

Research Article

Robust Object Tracking via Reverse Low-Rank Sparse Learning and Fractional-Order Variation Regularization

Dan Tian ^{1,2}, Guoshan Zhang,¹ and Shouyu Zang²

¹School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

²School of Information Engineering, Shenyang University, Shenyang 110044, China

Correspondence should be addressed to Dan Tian; www.slttd2008@163.com

Received 23 January 2020; Revised 5 July 2020; Accepted 4 August 2020; Published 25 August 2020

Academic Editor: Giacomo Falcucci

Copyright © 2020 Dan Tian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Object tracking based on low-rank sparse learning usually makes the drift phenomenon occur when the target faces severe occlusion and fast motion. In this paper, we propose a novel tracking algorithm via reverse low-rank sparse learning and fractional-order variation regularization. Firstly, we utilize convex low-rank constraint to force the appearance similarity of the candidate particles, so as to prune the irrelevant particles. Secondly, fractional-order variation is introduced to constrain the sparse coefficient difference in the bounded variation space, which allows the difference between consecutive frames to exist, so as to adapt object fast motion. Meanwhile, fractional-order regularization can restrain severe occlusion by considering more adjacent frames information. Thirdly, we employ an inverse sparse representation method to model the relationship between target candidates and target template, which can reduce the computation complexity for online tracking. Finally, an online updating scheme based on alternating iteration is proposed for tracking computation. Experiments on benchmark sequences show that our algorithm outperforms several state-of-the-art methods, especially exhibiting better adaptability for fast motion and severe occlusion.

1. Introduction

Visual object tracking is an important technique in computer vision with many applications, such as robotics, medical image analysis, human-computer interaction, and traffic control. The goal of tracking is to predict the motion state of the moving object in the video stream based on the initial state. Much progress has been made in this area, but many challenging tasks still remain caused by partial or full occlusion, fast motion, illumination and scale variation, deformation, background clutter, etc.

Low-rank constraint [1–4] on the candidate particles can reflect the subspace structure feature of the object appearance. This subspace representation is robust to handle the global appearance changes problem (e.g., illumination variations and pose changes). Furthermore, for the robustness of local appearance changes (e.g., deformation and partial occlusions), sparse representation [5–8] models the image observation by a linear combination of dictionary

templates, which can measure the importance of each target candidate. Therefore, low-rank constraint and sparse representation can be learned jointly for effective tracking [9, 10]. Zhong et al. [11] develop a sparse collaborative model for object tracking, which exploits a sparse discriminative classifier and sparse generative model to describe drastic appearance changes. Zhang et al. [12] learn the sparse representation and low-rank constraint in the particle filter framework and exploit temporal consistency simultaneously. Wang et al. [13] propose an inverse sparse representation based tracking algorithm with a locally weighted distance metric. Sui and Zhang [14] exploit low-rank constraint to describe the global feature of all the patches and capture the sparsity structure to reflect the local relationship between the neighboring patches. Sui et al. [15] formulate spatial-temporal locality under a discriminative dictionary learning structure for object tracking. Dash and Patra [16] propose an effective tracking framework by using a regularized robust sparse coding for representing the

multifeature templates of the candidate objects. These methods can successfully deal with the target appearance change problem caused by lighting variations and partial occlusions. Nevertheless, these formulations are not effective for handling fast motion challenges.

To solve this problem, we introduce reverse low-rank sparse learning with fractional-order variation regularization for visual object tracking. In comparison with the existing low-rank sparse trackers, we introduce fractional-order variation regularization to the representation model. Fractional-order variation has been widely used in static image analysis. We generalize it to overcome the challenging problems in dynamic video tracking because of the following two factors: (1) the variation method can model the tracking problem in the bounded variation space, which allows the difference among a few frames to exist, to adapt the object fast motion. (2) Fractional differential is a global operator, which can take more adjacent frames information into account and overcome the severe occlusion problem.

The main contributions of this work are four-fold: (1) the low-rank constraint is exploited to prune the irrelevant particles; (2) fractional-order variation regularization is introduced to learn the jump information generated by fast motion and complex occlusion; (3) an inverse sparse representation formulation is built to reduce the computation complexity for real-time tracking; and (4) an alternating iteration strategy is presented for online tracking optimization.

2. Problem Formulation

In this work, we formulate the target states within the particle filter framework [13]. Particle filter processing is built upon Bayesian inference rule, which can be used for predicting the posterior probability density function of the state variables in dynamic system. Object tracking as a classical dynamic variables estimation problem is suitable to be modeled in this framework. Based on this idea, the a posteriori probability of the target state can be inferred recursively as follows:

$$p(x_t | y_{1:t-1}) = \int p(x_t | x_{t-1}) p(x_{t-1} | y_{1:t-1}) dx_{t-1}, \quad (1)$$

$$p(x_t | y_{1:t}) \propto p(y_t | x_t) p(x_t | y_{1:t-1}),$$

where x_t is the motion state variable at time t , y_t is the observed image, $p(x_t | x_{t-1})$ denotes the state transition model, and $p(y_t | x_t)$ denotes the observation model. Thus, the target state can be found by maximizing a probability estimation model as

$$\hat{x}_t = \arg_{x_t^i} \max p(y_t | x_t^i) p(x_t^i | x_{t-1}), \quad (2)$$

where x_t^i is the i -th candidate at time t .

