66.0	**69.6**	67.2	49.1	29.9	61.0	33.3	**71.6**
kitesurf	35.9	49.2	**76.5**	**73.1**	71.6	34.6	24.8	34.8	65.3	68.7
skater	52.1	56.7	52.1	56.9	**73.0**	46.8	57.1	57.9	57.2	**60.9**
bird2	46.0	60.5	46.2	**75.9**	**68.5**	47.8	37.0	49.1	60.7	67.4
coke	55.9	51.1	**63.7**	58.9	51.5	58.2	58.0	**66.0**	53.3	60.4
rubik	42.3	33.1	51.7	**76.9**	37.9	65.5	41.9	46.1	50.6	**74.3**
average	46.5	44.5	59.4	**64.2**	61.4	46.9	40.2	46.3	51.8	**67.2**

in Figure 2(d). LMCF and SAMF-AT trackers drift sometimes, but can retrace the object finally. All targets in these sequences go through abrupt appearance changes. For these challenges, RPCF and our trackers can achieve favorable results. RPCF introduces the ROI pooled features to offer robust target representation. Our tracker penalizes the appearance information via spatial constraint strategy to overcome the effect.

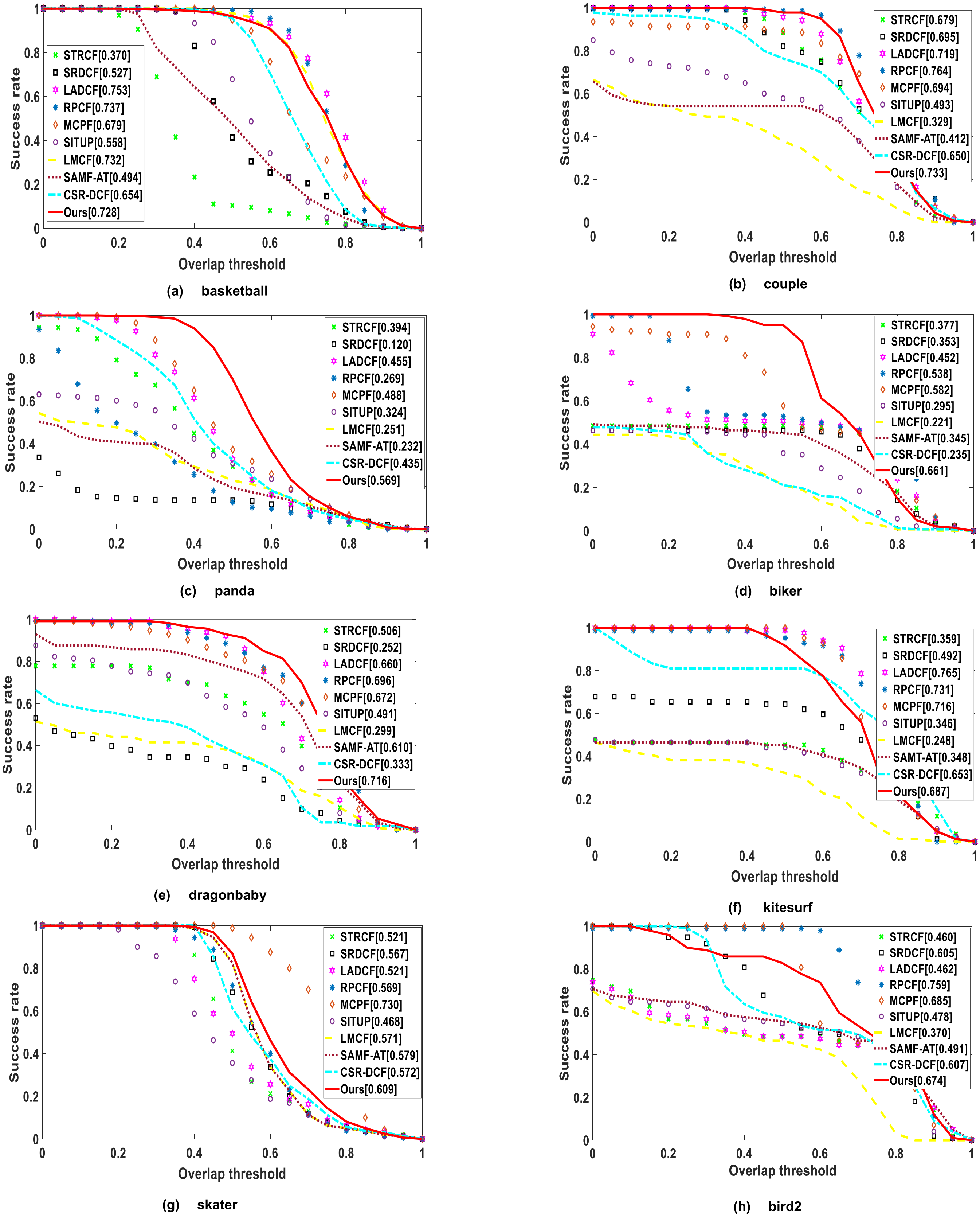
3) OCCLUSION

In the bird2 sequence, the fast-moving target not only suffers from frequent occlusion, but also has self-induced appearance changes. Figure 3(a) gives some representative results. STRCF, LADCF, SITUP, and SAMF-AT trackers lost the bird and track the shelter in turn (e.g. #73, #90). Furthermore, LMCF tracker cannot detect the target as it moves and turns. In the coke sequence, the coke bottle undergoes occlusion, illumination variation, and fast motion as shown in Figure 3(b). Most of the trackers can detect the target robustly, whereas the SRDCF tracker drifts off the object after occlusion (e.g. #291). Moreover, the target scales acquired