Robust Visual Tracking via Dirac-Weighted Cascading Correlation Filters

Cheng Peng, Fanghui Liu, Jie Yang, and Nikola Kasabov, *Fellow, IEEE*

***Abstract*—Correlation filter-based trackers (CFTs) have recently raised considerable attention in visual tracking and achieved competitive performance. Nevertheless, conventional structures of such CFTs build a shallow architecture with a single correlation filter, which cannot comprehensively depict the target appearance. Hence, these trackers lack strong discriminative ability and easily drift when the target suffers drastic appearance variations. To address the limitations, we propose Dirac-weighted cascading correlation filters (DWCCF) for visual tracking. It incorporates cascading characteristics of multiple filters to construct the target appearance model, and dynamically learns Dirac weights for each filter, which is accordingly robust to appearance variations. Besides, we design a boundary penalization strategy to adaptively reduce the boundary effects, which efficiently improves the detection precision for tracking. Qualitative and quantitative evaluations on OTB-2013 and OTB-2015 datasets demonstrate that the proposed DWCCF significantly outperforms other state-of-the-art methods.**

***Index Terms*—Correlation filter, cascading structure, Dirac weight, visual tracking.**

I. Introduction

VISUAL tracking is one of the most fundamental research topics in computer vision [1], [2]. It aims to estimate the locations of a target from a video sequence, only the initial position of which is given in the first frame. Although significant progress has been made in recent years, it remains a challenging problem due to several factors including occlusions (OCC), deformation (DEF) and background clutters (BC).

In general, existing tracking approaches are either generative or discriminative. Generative trackers typically regard the tracking task as finding the best image candidate with the minimal reconstruction error [3]–[10]. Comparably, discriminative tracking approaches often train a classifier to distinguish the target from the background [11]–[22]. Recent years have witnessed the rapid advance and competitive performance of Correlation Filter-based Trackers (CFTs) in visual tracking. They aim to train a correlation filter, and then efficiently detect the target location via the yielded response map by utilizing the Discrete Fourier Transform (DFT).

Since Bolme *et al.* [11] firstly introduce correlation filters to the visual tracking community, notable efforts have been devoted to its improvement, e.g., by utilizing spatial or temporal regularization [14], [16], [22], exploiting multi-dimensional features [15], [20], and interpreting the correlation filter as a differentiable layer in a deep neural network [21]. Although appealing results have been achieved, however, they usually apply a single correlation filter. As a result, the learned model is too shallow to comprehensively depict the target appearance, and is not robust to drastic appearance variations. Besides, conventional CFTs simply locate the target by searching for the position of the maximal response score. When the color and texture of background regions are similar to the target, multiple peaks existing in the response map may lead to false detection with a high probability.

Motivated by the above observations, in this letter, we propose a novel architecture termed as Dirac Weighted Cascading Correlation Filters (DWCCF). We exploit the cascading structure of multiple correlation filters and introduce Dirac parameterization [23] to visual tracking. The proposed DWCCF is both deep and broad to guarantee the implicit information to be shared among all filters, and accordingly enhances the model representation ability. Besides, a boundary penalization strategy is designed to improve the precision of the detection procedure, which adaptively reduces the boundary effects caused by BC and noises. We also develop a multi-scale estimation scheme with average peak-to-correlation energy (APCE) [24] as the confidence criterion to tackle scale variations (SV). Qualitative and quantitative evaluations are conducted on OTB-2013 [25] and OTB-2015 [26] benchmarks. The experimental results demonstrate that the proposed approach significantly outperforms other state-of-the-art trackers.

II. Related Work

In this section, we briefly review the training and detection procedures of conventional CFTs, which are closely related to the proposed approach.