2.1. State Transition Model. State transition model describes the change of the target state between two successive frames. We measure the transition of the target state by an affine motion formulation described as $x_t = [l_x, l_y, \theta, s, \alpha, \phi]^T$, whose parameters correspond to x, y coordinate translation,

rotation angle, scale, aspect ratio, and skew, respectively. To sample a group of candidate particles, we model the transition of the target state by a Gaussian distribution:

$$p(x_t | x_{t-1}) = N(x_t; x_{t-1}, \Sigma), \quad (3)$$

where Σ denotes a diagonal Gaussian distribution matrix, whose elements are the variances of the affine motion parameters.

2.2. Reverse Low-Rank Sparse Representation Model with Fractional-Order Variation Regularization. In this section, we utilize both reverse low-rank sparse learning and fractional-order variation regularization to formulate the object tracking. Firstly, we employ the local appearance representation based on patches to replace the holistic one for dealing with partial occlusion. Here, local patches are sampled sequentially in a nonoverlapping manner from the candidate particles as shown in Figure 1.

Secondly, we use a generative model based on statistical processing to select the optimal target candidate. The existing low-rank sparse optimization-based tracking methods usually make the drift phenomenon occur when the target faces complex occlusion and fast motion. Here, we build a reverse low-rank sparse learning formulation with fractional-order variation regularization for object tracking as follows:

$$\min_{Z_t} \frac{1}{2} \|T_t - D_t Z_t\|_2^2 + \lambda \|Z_t\|_* + \lambda_1 \|Z_t\|_1 + \lambda_2 \|\nabla^\alpha Z_t\|_1, \quad (4)$$

where

$$Z = [Z_t; Z_{t-1}^*; \dots; Z_{t-K}^*], \quad (5)$$

$$\nabla^\alpha = \sum_{k=0}^{K-1} (-1)^k C_k^\alpha. \quad (6)$$

At time t , T_t denotes the target template reshaped by the intensity vector of the observed target, whose initial value is drawn manually in the first frame and the current value is updated dynamically during tracking as shown in Section 3.2. D_t is a dictionary used for the sparse representation of target template, whose columns are formed by candidates particles $\{y_t^i\}_{i=1}^N$. y_t^i is the local patch vector of the candidate region in the current frame sampled by the state transition model. Z_t denotes the sparse coding. λ , λ_1 , and λ_2 are the adjustment parameters. $\|\cdot\|_*$ denotes the matrix nuclear norm. ∇^α is the fractional-order gradient operator. $K \geq 3$ is an integer constant, $C_k^\alpha = \Gamma(\alpha + 1) / (\Gamma(k + 1)\Gamma(\alpha - k + 1))$, and $\Gamma(\cdot)$ is the gamma function.

In model (4), the first three terms have already been used in existing trackers, which depict the reconstructed error, low-rank constraint, and sparse representation, respectively. The last term is our novel idea which represents fractional-order variation regularization.

In this optimization, we utilize low-rank constraint to force the appearance similarity of the candidate particles. This global restriction can help to acquire the structure feature of the object observations and prune the uncorrelated candidate particles. Since matrix rank minimization is



FIGURE 1: Nonoverlapping patches.

an NP-hard problem, we minimize a convex envelope of the rank function (nuclear norm) for alternative processing.

To realize robust tracking under fast motion and severe occlusion challenge, we introduce fractional-order variation regularization to the representation model. The variation method can model the variable selection problem in the bounded variation space. Functions in this space allow for the existence of jumping discontinuities. That is, the discontinuous features can be retained. Then, the appearance variation can be described effectively when the target undergoes fast motion. However, total variation regularization can only relate the information between two adjacent frames. Unlike this local processing, the fractional differential in equation (6) can involve more information from the front frames, which is helpful for acquiring more target feature information and dealing with the severe occlusion problems. Therefore, we employ fractional-order variation regularization under fractional-order bounded variation space to enhance the robustness of object tracking.

The candidate observation can be represented as a linear combination of target templates, and only a few templates are required to reliably represent the candidate image observation. Optimization problem (4) penalizes the representation matrix via L_1 norm, which can retain the useful information and remove the redundant information so that the optimal solution Z_t is sparse. We employ sparse representation to model the relationship between target candidates and target templates, which is helpful to deal with occlusion, because the residual error in the occlusion location is sparse. Currently, most of the sparse representation models utilize the target template to represent the candidate particles. These methods need to solve a large number of L_1 minimization problems. To reduce the computational cost, we use candidate particles' linear combination to represent the target template inversely. This is because the templates' number is smaller than that of the candidates. Then, the computational efficiency for tracking processing can be improved.

2.3. Observation Model. The observation model $p(y_t | x_t)$ measures the probability of the observed image y_t at the motion state x_t , which can describe the similarity between the target template and the candidate particle. Then, the candidate with maximal probability in equation (2) can be regarded as the tracking result. In this paper, we use the sparse coding coefficient Z_t in model (4) to estimate this similarity. The candidates with larger sparse coding coefficient have high probability to be the target, whereas the candidates with smaller coefficient are less likely to be the target. We define the observation model as

$$p(y_t^m | x_t^m) = \frac{Z_t^m}{\sum_{m=1}^M Z_t^m}, \quad m = 1, 2, \dots, M, \quad (7)$$

where the superscript m denotes the m -th candidate. In each frame, we crop out the optimal candidate as the tracking result.