by MCPF tracker have deviation (e.g., #252 and #291). In the rubik sequence, the rubik undergoes occlusion, scale variation and rotation as shown in Figure 3(c). When these challenging factors exist simultaneously, most of the trackers cannot acquire the target effectively except for RPCF, SITUP, and ours (e.g. #1997). In these sequences, our tracker can successfully track the target under occlusion mainly due to the structured sparse regularization.

C. QUANTITATIVE RESULTS

The tracking performance is evaluated quantitatively under two metrics: precision and success.

The precision is measured by the center location error (CLE), which is computed as the Euclidean distance between the center of the predicted tracked object and the ground truth bounding box. Figure 4 depicts the CLE curve for different trackers over sequences with deformation, rotation and occlusion respectively. By comparison, we can conclude that our tracker outperforms SRDCF, SITUP, and SAMF-AT



(a) basketball

(b) couple

(c) panda

(d) biker

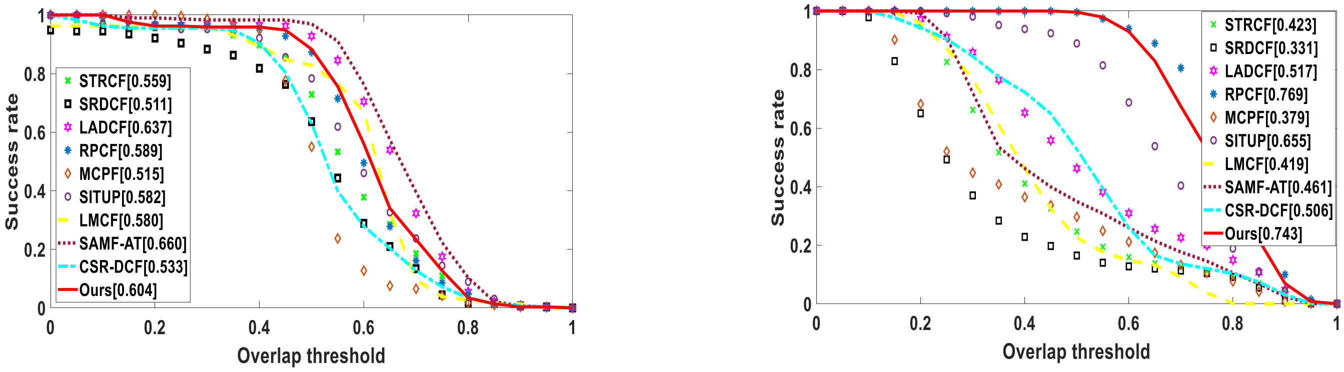
(e) dragonbaby

(f) kitesurf

(g) skater

(h) bird2

FIGURE 5. Performance in success over sequences with deformation, rotation and occlusion.



(i) coke

(j) rubik

FIGURE 5. ***(Continued.)*** **Performance in success over sequences with deformation, rotation and occlusion.**

trackers in dealing with deformation, outperforms STRCF, SRDCF, SITUP, LMCF and SAMF-AT trackers in dealing with rotation, and outperforms STRCF, SRDCF, LMCF, and SAMF-AT trackers in dealing with occlusion.

