Spatial-Temporal Context-Aware Tracking

Yuqi Han, *Student Member, IEEE*, Chenwei Deng, *Senior Member, IEEE*, Boya Zhao, *Student Member, IEEE*, and Baojun Zhao

***Abstract*—Discriminative correlation filters (DCFs) have recently achieved competitive performance in visual tracking benchmarks. However, most of the existing DCF trackers only consider the spatial features of the target and could hardly benefit from the inter-frame and historical information, which may degrade the tracking performance when occlusion and deformation occurs. To tackle the above-mentioned issues, in this letter, by introducing the temporal constrain into the DCF tracker, we advocate our spatial-temporal context-aware tracker. Through jointly modeling the spatial context and historical target information, our tracker could not only adapt the appearance change but also maintain a relatively stable filter due to the small target variation between inter-frames. Furthermore, we show that the proposed objective formula could be directly solved using the Alternating Direction Method of Multipliers (ADMM) technique with low computational cost. Experiments on the large-scale benchmark demonstrate that the proposed trackers perform favorably against other state-of-the-art methods.**

***Index Terms*—Visual tracking, correlation filter, spatial-temporal constrain.**

I. Introduction

VISUAL tracking is one of the fundamental task in computer vision with a plethora of applications. Numerous works [1]–[6] have been extensively studied to overcome tracking failures caused by the challenging attributes in tracking process. Recently, DCF-based trackers have attracted broad attention [7], [8] due to their simplicity and superior speed. Instead of randomly extracting the target and background patches in a small region, DCF trackers approximate dense sampling strategy by circularly shifting the training samples. Thus, taking advantage of the property of the circulant matrix, DCF-based trackers obtain the trained filter for each frame without solving the computational-cost matrix inversion.

In recent years, some works [9]–[14] focus on utilizing either the hand-crafted or the CNN-based feature to model the target appearance into the DCF tracking framework. While other works are dedicated to addressing some inherent limitations for DCF tracking. Danelljan et al. add an extra scale filter in [15] to address the scale issues. In [16]–[18], the authors design a re-detection scheme for long-term tracking. LBCF [19] addresses the problems resulting from learning with circular correlation from small patches. It proposes a learning framework that artificially increases the filter size by implicit zero-padding to the right and down. SRDCF [20] reformulates the loss function of to penalize the non-zero value near the template boundaries. In [21], [22], the author employ ensembling learning strategy to improve the tracking accuracy.

However, most of the aforementioned trackers predict the target state using the current target template and spatial contextual information. Thus, those trackers could hardly benefit from motion and inter-frame information and aggressively update the classifier to handle the target appearance variations over time such as partial occlusion and deformation, which lead model drifts and degrade the tracking performance. Although Martin et al. propose a learning formulation which could dynamically adjust the weight of the training samples by estimating their quality in SRDCFdecon [23] and CCOT [24] to model the inter-frame information of the target and upgrade tracking accuracy. However, such performance gain is not obtained without any additional cost. Since the aforementioned trackers both employ SRDCF as the base supervised learning algorithm, which destroy the closed-form solution of traditional DCF trackers and lose the characteristic speed and real-time capability. Moreover, both trackers have to store plenty of samples for optimization which occupy lots of memory and may hinder the practical use.

To balance the tradeoff between the long-term memory and sudden appearance variants, motivated by [25], we construct a framework which could jointly model the spatial and temporal information of the target and advocate our Spatial-Temporal Context-Aware tracker (STCAT) accordingly in this letter. Specifically, the surrounding contextual patches are employed into the training stage to resist tracking drift caused by the external distractions. Furthermore, by introducing the temporal constrain, our tracker could maintain a long-term memory of the target appearance compared with the baseline Context-Aware trackers (CA) [26] in the case of the significant appearance variation occurs. We validate the effectiveness of our Spatial-Temporal Context Learning framework on the Online Tracking Benchmark (OTB-100) [27] with the baseline CA trackers and other state-of-the-art tracking algorithms. Extensive experiments demonstrate the effectiveness of the proposed framework.