In the t-th frame, given a $M \times N$ image sample $\mathbf{X}_t \in \mathbb{R}^{M \times N}$, CFTs typically train a correlation filter based on D-channel feature maps $\{\psi^d(\mathbf{X}_t)\}_{d=1}^{D} \in \mathbb{R}^{M \times N \times D}$, where ψ denotes the feature extraction operation and d refers to the slice index. The correlation filter $\{\mathbf{H}_t^d\}_{d=1}^{D}$ can be trained by minimizing the following objective with a desired Gaussian-shaped response

Manuscript received August 7, 2018; revised September 19, 2018; accepted September 19, 2018. Date of publication September 24, 2018; date of current version October 2, 2018. This work was supported in part by the National Natural Science Foundation of China under Grants 61572315 and 6151101179 and in part by the 973 Plan of China under Grant 2015CB856004. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Joao Paulo Papa. *(Corresponding author: Jie Yang.)*

C. Peng, F. Liu, and J. Yang are with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: pc1899@outlook.com; lfhsgre@outlook.com; jieyang@sjtu.edu.cn).

N. Kasabov is with Knowledge Engineering and Discovery Research Institute, Auckland University of Technology, Auckland 1010, New Zealand (e-mail: nkasabov@aut.ac.nz).

This letter has supplementary downloadable material available at http://ieeexplore.ieee.org.

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2018.2871883

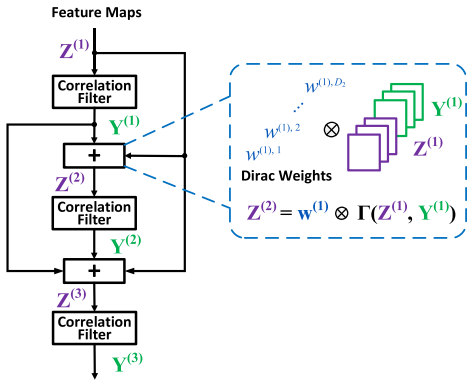

Fig. 1. The proposed cascading structure of multiple filters with Dirac weights. The function Γ denotes the concatenation operation, and $\otimes$ refers to the element-wise product between each channel of concatenated results with the Dirac weights.

map $\bar{\mathbf{Y}}$,

$$\arg\min_{\mathbf{H}_t} \left\| \sum_{d=1}^{D} \mathbf{H}_t^d \star \psi^d\left(\mathbf{X}_t\right) - \bar{\mathbf{Y}} \right\|^2 + \lambda \sum_{d=1}^{D} \left\| \mathbf{H}_t^d \right\|^2,$$

where λ is a regularization parameter, and $\star$ stands for the correlation operation. The closed-form solution is

$$\mathbf{H}_t^d = \mathcal{F}^{-1}\left(\frac{\mathcal{F}\left(\bar{\mathbf{Y}}\right)^* \odot \mathcal{F}\left(\psi^d\left(\mathbf{X}_t\right)\right)}{\sum_{d=1}^{D} \mathcal{F}\left(\psi^d\left(\mathbf{X}_t\right)\right)^* \odot \mathcal{F}\left(\psi^d\left(\mathbf{X}_t\right)\right) + \lambda} \right), \quad d \in \{1, \ldots, D\}.$$

Herein, $\mathcal{F}$ and $\mathcal{F}^{-1}$ refer to the DFT and its inverse operation, respectively. Besides, the superscript * denotes the complex conjugation, and $\odot$ represents the Hadamard product.

In the next frame, the target can be located by searching for the maximal value in the translation response map. The response map $\mathbf{Y}_{t+1}^{trans}$ can be obtained based on a new image sample $\mathbf{X}_{t+1}$ as

$$\begin{aligned} \mathbf{Y}_{t+1}^{trans} &= \sum_{d=1}^{D} \mathbf{H}_t^d \star \psi^d\left(\mathbf{X}_{t+1}\right) \\ &= \mathcal{F}^{-1}\left(\sum_{d=1}^{D} \mathcal{F}\left(\mathbf{H}_t^d\right) \odot \mathcal{F}\left(\psi^d\left(\mathbf{X}_{t+1}\right)\right) \right). \end{aligned} \quad (1)$$

III. Proposed DWCCF Tracker

In the following section, we firstly detail the cascading architecture with Dirac weights of multiple filters, and then introduce the tracking framework of the proposed approach.

