ROBUST OBJECT TRACKING VIA PART-BASED CORRELATION PARTICLE FILTER

Ning Wang, Wengang Zhou, Houqiang Li

CAS Key Laboratory of Technology in Geo-spatial Information Processing and Application System,
EEIS Department, University of Science and Technology of China
wn6149@mail.ustc.edu.cn {zhwg, lihq}@ustc.edu.cn

ABSTRACT

In this paper, a part-based correlation particle filter framework is proposed for robust visual tracking. Through managing target parts by correlation filters in a particle filter framework, we comprehensively model the target appearance using plentiful overlapped local parts with different positions and sizes. Further, we propose a particle re-sampling mechanism with appearance and geometry reliability consideration to resample the redundant particles, which guides our tracker to focus more on the discriminative and reliable local parts. Finally, to cope with the limited search range of local tracker and model corruption caused by unreliable samples, we introduce the top-down coarse-to-fine localization and bottom-up adaptive update strategies to further boost the performance. Extensive experimental results on three challenging datasets demonstrate that our tracking algorithm performs favorably against state-of-the-art methods. Specifically, our approach exhibits superior performance on tracking nonrigid objects with rotation and large deformation.

Index Terms— Visual tracking, correlation filter, particle filter, part-based model

1. INTRODUCTION

The goal of single object tracking is to predict the target state (*e.g.,* position and scale) in subsequent video frames with only the initiate target state. Visual tracking plays a basic and important role in many visual tasks, such as video surveillance, robotics and autonomous driving. Although significant progress [1, 2, 3] has been made, challenges such as occlusion, deformation, in-plane and out-of-plane rotation make visual tracking still a difficult task.

Considering the target representation strategy, the tracking model can be divided into two categories: global and part-based. Generally, global tracking model captures the holistic target information while part-based model mainly focuses on local appearance. As one of the most popular tracking algorithms, Discriminative Correlation Filter (DCF) [8, 2, 9] has been widely investigated due to its attractive performance and efficiency. Global DCF methods have been further developed by studying scale estimation [10, 11], multi-feature fusion [11, 12, 13, 14] and avoiding boundary effects [15]. Recently, Zhang *et al.* propose a multi-task correlation particle filter (MCPF) [16] framework to absorb the advantages of both correlation filter and particle filter. In MCPF, the particle filter aims to conduct better target scale estimation and cover multiple modes in the posterior density. On the other hand, correlation filter helps shepherd the particles, and thus less particles are required compared to traditional particle filter [17]. However, the MCPF and other algorithms that adopt global DCF model are prone to drift in case of occlusion and deformation.

In contrast, by dividing the target into several local parts and tracking each of them via an independent tracker, part-based DCF methods [5, 18, 4, 6, 7] demonstrate superior robustness in handling partial occlusion and deformation. The reliable patch tracker [18] considers motion trajectories of both foreground and background local parts in sequential

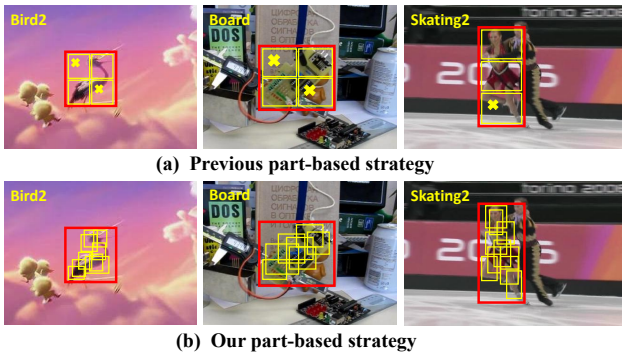


Fig. 1. In (a), previous part-based strategies [4, 5, 6, 7] usually adopt artificial and intuitive partition methods. Thus, some local parts with little target information are not reliable enough (labeled with "×"). In (b), our part-based strategy captures the discriminative and stable local parts.

This work was supported in part to Dr. Houqiang Li by 973 Pro-gram under contract No. 2015CB351803 and NSFC under contract No. 61390514, and in part to Dr. Wengang Zhou by NSFC under contract No. 61472378 and No. 61632019, the Fundamental Research Funds for the Central Universities, and Young Elite Scientists Sponsorship Program By CAST (2016QNRC001).

978-1-5386-1737-3/18/$31.00

Monte Carlo framework. In [7], the geometrically constrained constellation of local DCFs is proposed. In SCF algorithm [4], Liu *et al.* propose a structural DCF based model to exploit the relationship of local target parts. Fan *et al.* [6] further explore the temporal consistence and incorporate the global target information. Although promising performance has achieved, one major limitation is that they usually adopt artificial target partition strategy by simply dividing the object into several parts according to the target size ratio (*e*.g., 2×2 or 3×1 parts in Fig. 1 (a)). However, in real-world tracking tasks, the targets (especially nonrigid objects) are quite difficult to be tightly represented by a rectangle bounding box. Therefore, their simple and intuitive object partition strategy may include useless or corrupted local parts, which may cause the gradually drift problem.