The contributions of this letter are as follows:

- A STCAT model is presented by incorporating both the spatial and temporal constrain into the DCF framework. Based upon the Spatial-Temporal constrain, the STCAT model could provide a more robust appearance model compared with the baseline model.
- An efficient ADMM algorithm is employed to solving the proposed model, where each sub-problem has its element-wisely closed-form solution.

Manuscript received November 10, 2018; revised January 18, 2019; accepted January 20, 2019. Date of publication January 29, 2019; date of current version February 14, 2019. This work was partially supported by the National Natural Science Foundation of China under Grant 91438203. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yap-Peng Tan. *(Corresponding author: Chenwei Deng.)*

The authors are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: yuqi_han@bit.edu.cn; cwdeng@bit.edu.cn; zhaoboya@bit.edu.cn; zbj@bit.edu.cn).

Digital Object Identifier 10.1109/LSP.2019.2895962

II. Proposed Algorithm

A. Spatial-Temporal Context-Aware Tracker

We build the Spatial-Temporal framework upon the well-known Context-Aware tracker [26]. In [26], Muller et al. model the global contextual information during the filter training stage to resist the model drift caused by the limited use of context information. Specifically, k contextual patches around the target patch would be extracted around the estimated target patch. Afterwards, the feature circulant matrix would be calculated for both the target patch and the context patches, denoting as $\mathbf{A}_0$ and $\mathbf{A}_i$. Since the ideal filter would obtain a high response value for the target patch and a response close to zero for the background patch, the corresponding labels for the target and the background context would be assigned as $\mathbf{y}$ and zero. Thus, the objective function could be expressed as a ridge regression in the spatial domain as below, where the discriminative filter is denoted by the vector $\mathbf{w}$. For more details, please refer to [26].

$$\begin{aligned}\mathbf{w} &= \arg\min \|\mathbf{A}_0\mathbf{w}-\mathbf{y}\|_2^2 + \lambda_1\|\mathbf{w}\|_2^2 + \lambda_2\sum_{i=1}^{k}\|\mathbf{A}_i\mathbf{w}\|_2^2 \\ &= \arg\min \|\mathbf{B}\mathbf{w}-\mathbf{Y}\|_2^2 + \lambda_1\|\mathbf{w}\|_2^2 \end{aligned} \tag{1}$$

Here, $\mathbf{B}=\left[\mathbf{A_0}, \sqrt{\lambda_2}\mathbf{A_1}, \ldots, \sqrt{\lambda_2}\mathbf{A_k}\right]^T$ and $\mathbf{Y}=[\mathbf{y}, \mathbf{0}, \ldots, \mathbf{0}]^T$, which indicates the stacked feature matrix and regression label for the target and spatial context.

However, albeit from the above advantages, CA trackers still suffer from stability-plasticity dilemma since it updates the target template for each frame like other traditional DCF trackers. To tackle the above issues, motivated by the online Passive-Aggressive algorithm [28], we introduce a robust and novel updating strategy upon the baseline CA tracker. The objective formulation could be represented as below. For simplification, all derivations would be inferred under the single-channel feature condition.

$$\mathbf{w} = \arg\min \|\mathbf{B}\mathbf{w}-\mathbf{Y}\|_2^2 + \lambda_1\|\mathbf{w}\|_2^2 + \mu\|\mathbf{w}-\mathbf{w}_{t-1}\|_2^2 \tag{2}$$

The first two terms (the basic loss function for CA tracker) could be considered as the aggressive updating scheme, which aims at learning the target appearance for each frame. While the third term regularize the whole tracker by forcing the newly obtained filter close to the previous one in order to resist the interference caused by the corrupted samples. By introducing the temporal constrain, the proposed tracker would train the model at the batch level (considering target's historical appearance), instead of instance-wise updating in the baseline tracker, which naturally inherits the tradeoff between aggressive and passive model learning [28]. More comparison experiments and corresponding analysis would be shown in the Section III-C.