3. Numerical Implementation

3.1. Alternating Iterative Algorithm. To solve the optimization problem in (4) for online tracking, we present an alternating iterative algorithm based on three update steps as follows:

Step 1: acquire the low-rank matrix $Z_{1,t}$ by

$$\min \frac{1}{2} \|D_t - Z_{1,t}\|_2^2 + \lambda \|Z_{1,t}\|_* \quad (8)$$

We solve this problem with the FISTA algorithm. Define $f(Z_{1,t}) = 1/2 \|D_t - Z_{1,t}\|_2^2$, $g(Z_{1,t}) = \|Z_{1,t}\|_*$, and

$$p_L(Y) = \operatorname{argmin}_{Z_{1,t}} \left\| Z_{1,t} - \left(Y - \frac{1}{L} \nabla f(Y) \right) \right\|_F^2 + Lg(Z_{1,t}), \quad (9)$$

where L is the Lipschitz constant for the function ∇f . The details of the FISTA algorithm can be summarized as follows:

- (1) Initialization: $(Z_{1,t})^0 = Y^1$ and $\gamma^1 = 1$.
- (2) Iteration:

$$\left\{ \begin{array}{l} (Z_{1,t})^j = p_L(Y^j), \\ \gamma^{j+1} = \frac{1 + \sqrt{1 + 4(\gamma^j)^2}}{2}, \\ Y^{j+1} = (Z_{1,t})^j + \left(\frac{\gamma^j - 1}{\gamma^{j+1}} \right) \left((Z_{1,t})^j - (Z_{1,t})^{j-1} \right), \end{array} \right. \quad (10)$$

where $j = 1, 2, \dots, J$. The terminal condition is set by the duality gaps.

Step 2: introduce the fractional-order variation regularization by

$$\min \frac{1}{2} \|Z_{1,t} - Z_{2,t}\|_2^2 + \lambda_2 \|\nabla^\alpha Z_{2,t}\|_1. \quad (11)$$

We solve this model by an adaptive primal-dual algorithm [17] formulated as follows:

(1) Initialization: $\bar{Z}_{2,t}^0 = Z_{2,t}^0$, $p^0 = 0$, $\tau_0, \sigma_0 > 0$.

(2) Iteration:

$$\left\{ \begin{array}{l} p^{n+1} = \frac{(p^n + \sigma_n \nabla^\alpha \bar{Z}_{2,t}^n)}{\max(1, |p^n + \sigma_n \nabla^\alpha \bar{Z}_{2,t}^n|)}, \\ Z_{2,t}^{n+1} = \frac{(Z_{2,t}^n - \tau_n \nabla^{\alpha*} p^{n+1} + \tau_n \lambda_2 Z_{1,t})}{(1 + \tau_n \lambda_2)}, \\ \theta_n = \frac{1}{\sqrt{1 + 2\gamma\tau_n}}, \\ \tau_{n+1} = \theta_n \tau_n, \\ \sigma_{n+1} = \frac{\sigma_n}{\theta_n}, \\ \bar{Z}_{2,t}^{n+1} = Z_{2,t}^{n+1} + \theta_n (Z_{2,t}^{n+1} - Z_{2,t}^n). \end{array} \right. \quad (12)$$

(3) Termination condition:

$$\begin{aligned} \varsigma(Z_{2,t}, p) &= \max \langle p', \nabla^\alpha Z_{2,t} \rangle - F^*(p') + G(Z_{2,t}) \\ &\quad - \min \langle p, \nabla^{\alpha*} Z_{2,t}' \rangle - F^*(p) + G(Z_{2,t}'), \end{aligned} \quad (13)$$

where

$$\begin{aligned} G(Z_{2,t}) &= \frac{1}{2} \|Z_{1,t} - Z_{2,t}\|_2^2, \\ F^*(p) &= \begin{cases} 0, & p \in P, \\ +\infty, & p \notin P. \end{cases} \end{aligned} \quad (14)$$

P is the dual space. $\varsigma(Z, p)$ is termed as the primal-dual gap, which vanishes only if (Z, p) is the saddle point.

Step 3: update the coding Z_t by the inverse sparse representation:

$$\min \frac{1}{2} \|T_t - D_t Z_t\|_2^2 + \lambda_1 \|Z_t\|_1. \quad (15)$$

This is a traditional linear regression problem. The solution of this model can be calculated by the LARS algorithm. We utilize the SPAMS optimization toolbox to realize this numerical calculation.

This three-step iteration updates one variable at a time with the other variables fixed. Finally, the representation coefficient in (4) can be acquired.

3.2. *Template Update.* In model (4), a fixed target template is insufficient to account for appearance change among successive frames. In this work, we address this issue by a dynamic update scheme as

$$T_t = \mu T_{t-1} + (1 - \mu) r_{t-1}, \quad \text{if } \|T_{t-1} - r_{t-1}\|^2 < \tau. \quad (16)$$

The target template T_t is defined as the weighted sum of the target template T_{t-1} and the tracking result r_{t-1} . The contributions of these two terms can be balanced by the weight μ . The threshold τ is determined empirically by measuring the dissimilarity. We set $\mu = 0.95$ and $\tau = 0.1$.

This update mechanism can overcome the target appearance change due to partial occlusion. We can retain the unoccluded patches in the target template and prune the occluded ones.

The details of the numerical implementation are shown in Algorithm 1. In our reverse sparse learning framework, the computational cost of updating Z_t in step 3 is $O(np^2)$, where n is the number of target templates and p is the number of image feature, whereas, in the traditional sparse learning framework, n is the number of target candidates. Because the templates' number is smaller than that of the candidates, the computational complexity for tracking processing can be reduced linearly. In our algorithm, the average frame rate of the video sequences is about 7 frames per second.