Motivated by the above observations, we propose a novel Part-based Correlation Particle Filter (PCPF) algorithm. Our method samples multiple local parts through particle filter framework and tracks them independently via correlation filters. After estimating the appearance and geometric reliability of the multiple parts (particles), our tracker gradually discards and resamples the unstable particles. Therefore, our tracker is able to focus on the discriminative and stable local parts. Further, through coarse-to-fine localization and bottom-up adaptive update, our tracker achieves more robust performance.

Compared with MCPF algorithm [16], our method maintains its original metrics and demonstrates superior performance due to the following reasons. (1) Part-based target representation is less sensitive to partial occlusion and deformation. (2) MCPF resamples all the particles in each frame, while our method gradually resamples the particles depending on their appearance and geometry reliability, which preserves the object structure. (3) Different local parts are updated asynchronously, which alleviates the model corruption caused by inadequate update.

Furthermore, compared with previous part-based DCF trackers, our method has the following advantages. (1) Without manually designed object partition strategy [5, 4, 6, 7], our method models target appearance comprehensively through randomly sampling multiple overlapped local parts in a particle filter framework, which is more capable in tracking nonrigid object with heavy deformation. (2) The proposed resample mechanism drives the particles to pay more attention to trackable and discriminative local target parts. (3) Our tracker combines the global and part-based models in a coarse-to-fine fashion to cope with the limited search range of local trackers, which is mostly ignored [5, 18, 6]. Moreover, the occlusion estimation of the overall target is feasible after measuring the reliability of plentiful local parts, which in return guides the adaptive update of global model.

In brief, the main contributions of our work can be summarized as follows:

- We propose a Part-based Correlation Particle Filter (PCPF) framework for visual tracking, which captures the reliable and discriminative local parts.
- The top-down coarse-to-fine localization and bottom-up adaptive update strategies are introduced to further boost the tracking robustness.
- Extensive experiments on OTB-2013 [19], OTB-2015 [20] and Temple-Color [21] benchmarks demonstrate the state-of-the-art performance of our tracker.

2. METHOD

In this section, we introduce the part-based correlation particle filter in Section 2.1. How to estimate the reliability of the local parts and resample them is elaborated in Section 2.2. Finally, Section 2.3 describes our adaptive update strategy.

2.1. Part-Based Correlation Particle Filter

Correlation Filter. A typical tracker based on DCF [8, 2] is trained using an image patch $\mathbf{x}$ of size $M \times N$, which is centered around the target. All the circular shifts of the patch $\mathbf{x}(m, n) \in \{0, 1, ...M-1\} \times \{0, 1, ...N-1\}$ are generated as training samples with Gaussian function label $y(m, n)$. The filter $\mathbf{w}$ can be trained by minimizing the following regression error:

$$\min_{\mathbf{w}} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{w}\|_2^2, \tag{1}$$

where λ is a regularization parameter ($\lambda \geq 0$) and $\mathbf{X}$ is the data matrix by concatenating all the circular shifts. The close-form solution of Eq. (1) is defined by $\mathbf{w} = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$ [2]. Since $\mathbf{X}$ is circulant, the filter solution on the d-th ($d \in \{1, \cdots, D\}$) feature channel can be efficiently calculated as follows.

$$\hat{\mathbf{w}}_d^* = \frac{\hat{\mathbf{y}} \odot \hat{\mathbf{x}}_d^*}{\sum_{i=1}^{D} \hat{\mathbf{x}}_i^* \odot \hat{\mathbf{x}}_i + \lambda}, \tag{2}$$

where $\odot$ is the element-wise product, the hat symbol " $\hat{\bullet}$ " denotes the Discrete Fourier Transform (DFT) of a vector and " $\bullet^*$ " is the complex-conjugate operation. To avoid the boundary effects during learning, we apply Hann window to the signals [2].