B. Optimization

In this subsection, we show that the proposed loss function could be solved efficiently by introducing the constrain $\mathbf{w} \equiv \mathbf{h}$. Thus, we construct the Augmented Lagragian function $L(\mathbf{w}, \mathbf{h}, \mathbf{I}, \gamma)$ as:

$$\begin{aligned}L(\mathbf{w}, \mathbf{h}, \mathbf{I}, \gamma) = \|\mathbf{B}\mathbf{w}-\mathbf{Y}\|_2^2 + \lambda_1\|\mathbf{h}\|_2^2 + \mu\|\mathbf{w}-\mathbf{w}_{t-1}\|_2^2 \\ + \mathbf{I}(\mathbf{w}-\mathbf{h}) + \frac{\gamma}{2}\|\mathbf{w}-\mathbf{h}\|_2^2\end{aligned} \tag{3}$$

Here, $\mathbf{I}$ denotes the complex Lagrangian multiplier and γ is the penalty parameter. Furthermore, we could substitute the last two Augmented terms into $\frac{\gamma}{2}\|\mathbf{w}-\mathbf{h}+\rho\|_2^2$, by introducing the dual variable ρ with the equation $\rho = \frac{1}{\mathbf{I}}\gamma$.

Thus, according to the ADMM model [29], the reformulated Augmented Lagrangian function $L(\mathbf{w}, \mathbf{h}, \rho)$ could be solved using a series of iterations as:

$$\begin{cases}\mathbf{w}^{i+1} = \arg\min_{\mathbf{w}} L(\mathbf{w}^i, \mathbf{h}^i, \rho^i) \\ \mathbf{h}^{i+1} = \arg\min_{\mathbf{h}} L(\mathbf{w}^{i+1}, \mathbf{h}^i, \rho^i) \\ \rho^{i+1} = \rho^i + \mathbf{w}^{i+1} - \mathbf{h}^{i+1}\end{cases}$$

Accordingly, we detail the solution by solving the decomposed subproblems successively.

Subproblem for $\mathbf{w}$: The loss function over variable $\mathbf{w}$ could be reformulated as :

$$\|\mathbf{B}\mathbf{w}-\mathbf{Y}\|_2^2 + \mu\|\mathbf{w}-\mathbf{w}_{t-1}\|_2^2 + \frac{\gamma}{2}\|(\mathbf{w}-\mathbf{h}+\rho)\|_2^2 \tag{4}$$

Therefore, we derive the close-form solution for $\mathbf{w}$ as:

$$\mathbf{w} = \left(\mathbf{B}^H\mathbf{B} + \mu + \frac{\gamma}{2}\right)^{-1}\left(\mathbf{B}^H\mathbf{Y} + \mu\mathbf{w}_{t-1} + \frac{\gamma}{2}\mathbf{h} - \frac{\gamma}{2}\rho\right) \tag{5}$$

Recalling that $\mathbf{B}$ is the stacked feature matrix for the target patch $\mathbf{A}_0$ and its surrounding context patches $\mathbf{A}_k$. Meanwhile, $\mathbf{Y}$ indicates the corresponding label for those patches. Therefore, we have the formulations as below

$$\begin{cases}\mathbf{B}^H\mathbf{B} = \mathbf{A}_0^H\mathbf{A}_0 + \lambda_2\sum_{i=1}^{k}\mathbf{A}_i^H\mathbf{A}_i \\ \mathbf{B}^H\mathbf{Y} = \mathbf{A}_0^H\mathbf{y}\end{cases}$$