A. Dirac Weighted Cascading Correlation Filters Architecture

1) Cascading Structure With Dirac Weights: In this work, we exploit the cascading structure of multiple correlation filters, and introduce Dirac parameterization [23] into the visual tracking community to propose a novel architecture namely Dirac Weighted Cascading Correlation Filters (DWCCF).

Fig. 1 illustrates the proposed Dirac weighted structure with three correlation filters. We denote the input of the i-th filter as $\mathbf{Z}^{(i)} \in \mathbb{R}^{M \times N \times D_i}, i \in \{1, 2, 3\}$. Specifically, $\mathbf{Z}^{(1)} = \psi\left(\mathbf{X}\right)$ and the channel numbers in this letter are $D_1 = 64$, $D_2 = 128$, $D_3 = 192$, respectively. By utilizing the Dirac parameterization, we calculate $\mathbf{Z}^{(i)}$ as

$$\mathbf{Z}^{(i+1)} = \mathbf{w}^{(i)} \otimes \Gamma\left(\mathbf{Z}^{(1)}, \mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(i)}\right), i \in \{1, 2\}, \quad (2)$$

where the function Γ refers to concatenation, and $\otimes$ denotes the element-wise product between the weights and each channel of concatenated results. Besides, $\mathbf{w}^{(i)} \in \mathbb{R}^{D_{i+1}}$ is a vector of Dirac weights learned during training and updating. It contains D_{i+1} elements $\{w^{(i),c}\}_{c=1}^{D_{i+1}}$, corresponding to each channel of $\mathbf{Z}^{(i+1)}$, respectively.

2) Correlation Filters With Extended Layers: In the proposed DWCCF, we generalize the conventional correlation filter and extend its dimension. To be specific, the i-th filter $\mathbf{H}^{(i)}$ owns L_i layers, each of which contains D_i channels $\{\{\mathbf{H}^{(i),d}[l]\}_{d=1}^{D_i}\}_{l=1}^{L_i}$. Herein, we denote $\mathbf{H}[l]$ as the l-th filter layer, and the numbers of filter layers are set as $L_1 = L_2 = L_3 = 64$ in this letter. By doing so, the implicit information of interim response maps can be comprehensively exploited by all filters. The L_i-channel interim response maps $\{\mathbf{Y}^{(i),l}\}_{l=1}^{L_i} \in \mathbb{R}^{M \times N \times L_i}$ as the output of $\mathbf{H}^{(i)}$ are then yielded as

$$\mathbf{Y}^{(i),l} = \sum_{d=1}^{D_i} \mathbf{H}^{(i),d}[l] \star \mathbf{Z}^{(i),d}, \quad i \in \{1, 2, 3\}, l \in \{1, \ldots, L_i\}. \quad (3)$$

Besides, we slide the filter $\mathbf{H}^{(i)}$ in the spatial domain to obtain the correlation results, i.e.,

$$\mathbf{Y}^{(i),l}\Big|_{\mathbf{p}} = \sum_{d=1}^{D_i} \mathbf{H}^{(i),d}[l] \odot \Upsilon\left(\mathbf{Z}^{(i),d}; \mathbf{p}\right), \quad i \in \{1, 2, 3\}, l \in \{1, \ldots, L_i\}, \mathbf{p} \in \mathcal{P}, \quad (4)$$

where $\mathbf{Y}^{(i),l}|_{\mathbf{p}}$ denotes the response result of $\mathbf{Y}^{(i),l}$ at the position $\mathbf{p}$, and $\Upsilon(\mathbf{Z}; \mathbf{p})$ refers to the circular shift result of $\mathbf{Z}$ located at $\mathbf{p}$ in the position space $\mathcal{P}$.

3) Connection to Traditional CFTs: As formulated in Section II, conventional CFTs aim to train a single correlation filter, the structure of which is both shallow and narrow. Accordingly, the learned filter lacks strong model representation ability when the target suffers heavy OCC, and the tracker easily drifts.