4. Experimental Results

In this section, we assess the proposed tracking algorithm by qualitative and quantitative experiments. The experiments are conducted on a set of benchmark sequences (faceocc1, faceocc2, girl, boy, deer, jumping, singer1, car4, david, cardark) with MATLAB. These sequences are categorized by their main challenging factors including occlusion, fast motion, illumination and scale variation, deformation, and background clutters. For each sequence, the initial value of the affine parameter can be acquired from the bounding box in the first frame, which is drawn manually, and then the affine parameter varies accordingly in the tracking process. We sample 300 candidate particles and regularize the target templates size to $32 * 32$. Furthermore, we set weight coefficients $\lambda = 0.5$, $\lambda_1 = 0.2$, $\lambda_2 = 0.05$, respectively.

Comparative studies with 7 state-of-the-art trackers including SCM [11], IST [13], LLR [14], DDL [15], CNT [18], MCPF [19], and VITAL [20] are carried out. In [11, 13–15], object tracking is modeled by low-rank and sparse representation. In [18], a convolutional neural network is incorporated into object tracking without training. In [19], object tracking is described in a multitask correlation particle filter framework. In [20], object tracking by detection framework is realized via adversarial learning. These comparisons mainly consider deep network [21], correlation

filter and adversarial learning have attracted much attention in complicated tasks of visual tracking.

4.1. Qualitative Results. Figures 2–6 compare the tracking results of the 8 trackers on 10 benchmark sequences qualitatively. In the following, we analyze the results according to the main challenging factors in each sequence.

Occlusion: in the faceoccl sequence, the target face undergoes frequent occlusion, which causes serious appearance changes. Figure 2(a) shows some representative tracking results in the sequence. These trackers can complete the tracking successfully. In the faceoccl2 sequence, the target face not only suffers from heavy partial occlusion, but also undergoes rotation. As shown in Figure 2(b), these trackers overcome the effect of occlusion to different degrees. When the face is occluded by a magazine heavily (e.g., #181 and #726), all the trackers can still achieve favorable results. But when the face undergoes both severe occlusion and in-plane rotation simultaneously around #481, most sparse trackers can detect the target well, whereas the CNT tracker deviates from the target. In the girl sequence, the target face involves heavy occlusion and out-of-plane and in-plane rotation simultaneously, as shown in Figure 2(c). When a man occludes the target girl around #500, the IST tracker drifts away from the target girl and tracks the man in turn. The MCPF tracker loses locating the target accurately as the influence of occlusion and scale variation after #428. The VITAL tracker loses the object around #428 and #457 but retraces the object finally. The DDL tracker starts to drift around #428. The SCM tracker fails to track the object as a result of rotation while our tracker can track the girl reliably in the entire sequence.

Fast motion: Figure 3 presents some tracking results over sequences whose target suffers from fast motion and motion blur. The ground truth indicates that the motion in these sequences is larger than 20 pixels. It is hard to locate the object, and it is rather challenging to describe the appearance changes caused by motion blur. Our tracker can achieve robust tracking in these sequences. But not all of the other trackers get promising results when the target faces these conditions. The boy sequence contains scenes with fast motion and motion blur, as well as out-of-plane and in-plane rotation. The DDL and LLR trackers cannot keep track of the object and drift to the other areas around #360, #490, and #602. The IST tracker outperforms the other trackers but also with some errors (e.g., #117). In the jumping sequence, the DDL and IST trackers cannot detect the target around #124, #180, #248, and #310, and the LLR tracker fails around #180, #248, and #310. In the deer sequence, the deer head undergoes fast motion, background clutter, and rotation. The DDL and LLR trackers lose the object from the start, and the IST tracker makes the drift phenomenon exist around #32 and #48.

Illumination and scale variation: Figure 4 shows some tracking results over sequences with severe illumination and scale variation. In the singer1 sequence, the stage light changes frequently. In the car4 sequence, the car crosses the overpass undergoing drastic illumination and scale changes. Most trackers can overcome the influences to obtain the object region based on low-rank constraint. The CNT tracker utilizes the normalized local image features to overcome this challenge. The MCPF tracker employs a particle sampling strategy to deal with large-scale variation problems. The VITAL tracker handles the scale variance sequence by acquiring the discriminative features based on the weight mask.

Deformation: in the david sequence, a moving face experiences strong nonrigid deformation due to pose variation and out-of-plane and in-of-plane rotations. We show some significance tracking results in Figure 5. Our tracker can track the target effectively on all the frames. It is attributed to the low-rank and reverse sparse characteristics of the tracking framework, which can learn the robust discriminative subspace. The IST and VITAL trackers also perform well with stable tracking results while the DDL and LLR trackers fail at different times. The CNT tracker deviates away in certain frames (e.g., #375 and #460). The MCPF tracker cannot locate the object effectively as scale variation (e.g., #460).

Background clutter: The cardark sequence includes scenes with background clutter and illumination variation. The car and the surrounding scene have similar color and texture as shown in Figure 6. Overall, most trackers can achieve better performance, whereas the LLR tracker drifts away from the car when the similar color or texture draws near to the car, such as around #60. The MCPF tracker cannot locate the car effectively as scale variation (e.g., #284 and #351).