According to the property of circulant matrix that, $\mathbf{X} = \mathbf{F}diag(\hat{\mathbf{x}})\mathbf{F}^H$. The variable above could be represented element-wisely in Fourier domain. Thus, we have the formulations as below:

$$\begin{cases}\mathbf{B}^H\mathbf{B} = \mathbf{F}diag(\hat{a}_0^* \odot \hat{a}_0 + \lambda_2\sum_{i=1}^{k}\hat{a}_i^* \odot \hat{a}_i)\mathbf{F}^H \\ \mathbf{B}^H\mathbf{Y} = \mathbf{F}diag(\hat{a}_0^* \odot \hat{y})\mathbf{F}^H\end{cases}$$

$\hat{a}_0$ and $\hat{a}_i$ indicate the complex feature vector for target patch and context patches, in conjunction with the circulant feature matrix $\mathbf{A}_0$ and $\mathbf{A}_i$, respectively. Therefore, by substituting $\mathbf{B}^H\mathbf{Y}$ and $\mathbf{B}^H\mathbf{B}$, we obtain the element-wise closed-form solution for the filter $\mathbf{w}$:

$$\mathbf{w} = \frac{\hat{a}_0^* \odot \hat{y} + \mu\mathbf{w}_{t-1} + \frac{\gamma}{2}\mathbf{h} - \frac{\gamma}{2}\rho}{\hat{a}_0^* \odot \hat{a}_0 + \lambda_2\sum_{i=1}^{k}\hat{a}_i^* \odot \hat{a}_i + \mu + \frac{\gamma}{2}} \tag{6}$$

Subproblem for $\mathbf{h}$: Similarly, the subproblem for variable $\mathbf{h}$ could be seperated from the original loss function as:

$$\lambda_1\|\mathbf{h}\|_2^2 + \frac{\gamma}{2}\|(\mathbf{w}-\mathbf{h}+\rho)\|_2^2 \tag{7}$$

Thus, by setting its complex gradient $\bigtriangledown_{\mathbf{h}^H} L(\mathbf{h})$ to zero, we acquire the optimal $\mathbf{h}$:

$$\mathbf{h} = \left(\lambda_1 + \frac{\gamma}{2}\right)^{-1}\left(\frac{\gamma}{2}\mathbf{w} + \frac{\gamma}{2}\rho\right) \tag{8}$$

Convergence: It should be mentioned that, since the overall STCAT model is convex and each subproblem has a close-form solution. According to the theoretical property of ADMM model [30], its convergence could be guaranteed on condition

Algorithm 1: STCAT Tracking Algorithm.

Input: Current Image $\mathbf{I}_t$, Previous Position P_{t-1}, Learned Filter $\mathbf{w}_{t-1}$.

Output: Estimated Target Position P_t, Updated Filter $\mathbf{w}_t$.

1. **Repeat:**
2. Extract target patch's feature $\mathbf{a}_0$ and context's features $\mathbf{a}_i$.
3. Initialize the Lagrangian multiplier $\mathbf{I}$ and γ
4. Iteratively train the filter $\mathbf{w}_t$ using the ADMM model (Eq. 6 to Eq. 9) with the discriminative features $\mathbf{a}_0$, $\mathbf{a}_i$ and previous learned filter $\mathbf{w}_{t-1}$.
5. Estimate the current target position P_t by computing response map as other standard DCF trackers.
6. **Until** the end of video sequence.

that the quadratic penalty parameter γ is non-decreasing and $\sum_{i=1}^{+\infty}\gamma^i = +\infty$. Thus, we design the updating scheme for γ as below to reach the global optimum:

$$\gamma^{i+1} = \min(\gamma_{max}, \beta\gamma^i) \tag{9}$$

Here, γ_{max} indicates the maximum value for γ and β is the scale step factor. The residual term $(\mathbf{w}^{i+1} - \mathbf{h}^{i+1})$ is used for contriolling the optimization procedure. According to the experiment, the iteration number is set to 3 for balancing the performance and efficiency.