Comparably, our architecture contains multiple correlation filters, and dynamically learns Dirac weights for each filter. The conventional CFT is a special case of DWCCF. When the learned Dirac weights $w^{(2),c} = 0, c \in \{D_1 + 1, D_1 + 2, \ldots, D_3\}$ and the extended filter layer $L_3 = 1$, DWCCF is degraded to the conventional CFT with a single filter. The translation response map is then yielded as

$$\mathbf{Y}^{trans} = \mathbf{Y}^{(3),1} = \sum_{d=1}^{D_1} \mathbf{H}^{(3),d}[1] \star \psi^d\left(\mathbf{X}\right),$$

which is equivalent to (1). In most cases, the proposed DWCCF can significantly exploit the implicit information carried out by all filters, which contributes to robustness against drastic appearance variations.

B. Tracking Framework

In this section, we investigate the training, updating and detection scheme of the proposed DWCCF.

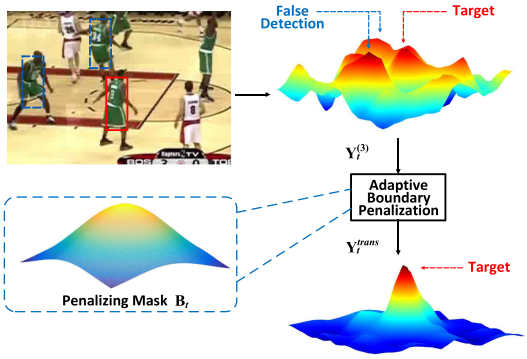

Fig. 2. Illustration of adaptive boundary penalization scheme. Due to background clutters in sequence *basketball* from OTB-2013, there exist multiple peaks in the response map. The produced peak of the target is inferior to those of the background regions, which leads to inaccurate detection and tracking drifts. The proposed boundary penalization strategy will adaptively reduce the boundary effects and correctly re-detect the target position.

1) Training and Updating Scheme: The proposed DWCCF is capable of learning a data-specific model without off-line training on large-scale datasets. In the first frame, based on the extracted feature maps $\psi(\mathbf{X}_1)$ as the input $\mathbf{Z}_1^{(1)}$, we can obtain the corresponding output $\mathbf{Y}_1^{(3)}$ via the forward pass of DWCCF with (2), (3) and (4). The Adam optimiser [27] is utilized to train the model by backwards propagating and minimizing the ℓ_2-loss errors between the yielded response map and the desired output $\bar{\mathbf{Y}}$, i.e.,

$$\min_{\mathbf{H}_1,\mathbf{w}_1}\left\|\sum_{l=1}^{L_3}\mathbf{Y}_1^{(3),l}-\bar{\mathbf{Y}}\right\|^2+\lambda\sum_{i=1}^{3}\sum_{l=1}^{L_i}\sum_{d=1}^{D_i}\left\|\mathbf{H}_1^{(i),d}[l]\right\|^2+\lambda\sum_{i=1}^{2}\left\|\mathbf{w}_1^{(i)}\right\|^2.$$

In order to tackle appearance variations, we also establish an online updating strategy after detection in each frame. Based on the model parameters generated in the last frame, the DWCCF structure is updated in the similar way of training with lower learning rates. The updating procedure is terminated in finite iterations, which avoids over-fitting and achieves a trade-off between adaptivity and stability.

2) Translation Detection via Adaptive Boundary Penalization: As illustrated in Fig. 2, we propose a novel adaptive boundary penalization scheme to reduce the boundary effects during the detection procedure. Specifically, a penalizing mask $\mathbf{B}_t \in \mathbb{R}^{M\times N}$ in the t-th is adopted to yield the translation response map $\mathbf{Y}_t^{trans}$ as

$$\mathbf{Y}_t^{trans}=\mathbf{B}_t\odot\sum_{l=1}^{L_3}\mathbf{Y}_t^{(3),l}.$$

The penalizing mask can be generated as

$$\mathbf{B}_t|_{\mathbf{p}}=\exp\left(-\left\|\frac{\mathbf{p}-\mathbf{p}_{t-1}}{\boldsymbol{\sigma}_t}\right\|^2\right),\ \mathbf{p}\in\mathcal{P},$$

where $\mathbf{p}_{t-1}$ refers to the target position detected in the last frame, and $\boldsymbol{\sigma}_t$ is a variance vector associated with current decay rate of the APCE value. The APCE value of a response map $\mathbf{Y}\in\mathbb{R}^{M\times N}$ is defined as

$$APCE(\mathbf{Y})=\frac{|\max(\mathbf{Y})-\min(\mathbf{Y})|^2}{\|\mathbf{Y}-\min(\mathbf{Y})\|^2/(MN)}.$$