4.2. Quantitative Results

4.2.1. Central-Pixel Error Comparison. This subsection compares the central-pixel error (CPE) of the 8 trackers on 10 sequences quantitatively as shown in Table 1. CPE records the Euclidean distance between the manually labeled ground truth and the central location of the tracked bounding box. The smaller the error is, the more accurate the tracking result will be. In Table 1, the smallest and the second smallest errors are marked in bold font for each sequence, and the last row presents the average performance of these trackers. From the results, it is clear that our tracker achieves the best or second best performance in terms of the CPE. CNT and VITAL trackers perform relatively well as well. Among these trackers, SCM, IST, LLR, and DDL are the most relevant trackers with us. However, our tracker outperforms the SCM tracker in deformation and occlusion sequences and outperforms IST, LLR, and DDL trackers in fast motion sequences. Furthermore, compared with the CNT tracker which models tracking in a convolutional neural network

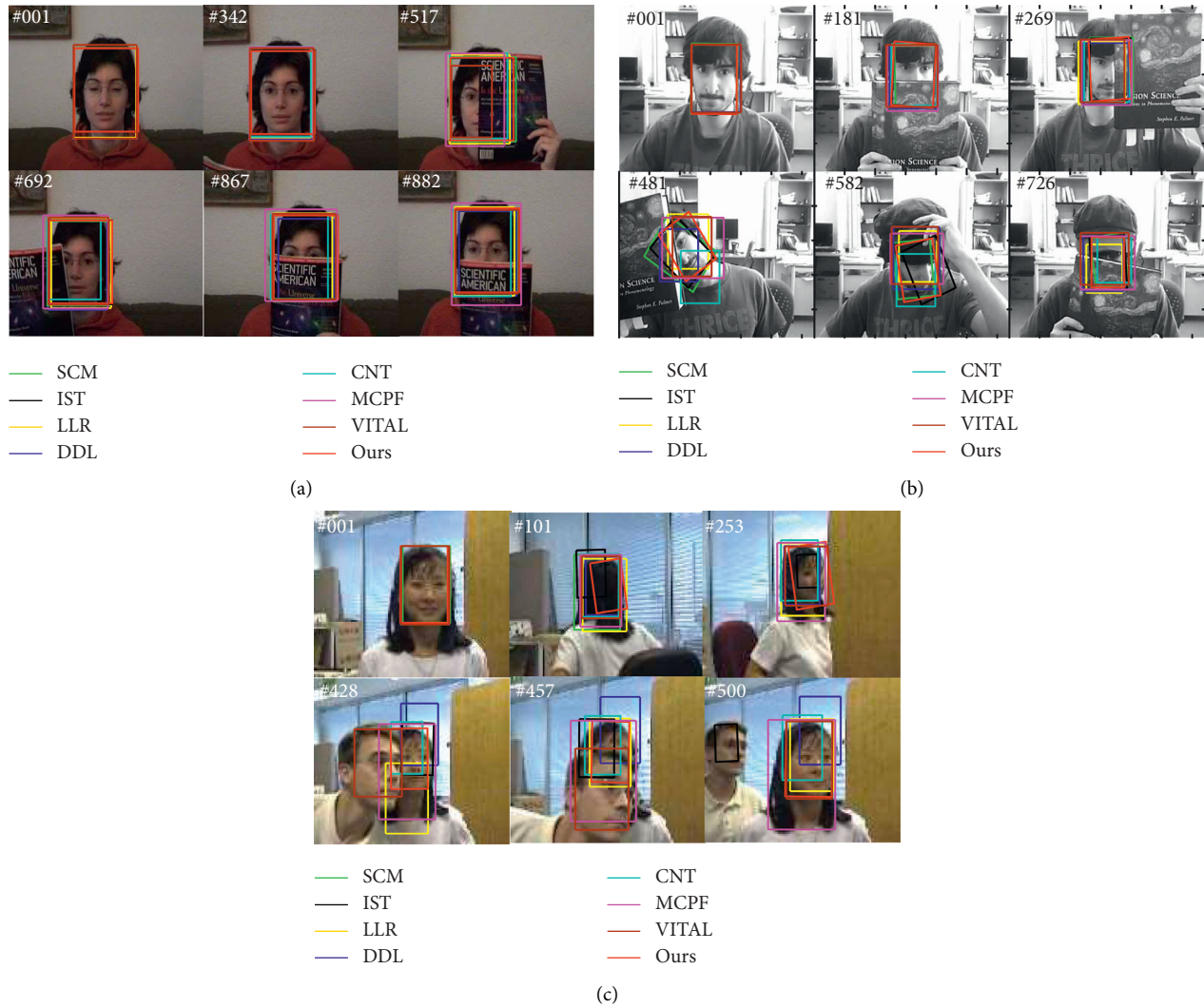


FIGURE 2: Qualitative results on three sequences with occlusion: (a) faceocc1, (b) faceocc2, and (c) girl.

Input: template matrix T_t , dictionary D_t , weight coefficients λ , λ_1 , λ_2 , and $\alpha = 1.8$.

Output: Z_t .

- (1) Initiate parameters: λ , λ_1 , $\lambda_2 > 0$.
- (2) While not converged do
- (3) Fixing other variables to update $Z_{1,t}$; (equation (8))
- (4) Fixing other variables to update $Z_{2,t}$; (equation (11))
- (5) Fixing other variables to update Z_t ; (equation (15))
- (6) Updating target template T_t ; (equation (16))
- (7) end

ALGORITHM 1: Numerical implementation for reverse low-rank sparse learning and fractional-order variation regularization.

framework, our tracker is more efficient in occlusion and deformation conditions. Compared with the MCPF tracker which models tracking in multitask correlation particle filter framework, our tracker is more efficient in

deformation and background clutter conditions. Compared with the VITAL tracker which models tracking via adversarial learning, our tracker is more efficient in occlusion conditions. These results indicate the robustness