Computational Analysis: It should be mentioned that even both BACF and our STCAT tracker employ ADMM model to optimize the objective function, the computation of the proposed STCAT algorithm are fully carried out element-wisely in Fourier domain taking advantage of the property of circulant matrix, resulting a low computational burden $\mathcal{O}(NlogN)$. On the other hand, BACF has to utilize the Sherman-Morrision formula to solve the matrix inversion for several independent linear systems since it employs the densely cropped contextual patches into filter training. Thus, the trained filter in BACF tracker could not be represented element-wisely as our tracker, which demonstrate that our model is more suitable for the time-critical tracking application. For better understanding and details about our tracker, please refer to Algorithm 1.

III. Experiments

To evaluate the effectiveness and robustness of the proposed STCAT framework, we validate it with five popular DCF based trackers (MOSSE [31], DCF [9], DSST [15], SAMF [32] and Staple [33]) and denote them as MOSSE-STCA, DCF-STCA, DSST-STCA, SAMF-STCA and Staple-STCA respectively. We employ the same hand-crafted features (Grayscale, HOG and CN) as their baseline CA trackers [26] instead of utilizing the deep features to investigate the performance gain due to the employment of the spatial-temporal constrain. We also compare our tracker with other state-of-the-art tracking algorithms reported in recent years, namely SRDCF [20], SRDCFdecon [23], BACF [34], LCT [16] and Muster [17].

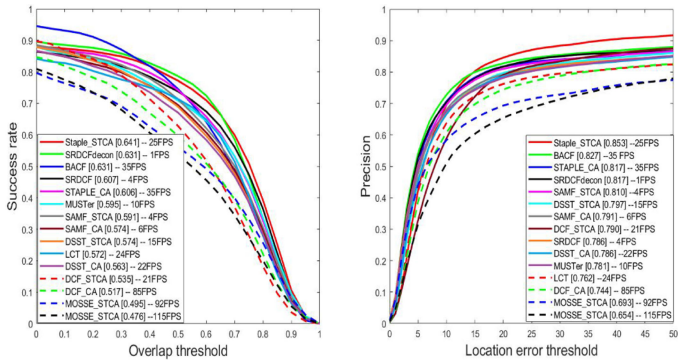

Fig. 1. The overall performance of success and precision plot on OTB100. The scores and running speed of each tracker are reported in the legend. Best viewed in color.

A. Parameter Setting

For fair comparison, same parameters (i.e., λ_1 and λ_2) are employed for our STCAT trackers as their corrsponding baseline trackers. Furthermore, we also utilize the same context sampling strategy as CA in our tracking models. The penalty parameter μ is set to 25 throughout all the experiments. As for the ADMM optimization, the initial parameter γ_0 and the upper-bound parameter γ_{max} are set to 5 and 25, severally. The scale factor β is fixed to 3, while the model iterates 3 times for each frame to balance the accuracy and efficiency. Please refer to CA [26] for more implementation details. In addition, we employ the public available codes or results of other trackers for fair comparison.

B. Benchmark and Evaluation Metrics

We conduct the experiments on OTB-100 [27], which contains 100 fully annotated challenging sequences with 11 challenging attributes. Two common metrics (precision and overlap ratio) are utilized to measure the performance for each candidate trackers. Both the precision and success plots show the mean scores over all the sequences in the tracking benchmark.

C. Quantitative Results

1) Overall Performance: Fig. 1 illustrates the overall performance of the proposed trackers against their baseline trackers and other state-of-the-art DCF trackers. One can see that the proposed trackers outperform their baseline methods by a large margin while running at approximately 0.7 times of the speed of their corresponding baseline trackers. Meanwhile, the proposed Staple-STCA tracker ranks first against other representative state-of-the-art trackers in terms of success plot and precision plot.