We denote the APCE value of the response map $\mathbf{Y}_t^{(3)}$ in the t-th frame as ξ_t, i.e., $\xi_t=APCE(\sum_{l=1}^{L_3}\mathbf{Y}_t^{(3),l})$. A low APCE value ξ_t usually implies the image $\mathbf{X}_t$ suffers heavy noises. To tackle this problem, we adaptively calculate the variance vector $\boldsymbol{\sigma}_t$ of the distribution via a sigmoid function as

$$\boldsymbol{\sigma}_t=\frac{\boldsymbol{\kappa}}{1+\exp\left(-\tau(\frac{\xi_t}{\xi_1}-\alpha)\right)}+\boldsymbol{\beta},$$

with two vectorial constants $\boldsymbol{\kappa}$ and $\boldsymbol{\beta}$ relevant to the target size. By doing so, the response scores residing in the background regions are penalized by assigning lower weights according to the APCE value decay, which significantly improves localization precision. The target position $\mathbf{p}_t$ in the t-th frame is then estimated to locate at the maximal score in $\mathbf{Y}_t^{trans}$.

3) Multiple Scale Estimation: Following [28], a multi-scale estimation scheme is developed to account for SV problems. We search the scale space $\mathcal{S}$ in parallel with different scales, and introduce APCE as the confidence criterion of the scale results. The target scale s_t in the t-th frame is estimated from scale results with the maximal APCE value.

IV. Experiments and Evaluations

We provide qualitative and quantitative comparisons to evaluate the proposed DWCCF. Experiments are conducted OTB-2013 [25] and OTB-2015 [26] datasets compared with other state-of-the-art tracking approaches in recent years.

A. Details and Parameters

The proposed DWCCF is implemented on MATLAB with an Intel Xeon E5-2695 CPU and a GTX TITAN X GPU. Our implementation runs at the speed of 6.1 frames per second on average. We used fixed parameters for all sequences as follows. The $M\times N$ search window was 4 times of the initial target size. The feature extraction was same to that of [12], and we generated the outputs of *conv4-4*, *conv3-4*, *conv5-4* layers of VGG-Net [29] as the feature maps. The spatial size of each filter was set as $(\lfloor\frac{M}{4}\rfloor-\lfloor\frac{M}{4}\rfloor \bmod 2+1)\times(\lfloor\frac{N}{4}\rfloor-\lfloor\frac{N}{4}\rfloor \bmod 2+1)$. The regularization parameter $\lambda=0.001$. In Section III-B2, the constant vectors $\boldsymbol{\kappa}$ and $\boldsymbol{\beta}$ were set as $\frac{1}{10}(M,N)$ and $\frac{1}{4}(M,N)$, respectively, and $\alpha=0.5$, $\tau=6$. The scale space was set as $\mathcal{S}=\{0.995,1,1.005\}$.

B. Benchmarks and Evaluation Metrics

We evaluate the proposed method on OTB-2013 [25] and OTB-2015 [26] datasets which contain 51 and 100 sequences, respectively. All the sequences are annotated with eleven attributes, namely: BC, DEF, fast motion (FM), in-plane rotation (IPR), illumination variation (IV), low resolution (LR), motion blur (MB), OCC, out-of-plane rotation (OPR), out-of-view (OV) and SV.