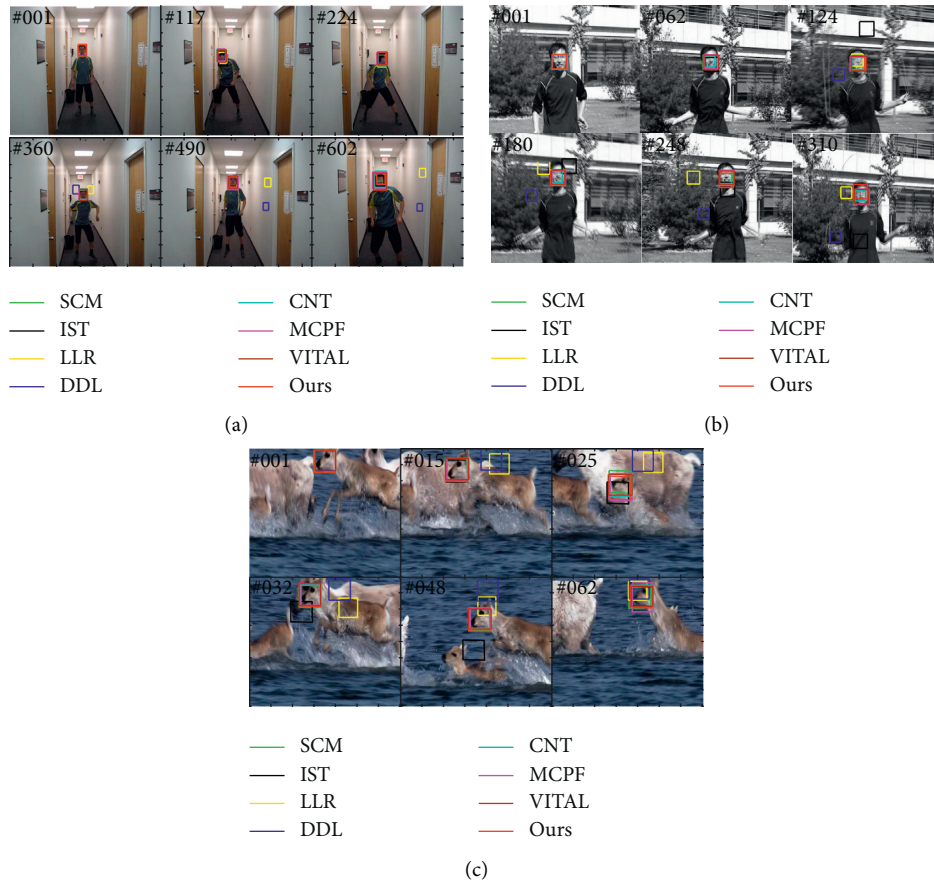


FIGURE 3: Qualitative results on three sequences with fast motion: (a) boy, (b) jumping, and (c) deer.

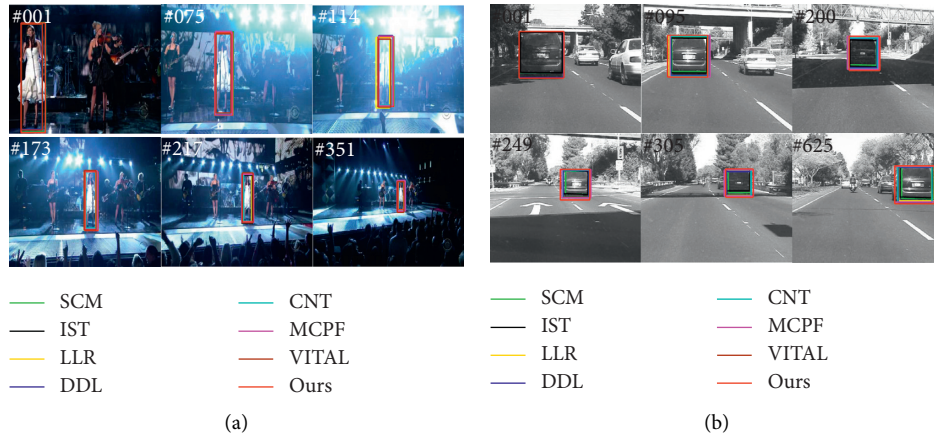


FIGURE 4: Qualitative results on two sequences with illumination and scale variation: (a) singer1 and (b) car4.

of our tracker to occlusion, illumination and scale variation, fast motion, deformation, and background clutter.

4.2.2. *Influence of Fractional-Order Variation.* This subsection compares the influence of fractional-order variation with first-order variation on the tracking results. Figure 7 draws the evolution curves of CPE versus frame

numbers on different differential orders. The experimental sequences are selected according to their main challenging factors. In most sequences, the fractional-order regularization is similar to the one obtained by using first-order regularization. But in complex occlusion condition, fractional-order regularization has an obvious advantage. In the faceocc2 sequence, especially from #576 to #819, the object face undergoes heavy appearance changes,

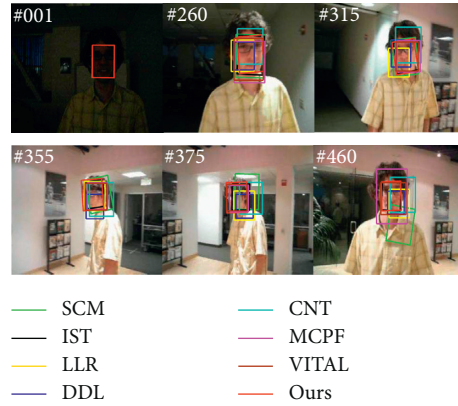


FIGURE 5: Qualitative results on david sequence with deformation.

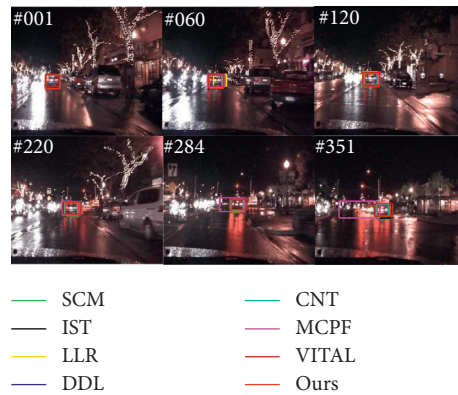


FIGURE 6: Qualitative results on cardark sequence with background clutters.