2) Attribute-based Evaluation: Since the target undergoes several different attributes during tracking procedure, it's crucial to investigate the tracking performance of the proposed trackers over such factors. Fig. 2 illustrates the precision scores for the typical proposed trackers (Staple-STCA, DCF-STCA, MOSSE-STCA) on 11 challenging factors. One can see that, the Staple-STCA tracker ranks the first place in most cases. Furthermore, DCF-STCA perform favorably against other sophisticated trackers (e.g., SRDCF and SRDCFdecon). It should be mentioned that even we don't deliberately add any occlusion detection mechanisms [35], [36], our tracker could tackle

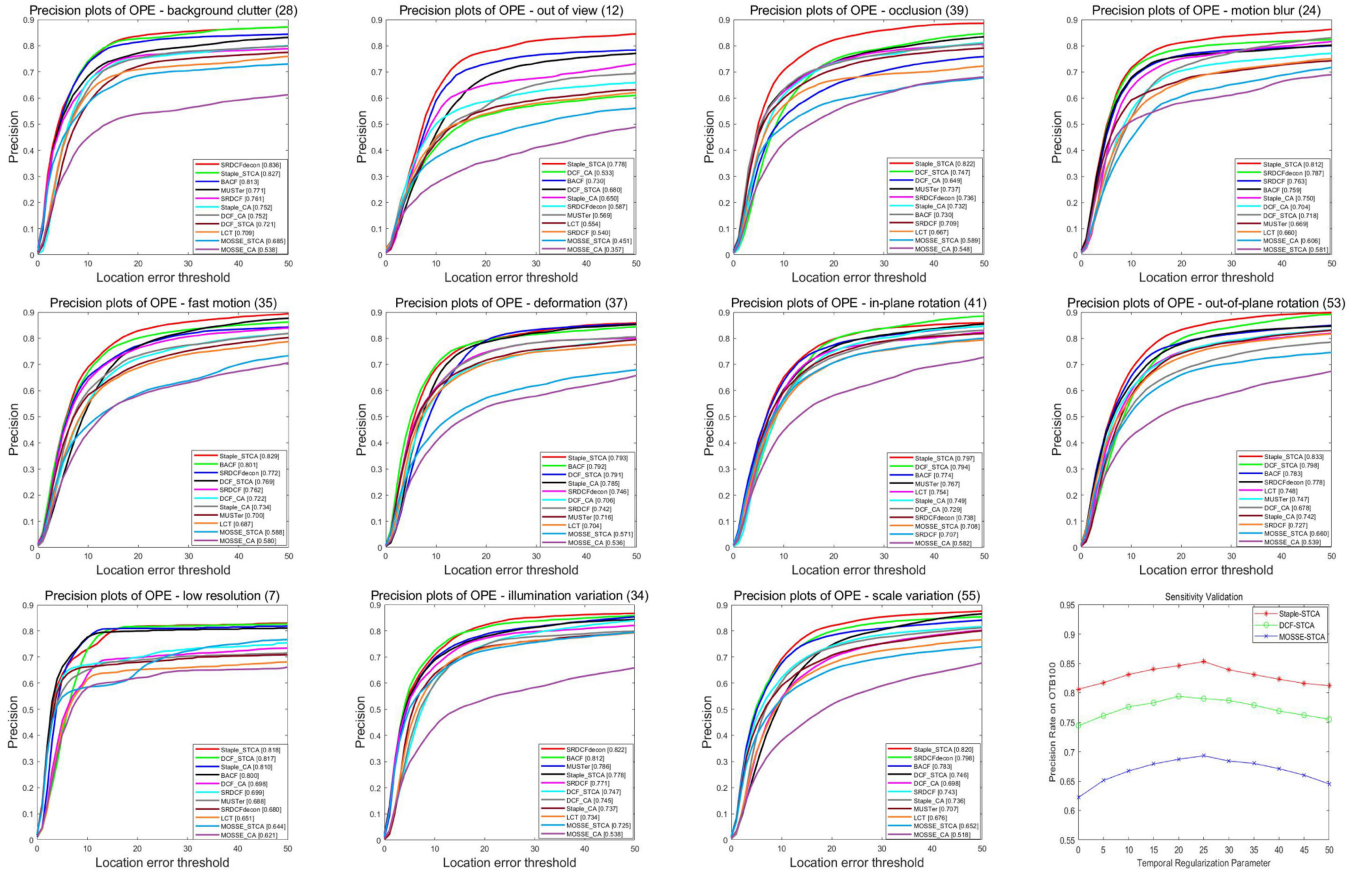

Fig. 2. The precision plots over 11 challenging attributes and the impacts of the temporal regularization parameter μ on OTB-100.