Two standard evaluation metrics are used to measure the tracking performance. The *precision* of a tracker is associated with the center location error, which is defined as the average Euclidean distance between the center locations of the tracked result r_T and the ground truth r_G. Tracking methods are compared in precision for the distance = 20 pixels. The *success rate* of a tracker refers to the ratio of successfully tracked frames to the total frames. The successfully tracked frames are those the intersection-over-union overlaps of which (defined as $\frac{|r_T\cap r_G|}{|r_T\cup r_G|}$)

TABLE I
RANKED AUC SCORES (%) UNDER ELEVEN ATTRIBUTES ON OTB-2015

Tracker \ Attribute	FM (39)	BC (31)	MB (29)	DEF (44)	IV (38)	IPR (51)	LR (9)	OCC (49)	OPR (63)	OV (14)	SV (64)
DWCCF (Ours)	63.5	62.3	65.6	58.0	65.4	62.0	57.1	60.0	61.5	57.3	58.2
CFNet (2017)	58.7	57.0	57.5	50.1	59.2	58.6	61.4	54.5	56.3	51.1	56.7
LMCF (2017)	54.7	60.0	55.5	52.8	60.2	53.9	45.0	55.2	55.7	52.5	52.2
fDSST (2017)	54.7	58.0	53.8	46.7	56.8	54.6	45.8	47.9	50.1	44.4	50.1
ACFN (2017)	56.1	53.6	56.1	53.5	56.7	54.3	42.5	53.8	54.3	49.4	54.7
CSR-DCF (2017)	57.9	56.3	57.8	56.4	59.4	54.4	41.7	53.2	54.7	49.9	52.8
SRDCFdecon (2016)	60.1	63.5	63.3	55.3	64.6	56.9	49.2	58.5	59.1	49.6	60.4
SiamFC (2016)	56.8	52.3	55.0	50.6	56.8	55.7	59.2	54.3	55.8	50.6	55.2
DLSSVM (2016)	53.8	54.7	57.6	51.3	56.1	55.6	43.3	53.3	54.7	46.7	48.8
Staple (2016)	53.4	56.1	53.9	55.0	59.4	54.9	39.9	54.3	53.1	46.8	52.2
SRDCF (2015)	59.3	57.8	58.9	54.4	61.3	54.1	49.4	55.6	55.0	45.0	55.9
HCFT (2015)	56.7	58.0	58.0	53.0	54.0	55.7	43.9	52.3	53.4	46.4	48.3
KCF (2015)	45.9	49.8	45.9	43.6	47.9	46.9	30.7	44.3	45.3	39.3	39.4

Sequence numbers associated with corresponding attributes are shown in parenthesis. The best, second and third performance are indicated by colors.

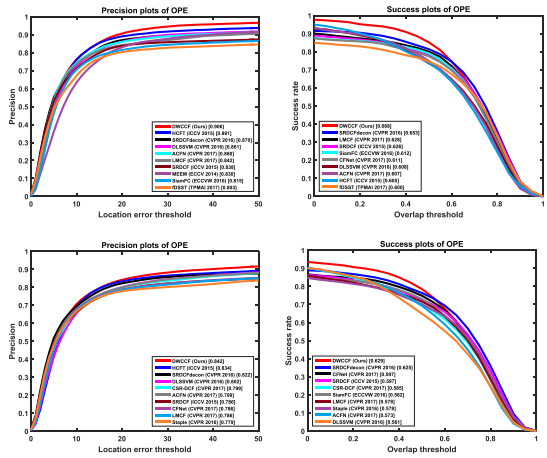

Fig. 3. Comparisons of the proposed DWCCF with other state-of-the-art approaches (first row on OTB-2013 and second row on OTB-2015). Only the top ten trackers are listed for clarity.

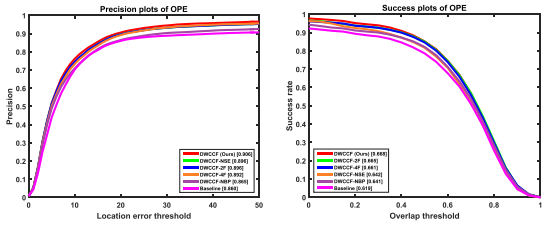

Fig. 4. Evaluation of key components from proposed DWCCF on OTB-2013.

are larger than a threshold. Trackers are ranked by the Area Under Curve (AUC) scores of the success plots.