TABLE 1: CPE of the trackers.

	SCM	IST	LLR	DDL	CNT	MCPF	VITAL	Ours
faceocc1	14.4	14.7	15.0	13.1	16.8	22.0	16.7	14.2
faceocc2	8.3	8.7	10.6	5.0	18.0	9.7	11.3	7.8
girl	169.5	7.9	10.5	6.6	5.2	4.7	6.1	3.7
boy	2.8	3.8	68.4	60.7	2.4	4.2	2.4	2.6
jumping	4.4	41.9	46.2	63.8	5.6	3.1	3.5	7.7
deer	15.5	33.6	86.4	98.6	4.7	9.1	11.9	7.9
singer1	5.1	5.3	8.9	7.2	3.7	8.2	7.7	4.6
car4	4.3	2.8	14.6	13.5	1.5	3.7	7.7	2.7
david	30.0	2.3	9.5	3.2	16.1	17.7	4.8	3.0
cardark	2.7	2.8	3.6	1.4	1.0	21.9	3.8	2.2
average	25.7	12.4	27.4	27.3	7.5	10.4	7.6	5.6

occluded by a magazine and a hat. The tracking performance based on fractional-order regularization is much better than that of the first-order regularization. Similarly, in the girl sequence, especially from #90 to #110, the object face is occluded from local to global gradually, the fractional-order operator also performs better with smaller error. This implies that fractional-order regularization

should be used to take more neighboring frames information into account. This is mainly because the fractional differential is a global operation. Theoretically, the number of its expansion terms should be very large, but we take $K = 4$ for our tracking because the fractional-order computation will cost more time. Based on the average CPE, we set $\alpha = 1.8$ in Figures 2–6.

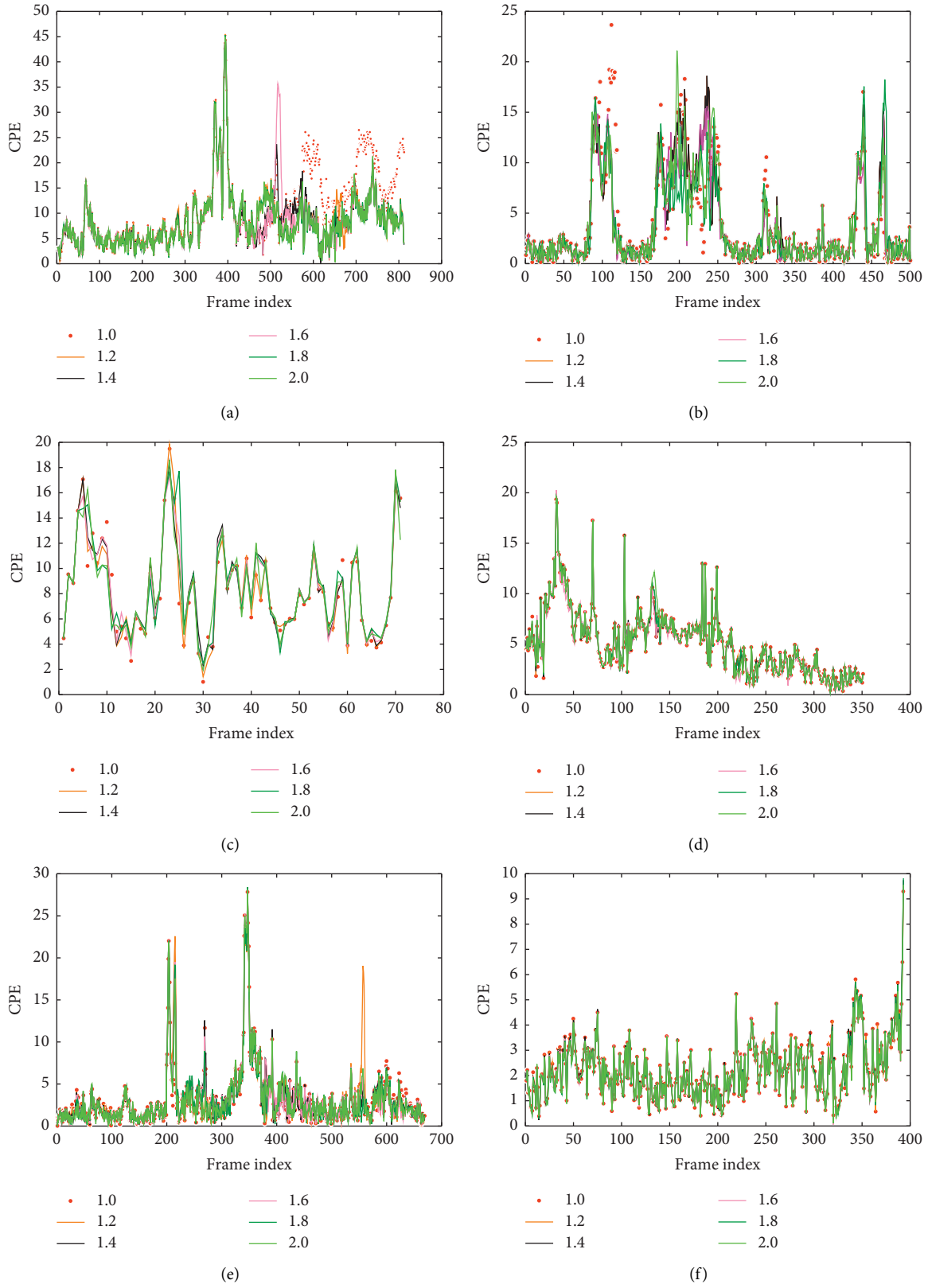


FIGURE 7: CPE under different differential orders. (a) faceocc2, (b) girl, (c) deer, (d) singer1, (e) david, and (f) cardark.