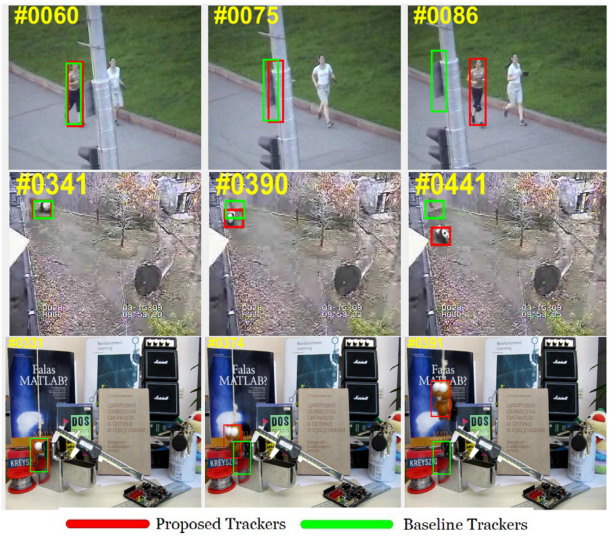

Fig. 3. Tracking results of the proposed trackers compared to their baseline methods. The baselune trackers from top to bottom are: MOSSE-CA, DCF-CA and Staple-CA, respectively.

such issue favorably. This could be attributed to the consideration for the spatial-temporal information. Specifically, when the target is partial or fully occluded, instead of aggressively updating the model like other CF-based trackers, our tracker would conservatively update the template in order to maintain the stability of the filter as well as keeping purity for the training sample set.

3) Parametric sensitivity: In addition, we conduct the comparison experiments on the same typical trackers (Staple-STCA, DCF-STCA, MOSSE-STCA) with different configurations of μ to investigate the parametric sensitivity of the regularization parameter. The precision rate is shown in the bottom-right of Fig. 2. One can see that the best performance is achieved around $\mu = 25$. It should also be mentioned that when $\mu = 0$, the tracker would degenerate into the baseline CA tracker.

D. Qualitative Results

In this subsection, we show the Qualitative comparisons between proposed trackers against their baselines in Fig. 3. Even though the CA trackers incorporate the context into filter training stage, they suffer from model drift when serve occlusion, rotation and deformation occurs. While with the employment of the temporal constrain, our trackers could resist these distraction while locating on the target precisely.

IV. Conclusion

In this letter, we propose a generic framework for the well-known DCF trackers which jointly model the spatial and temporal information synchronously. Extensive experiments on a large-scale benchmark have demonstrated that the proposed framework improved the tracking accuracy for all the baseline DCF-based trackers, while maintaining a computational cost.

REFERENCES

[1] X. Dong, J. Shen, W. Wang, Y. Liu, L. Shao, and F. Porikli, "Hyperparameter optimization for tracking with continuous deep Q-learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 518–527.

[2] X. Dong and J. Shen, "Triplet loss in siamese network for object tracking," in *Proc. IEEE Eur. Conf. Comput. Vis.*, 2018, pp. 459–474.

[3] B. Ma, L. Huang, J. Shen, L. Shao, M.-H. Yang, and F. Porikli, "Visual tracking under motion blur," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5867–5876, Dec. 2016.