Besides, we evaluate our tracker on the VOT-2016 benchmark [30]. Due to the space limitation, more details are provided in the supplementary material.

C. Qualitative and Quantitative Evaluations

1) Overall Comparisons: We provide evaluations of DWCCF compared with 21 state-of-the-art tracking methods, including fDSST [20], CSR-DCF [31], CFNet [21], ACFN [32], LMCF [24], SRDCFdecon [16], Staple [15], Struck [33], DLSSVM [34], SiamFC [35], SRDCF [14], HCFT [12], KCF [13], SAMF [28], MEEM [36], CN [37] and TLD [38].

Fig. 3 illustrates the precision and success plots under One Pass Evaluation (OPE) on OTB-2013 [25] and OTB-2015 [26]. Our DWCCF provides the best tracking performance with mean precisions of 90.6% and 84.2%, and mean success rates of 66.8% and 62.9%, respectively. The overall results demonstrate that the proposed approach significantly outperforms conventional CFTs including fDSST, CFNet, SRDCFdecon, Staple and HCFT.

2) Attribute-Based Evaluation: The attribute-based evaluation under eleven attributes on OTB-2015 is shown in Table I. The experimental results demonstrate that the proposed DWCCF outperforms other trackers under most challenging situations. DWCCF is especially robust against DEF, BC and OCC compared with state-of-the-art methods. It is because that the proposed structure efficiently exploits the implicit information of features and interim response maps from various filters, and the boundary penalization strategy adaptively reduces the boundary effects caused by noises, which enhances the discriminative ability and thus improves the tracking performance.

3) Ablation Analysis: An ablation analysis on OTB-2013 is provide to validate the effectiveness of the key components of the proposed DWCCF. The comparisons among the following variants are shown in Fig. 4: (i) a simple cascade of three filters as the baseline; (ii) Dirac weighted cascading structure with 2 correlation filters, denoted as DWCCF-2F; (iii) Dirac weighted cascading structure with 4 filters, denoted as DWCCF-4F; (iv) the proposed DWCCF tracker; (v) DWCCF without adaptive boundary penalization strategy, denoted as DWCCF-NBP; (vi) DWCCF without multi-scale estimation scheme, denoted as DWCCF-NSE.

The results demonstrate that the structure with fewer filters is too shallow and narrow to ensure a strong discriminative ability, while more filters are sensitive to target appearance variations as a result of over-fitting problems, and also make the training, detection and updating procedures inefficient. The comparisons show that the components of the proposed DWCCF are significantly conducive to the improvement of tracking performance.

V. CONCLUSION

In this letter, we propose a novel cascading architecture of multiple correlation filters with Dirac weights for robust visual tracking. The proposed DWCCF comprehensively exploits the implicit information carried out by all filters, and significantly enhances the discriminative ability of the learned model. Besides, the introduced adaptive boundary penalization scheme in our framework efficiently helps to reduce noise effects, and accordingly prevents the tracker from drifting. Evaluations on OTB-2013 and OTB-2015 datasets demonstrate the effectiveness and robustness of the proposed DWCCF tracker compared with existing state-of-the-art methods.

REFERENCES

[1] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1442–1468, Jul. 2014.

[2] P. Liang, Y. Pang, C. Liao, X. Mei, and H. Ling, "Adaptive objectness for object tracking," *IEEE Signal Process. Lett.*, vol. 23, no. 7, pp. 949–953, Jul. 2016.

[3] X. Mei and H. Ling, "Robust visual tracking using ℓ_1 minimization," in *Proc. IEEE 12th Int. Conf. Comput. Vision*, 2009, pp. 1436–1443.

[4] T. Zhou, X. He, K. Xie, K. Fu, J. Zhang, and J. Yang, "Robust visual tracking via efficient manifold ranking with low-dimensional compressive features," *Pattern Recognit.*, vol. 48, no. 8, pp. 2459–473, 2015.