5. Conclusion

In this paper, we proposed a novel object tracking method based on reverse low-rank sparse learning and fractional-order variation regularization. Our tracker comprised some effective technical elements as follows. We utilized low-rank constraint to prune the uncorrelated candidate particles. We introduced fractional-order variation regularization to retain the discontinuous features and conquer the fast motion problem. Meanwhile, this regularization could also relate adjacent frame feature information to repress occlusion. Furthermore, we built an inverse sparse representation to reduce the computational cost for tracking processing. We gave an alternating iteration strategy for online tracking optimization. Qualitative and quantitative evaluation on benchmark sequences have demonstrated the robustness of our tracking algorithm, especially in complex occlusion and fast motion challenges. In the future, we will extend our tracker to deep learning for enhancing its discriminatory ability.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by National Nature Science Foundation of China under Grant 61703285 and Liaoning Natural Science Foundation of China under Grant 2019-MS-237.

References

- [1] X. Shi, H. Ling, Y. Pang, W. Hu, P. Chu, and J. Xing, "Rank-1 tensor approximation for high-order association in multi-target tracking," *International Journal of Computer Vision*, vol. 127, no. 8, pp. 1063–1083, 2019.
- [2] B. Fan, X. Li, Y. Cong, and Y. Tang, "Structured and weighted multi-task low rank tracker," *Pattern Recognition*, vol. 81, pp. 528–544, 2018.
- [3] H. Kasai, "Fast online low-rank tensor subspace tracking by CP decomposition using recursive least squares from incomplete observations," *Neurocomputing*, vol. 347, pp. 177–190, 2019.
- [4] Y. Sui, Y. Tang, L. Zhang, and G. Wang, "Visual tracking via subspace learning: a discriminative approach," *International Journal of Computer Vision*, vol. 126, no. 5, pp. 515–536, 2018.
- [5] B. Kang, W.-P. Zhu, D. Liang, and M. Chen, "Robust visual tracking via nonlocal regularized multi-view sparse representation," *Pattern Recognition*, vol. 88, pp. 75–89, 2019.
- [6] T. Zhang, C. Xu, and M. H. Yang, "Robust structural sparse tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 473–486, 2018.
- [7] Z. He, S. Yi, Y. M. Cheung, X. You, and Y. Y. Tang, "Robust object tracking via key patch sparse representation," *IEEE Transactions on Cybernetics*, vol. 47, no. 2, pp. 354–364, 2017.
- [8] Y. Qi, L. Qin, J. Zhang, S. Zhang, Q. Huang, and M.-H. Yang, "Structure-aware local sparse coding for visual tracking," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3857–3869, 2018.
- [9] A. Elnakeeb and U. Mitra, "Line constrained estimation with applications to target tracking: exploiting sparsity and low-rank," *IEEE Transactions on Signal Processing*, vol. 66, no. 24, pp. 6488–6502, 2018.
- [10] T. Zhou, F. Liu, H. Bhaskar, and J. Yang, "Robust visual tracking via online discriminative and low-rank dictionary learning," *IEEE Transactions on Cybernetics*, vol. 48, no. 9, pp. 2643–2655, 2017.
- [11] W. Zhong, H. Lu, and M. H. Yang, "Robust object tracking via sparse collaborative appearance model," *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 23, no. 5, pp. 2356–2368, 2014.
- [12] T. Zhang, S. Liu, N. Ahuja, M.-H. Yang, and B. Ghanem, "Robust visual tracking via consistent low-rank sparse learning," *International Journal of Computer Vision*, vol. 111, no. 2, pp. 171–190, 2015.
- [13] D. Wang, H. Lu, Z. Xiao, and M. H. Yang, "Inverse sparse tracker with a locally weighted distance metric," *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 24, no. 9, pp. 2646–2657, 2015.
- [14] Y. Sui and L. Zhang, "Robust tracking via locally structured representation," *International Journal of Computer Vision*, vol. 119, no. 2, pp. 110–144, 2016.
- [15] Y. Sui, G. Wang, L. Zhang, and M.-H. Yang, "Exploiting spatial-temporal locality of tracking via structured dictionary learning," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1282–1296, 2018.
- [16] P. P. Dash and D. Patra, "Efficient visual tracking using multi-feature regularized robust sparse coding and quantum particle filter based localization," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 2, pp. 449–462, 2019.
- [17] D. Tian, D. Xue, and D. Wang, "A fractional-order adaptive regularization primal-dual algorithm for image denoising," *Information Sciences*, vol. 296, no. 1, pp. 147–159, 2015.
- [18] K. Zhang, Q. Liu, Y. Wu, and M. Yang, "Robust visual tracking via convolutional networks without training," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1779–1792, 2016.
- [19] T. Zhang, C. Xu, and M. H. Yang, "Multi-task correlation particle filter for robust object tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4335–4343, Honolulu, HI, USA, July 2017.
- [20] Y. Song, C. Ma, X. Wu et al., "Vital: visual tracking via adversarial learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8990–8999, Salt Lake City, UT, USA, June 2018.
- [21] P. Li, B. Chen, W. Ouyang et al., "Gradnet: gradient-guided network for visual object tracking," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 6162–6171, Seoul, Republic of Korea, October–November 2019.