[4] B. Ma, L. Huang, J. Shen, and L. Shao, "Discriminative tracking using tensor pooling," *IEEE Trans. Cybern.*, vol. 46, no. 11, pp. 2411–2422, Nov. 2016.

[5] C. Deng, Y. Han, and B. Zhao, "High-performance visual tracking with extreme learning machine framework," *IEEE Trans. Cybern.*, to be published.

[6] B. Ma, H. Hu, J. Shen, Y. Liu, and L. Shao, "Generalized pooling for robust object tracking," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4199–4208, Sep. 2016.

[7] Z. Chen, Z. Hong, and D. Tao, "An experimental survey on correlation filter-based tracking," 2015, arXiv:1509.05520.

[8] F. Liu, C. Gong, X. Huang, T. Zhou, J. Yang, and D. Tao, "Robust visual tracking revisited: From correlation filter to template matching," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2777–2790, Jun. 2018.

[9] 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.

[10] M. Danelljan, F. S. Khan, M. Felsberg, and J. V. D. Weijer, "Adaptive color attributes for real-time visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1090–1097.

[11] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Hierarchical convolutional features for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 3074–3082.

[12] Y. Qi *et al.*, "Hedged deep tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4303–4311.

[13] C. Ma, Y. Xu, B. Ni, and X. Yang, "When correlation filters meet convolutional neural networks for visual tracking," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1454–1458, Oct. 2016.

[14] Y. Han, C. Deng, Z. Zhang, J. Li, and B. Zhao, "Adaptive feature representation for visual tracking," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 1867–1870.

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

[16] C. Ma, X. Yang, C. Zhang, and M.-H. Yang, "Long-term correlation tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5388–5396.

[17] C. Wang, X. Mei, D. Prokhorov, Z. Hong, Z. Chen, and D. Tao, "Multi-store tracker(MUSTer): A cognitive psychology inspired approach to object tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 749–758.

[18] X. Dong, J. Shen, D. Yu, W. Wang, J. Liu, and H. Huang, "Occlusion-aware real-time object tracking," *IEEE Trans. Multimedia*, vol. 19, no. 4, pp. 763–771, Apr. 2017.

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

[20] 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.*, 2015, pp. 4310–4318.

[21] J. Li, C. Deng, R. Y. Xu, D. Tao, and B. Zhao, "Robust object tracking with discrete graph based multiple experts," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2736–2750, Jun. 2017.

[22] J. Guo and T. Xu, "Deep ensemble tracking," *IEEE Signal Process. Lett.*, vol. 24, no. 10, pp. 1562–1566, Oct. 2017.

[23] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Adaptive decontamination of the training set: A unified formulation for discriminative visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1430–1438.

[24] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Convolutional features for correlation filter based visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop*, 2015, pp. 58–66.

[25] F. Li, C. Tian, W. Zuo, L. Zhang, and M.-H. Yang, "Learning spatial-temporal regularized correlation filters for visual tracking," 2018, arXiv:1803.08679.

[26] M. Mueller, N. Smith, and B. Ghanem, "Context-aware correlation filter tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1396–1404.

[27] Y. Wu, J. Lim, and M.-H. Yang, "Object tracking benchmark," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1834–1848, Sep. 2015.

[28] S. Shalev-Shwartz, K. Crammer, O. Dekel, and Y. Singer, "Online passive-aggressive algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 1229–1236.

[29] S. Boyd *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[30] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2010, arXiv:1009.5055.

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

[32] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *Proc. IEEE Eur. Conf. Comput. Vis.*, 2014, pp. 254–265.

[33] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. Torr, "Staple: Complementary learners for real-time tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1401–1409.

[34] H. K. Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 21–26.

[35] W. Wang, J. Shen, R. Yang, and F. Porikli, "Saliency-aware video object segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 20–33, Jan. 2018.

[36] J. Shen, J. Peng, and L. Shao, "Submodular trajectories for better motion segmentation in videos," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2688–2700, Jun. 2018.