[5] K. Zhang, Q. Liu, Y. Wu, and M.-H. Yang, "Robust visual tracking via convolutional networks without training," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1779–1792, Apr. 2016.

[6] F. Liu, C. Gong, T. Zhou, K. Fu, X. He, and J. Yang, "Visual tracking via nonnegative multiple coding," *IEEE Trans. Multimedia*, vol. 19, no. 12, pp. 2680–2691, Dec. 2017.

[7] T. Zhou, H. Bhaskar, F. Liu, and J. Yang, "Graph regularized and locality-constrained coding for robust visual tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 10, pp. 2153–2164, Oct. 2017.

[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] F. Liu, T. Zhou, C. Gong, K. Fu, L. Bai, and J. Yang, "Inverse nonnegative local coordinate factorization for visual tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 8, pp. 1752–1764, Aug. 2018.

[10] K. Zhang, Q. Liu, J. Yang, and M.-H. Yang, "Visual tracking via boolean map representations," *Pattern Recognit.*, vol. 81, pp. 147–160, 2018.

[11] D. S. 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. Vision Pattern Recognit.*, 2010, pp. 2544–2550.

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

[13] 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.

[14] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vision*, 2015, pp. 4310–4318.

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

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

[17] 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.

[18] J. Yang, K. Zhang, and Q. Liu, "Robust object tracking by online Fisher discrimination boosting feature selection," *Comput. Vision Image Understanding*, vol. 153, pp. 100–108, 2016.

[19] C. Peng, F. Liu, H. Yang, J. Yang, and N. Kasabov, "Correlation filters with adaptive memories and fusion for visual tracking," in *Proc. Int. Conf. Neural Inf. Process.*, 2017, pp. 170–179.

[20] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Discriminative scale space tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 8, pp. 1561–1575, Aug. 2017.

[21] J. Valmadre, L. Bertinetto, J. Henriques, A. Vedaldi, and P. H. S. Torr, "End-to-end representation learning for correlation filter based tracking," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 5000–5008.

[22] K. Zhang, X. Li, H. Song, Q. Liu, and W. Lian, "Visual tracking using spatio-temporally nonlocally regularized correlation filter," *Pattern Recognit.*, vol. 83, pp. 185–195, 2018.

[23] S. Zagoruyko and N. Komodakis, "Diracnets: Training very deep neural networks without skip-connections," 2017, arXiv:1706.00388.

[24] M. Wang, Y. Liu, and Z. Huang, "Large margin object tracking with circulant feature maps," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 4800–4808.

[25] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2013, pp. 2411–2418.

[26] 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.

[27] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, arXiv:1412.6980.

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

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, arXiv:1409.1556.

[30] M. Kristan *et al.*, "*The visual object tracking VOT2016 challenge results*," in *Proc. IEEE Eur. Conf. Comput. Vision*, 2016, pp. 777–823.

[31] A. Lukežič, T. Vojíř, L. C. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 4847–4856.

[32] J. Choi, H. J. Chang, S. Yun, T. Fischer, Y. Demiris, and J. Y. Choi, "Attentional correlation filter network for adaptive visual tracking," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 4828–4837.

[33] S. Hare, A. Saffari, and P. H. S. Torr, "Struck: Structured output tracking with kernels," in *Proc. IEEE Int. Conf. Comput. Vision*, 2011, pp. 263–270.

[34] J. Ning, J. Yang, S. Jiang, L. Zhang, and M.-H. Yang, "Object tracking via dual linear structured SVM and explicit feature map," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 4266–4274.

[35] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, "Fully-convolutional siamese networks for object tracking," in *Proc. Eur. Conf. Comput. Vision Workshops*, 2016, pp. 850–865.

[36] J. Zhang, S. Ma, and S. Sclaroff, "Meem: Robust tracking via multiple experts using entropy minimization," in *Proc. Eur. Conf. Comput. Vision*, 2014, pp. 188–203.

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

[38] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1409–1422, Jul. 2012.