The success is measured by the overlap rate (OR), which is defined as

$$OR = \frac{area(ROI_T \cap ROI_G)}{area(ROI_T \cup ROI_G)}, \tag{15}$$

where ROI_T and ROI_G denote the areas of the tracked bounding box and the groundtruth box, $area(\cdot)$ denotes the area of the box in pixels, $\cap$ and $\cup$ denote intersection and union operations of boxes. The success rate is the percentage of frames where OR value exceeds the specified threshold. Usually, overlap threshold is set to 0.5. Figure 5 demonstrates the success plot for the compared trackers. The results tell us that our tracker has the superior performance over STRCF, SRDCF, SITUP, SAMF-AT, and CSR-DCF trackers when facing deformation challenge, surpasses STRCF, SRDCF, SITUP, LMCF, SAMF-AT, and CSR-DCF trackers when facing rotation challenge, and surpasses STRCF, SRDCF, SITUP, LMCF, and CSR-DCF trackers when facing occlusion challenge.

The detailed quantitative comparison results are presented in Table 2 and Table 3. The best and the second best results for the metric are marked in bold-face, and the last row lists the average tracking performance of the trackers. For CLE, the smaller the result, the more precise the tracker is. However, in terms of overlap success (OS) rate, the higher the value, the more successful the tracking. From the results, it can be seen that our tracker basically obtains the best or the second best performance in the light of both precision and success. For the average performance, the top 3 trackers are our tracker with 9.1 in precision and 67.2% in success, RPCF with 9.9 in precision and 64.2% in success, MCPF with 15.9 in precision and 61.4% in success. Therefore, it is reasonable to conclude that our tracker is more robust to above challenges than many other state-of-the-art trackers.

V. CONCLUSION

In this paper, we develop a novel correlation tracking algorithm based on spatial-temporal constraints and structured sparse regularization. Our tracker involves several key technical elements as follows. We introduce the background-aware selection strategy to extract real negative examples from background, and learn more reliable features in central regions via weighted spatial constraint, which all alleviate the boundary effects problem. Furthermore, we constrain the filters with structured sparse regularization to cope with the local appearance changes, and employ temporal consistent constraint on the filters to adapt the global appearance changes. Finally, we utilize the ADMM technique in the Fourier domain for online tracking optimization. To enhance the tracking efficiency, we penalize spatial constraint and structured sparse regularization alternatively via occlusion detection. Qualitative and quantitative results on benchmark sequences have verified the robustness of our tracker, especially for complex deformation, rotation, and occlusion challenges.

REFERENCES

[1] Y. Zheng, X. Liu, X. Cheng, K. Zhang, Y. Wu, and S. Chen, "Multi-task deep dual correlation filters for visual tracking," *IEEE Trans. Image Process.*, vol. 29, no. 10, pp. 9614–9626, Oct. 2020.

[2] F. Li, H. Zhang, and S. Liu, "Correlation filters with adaptive multiple contexts for visual tracking," *IEEE Access*, vol. 8, pp. 94547–94559, May 2020.

[3] D. Yuan, W. Kang, and Z. He, "Robust visual tracking with correlation filters and metric learning," *Knowl.-Based Syst.*, vol. 195, May 2020, Art. no. 105697.

[4] N. Wang, W. Zhou, Y. Song, C. Ma, and H. Li, "Real-time correlation tracking via joint model compression and transfer," *IEEE Trans. Image Process.*, vol. 29, no. 4, pp. 6123–6135, Apr. 2020.

[5] D. Yuan, X. Chang, P.-Y. Huang, Q. Liu, and Z. He, "Self-supervised deep correlation tracking," *IEEE Trans. Image Process.*, vol. 30, no. 12, pp. 976–985, Dec. 2020.

[6] S. Moorthy, J. Y. Choi, and Y. H. Joo, "Gaussian-response correlation filter for robust visual object tracking," *Neurocomputing*, vol. 411, pp. 78–90, Oct. 2020.

[7] N. Fan, J. Li, Z. He, C. Zhang, and X. Li, "Region-filtering correlation tracking," *Knowl.-Based Syst.*, vol. 172, pp. 95–103, May 2019.

[8] D. Yuan, X. Lu, D. Li, Y. Liang, and X. Zhang, "Particle filter re-detection for visual tracking via correlation filters," *Multimedia Tools Appl.*, vol. 78, no. 11, pp. 14277–14301, Jun. 2019.

[9] D. Yuan, X. Shu, and Z. He, "TRBACF: Learning temporal regularized correlation filters for high performance online visual object tracking," *J. Vis. Commun. Image Represent.*, vol. 72, Oct. 2020, Art. no. 102882.

[10] B. Huang, T. Xu, S. Jiang, Y. Chen, and Y. Bai, "Robust visual tracking via constrained multi-kernel correlation filters," *IEEE Trans. Multimedia*, vol. 22, no. 11, pp. 2820–2832, Nov. 2020.

[11] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2014, pp. 1–5.

[12] H. K. Galoogahi, T. Sim, and S. Lucey, "Correlation filters with limited boundaries," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4630–4638.

[13] M. Zhang, J. Xing, J. Gao, and W. Hu, "Robust visual tracking using joint scale-spatial correlation filters," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 1468–1472.

[14] D. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2544–2550.

[15] Z. Ji and W. Wang, "Correlation filter tracker based on sparse regularization," *J. Vis. Commun. Image Represent.*, vol. 55, pp. 354–362, Aug. 2018.

[16] W. Feng, R. Han, Q. Guo, J. Zhu, and S. Wang, "Dynamic saliency-aware regularization for correlation filter-based object tracking," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3232–3245, Jul. 2019.

[17] H. K. Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1135–1143.

[18] F. Li, C. Tian, W. Zuo, L. Zhang, and M.-H. Yang, "Learning spatial-temporal regularized correlation filters for visual tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4904–4913.

[19] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4310–4318.

[20] T. Xu, Z.-H. Feng, X.-J. Wu, and J. Kittler, "Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5596–5609, Nov. 2019.

[21] Y. Sun, C. Sun, D. Wang, Y. He, and H. Lu, "ROI pooled correlation filters for visual tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5783–5791.

[22] Y. Sui, Z. Zhang, G. Wang, Y. Tang, and L. Zhang, "Exploiting the anisotropy of correlation filter learning for visual tracking," *Int. J. Comput. Vis.*, vol. 127, no. 8, pp. 1084–1105, Aug. 2019.

[23] Z. Huang, C. Fu, Y. Li, F. Lin, and P. Lu, "Learning aberrance repressed correlation filters for real-time UAV tracking," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2891–2900.

[24] K. Dai, D. Wang, H. Lu, C. Sun, and J. Li, "Visual tracking via adaptive spatially-regularized correlation filters," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4670–4679.

[25] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.

[26] T. Zhang, C. Xu, and M.-H. Yang, "Multi-task correlation particle filter for robust object tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4335–4343.

[27] H. Ma, S. T. Acton, and Z. Lin, "SITUP: Scale invariant tracking using average peak-to-correlation energy," *IEEE Trans. Image Process.*, vol. 29, pp. 3546–3557, Jan. 2020.

[28] M. Wang, Y. Liu, and Z. Huang, "Large margin object tracking with circulant feature maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4021–4029.

[29] A. Bibi, M. Mueller, and B. Ghanem, "Target response adaptation for correlation filter tracking," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 419–433.

[30] A. Lukezic, T. Vojir, L. C. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6309–6318.

DAN TIAN received the M.S. degree in control theory and control engineering and the Ph.D. degree in system simulation and application from Northeastern University, China, in 2006 and 2015, respectively. Since 2006, she has been with Shenyang University, where she is currently a Professor with the Information and Engineering College. From 2015 to 2020, she held a postdoctoral position with Tianjin University. She is the first author of more than 20 articles and hosts Natural Science Foundation of China, in 2017. Her main research interests include computer vision and digital image processing. She is a member of China Simulation Society. She received the Second Prize of the Municipal Natural Science Academic Achievement, in 2020.

SHOUYU ZANG received the B.S. degree in logistics engineering from the Huaiyin Institute of Technology, China, in 2019. She is currently pursuing the degree in logistics engineering with Shenyang University. Her main research interest includes video image processing.

BINBIN TU received the M.S. degree in computer technology from Shenyang Aerospace University, China, in 2011, and the Ph.D. degree in measurement techniques and instruments from the Shenyang University of Technology, China, in 2019. Since 2003, she has been with Shenyang University, where she is currently an Associate Professor with the Information and Engineering College. She is the first author of more than five articles and hosts Natural Science Foundation of Liaoning Province, in 2020. Her main research interests include signal processing and pattern recognition. She received the third prize of the Municipal Natural Science Academic Achievement, in 2019.