In the next frame, a Region of Interest (ROI) patch $\mathbf{z}$ with the same size of $\mathbf{x}$ is cropped out for predicting the target position. The response map R of the search patch $\mathbf{z}$ is calculated by Eq (3).

$$R = \mathcal{F}^{-1}(\sum_{d=1}^{D} \hat{\mathbf{w}}_d \odot \hat{\mathbf{z}}_d^*). \tag{3}$$

Particle Filter. In particle filter, given the target observations $\mathbf{z}_{1:t-1} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{t-1}\}$ up to time $t-1$, the posterior density function $p(\mathbf{x}_t|\mathbf{z}_{1:t-1})$ can be calculated as follows.

$$p(\mathbf{x}_t|\mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}) d\mathbf{x}_{t-1}, \tag{4}$$

where $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ is the state prediction and $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})$ is the state density function. When the observation $\mathbf{z}_t$ is available, according to the Bayes rule, the posterior probability $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ can be inferred recursively by Eq (5).

$$p(\mathbf{x}_t|\mathbf{z}_{1:t}) = \frac{p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{z}_{1:t-1})}{p(\mathbf{z}_t|\mathbf{z}_{1:t-1})}, \tag{5}$$

where $p(\mathbf{z}_t|\mathbf{x}_t)$ is the observation likelihood. The posterior $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ is approximated by N particles $\{\mathbf{x}_t^i\}_{i=1}^N$ as follows.

$$p(\mathbf{x}_t|\mathbf{z}_{1:t}) \approx \sum_{i=1}^{N} w_t^i \delta(\mathbf{x}_t - \mathbf{x}_t^i), \tag{6}$$

where w_t^i represents the weight of i-th particle and $\delta(\cdot)$ is the Dirac delta function. Through defining the importance density function as $p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i)$, the particle weight can be calculated by $w_t^i \propto w_{t-1}^i p(\mathbf{z}_t|\mathbf{x}_t^i)$. To avoid the particle degeneracy problem [17], the weights are set to $w_{t-1}^i = 1/N$ for all particles and this leads to $w_t^i \propto p(\mathbf{z}_t|\mathbf{x}_t^i)$

Part-Based Correlation Particle Filter. In our part-based correlation particle filter framework, each particle is formulated by a local DCF tracker. In frame t, the i-th particle position p_t^i is identified by searching for the maximum value of its correlation response map R_t^i. As for the particle weight, different from the methods [16, 6] that use the maximal value of the response map, we define $p(\mathbf{y}_t|\mathbf{s}_t^i)$ as the peak-to-sidelobe ratio (PSR) [8] of the correlation response by Eq (7), which is more robust.

$$PSR(R_t^i) = \frac{max(R_t^i) - \mu(R_t^i)}{\sigma(R_t^i)}, \tag{7}$$

where μ and σ are the mean value and standard deviation of the response map. Then, the weight of the i-th particle in frame t is proportional to the normalized response peak strength, *i.e.,* $w_t^i \propto norm\left(PSR(R_t^i)\right)$.

Similar to Eq. (6), The global target location is estimated using local particle position p_t^i through

$$P_{target} \approx \sum_{i=1}^{N} w_t^i (p_t^i + \Delta_t^i), \tag{8}$$

where N is the total particle number, Δ_t^i denotes deformation vector between i-th part and object center in the previous frame. Besides, the unreliable particles (*e.g.,* green boxes in Fig. 2) do not contribute to this target position estimation, and the reliability measure will be discussed in Section 2.2.

Coarse-to-Fine Localization. In DCF tracker, the ROI is usually 2.5 times the size of the target [2], and a too large ROI will decrease the discriminative power of the filter. Therefore, the small size of local parts will restrict their search range, which makes it hard for them to keep pace with the fast motion of global target.

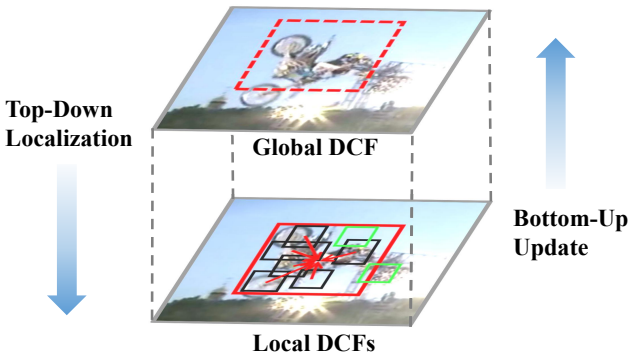


Fig. 2. **Top-down localization:** Global DCF tracker performs well when facing fast motion while local DCF trackers are less sensitive to occlusion and deformation, and the coarse-to-fine combination of them helps refine the tracking result. **Bottom-up update:** After selectively updating the local DCFs, the reliability estimation of plentiful local parts is quantified to adaptively update the global model.

To cope with this issue, we use a global DCF tracker to coarsely predict the target state and obtain the rough target location P'_{target}. Then the coarse position of each local part in frame t is computed by $p'^i_t = P'_{target} - \Delta_t^i$. Through drawing an ROI at position p'^i_t, the accurate particle position is estimated through local DCF tracker: $p'^i_t \rightarrow p_t^i$. Finally, the accurate local locations p_t^i together contribute to the final target position through Eq. (8). As for target scale estimation, we utilize the global DCF model to measure the overall object scale changes [10].

2.2. Particle Resample

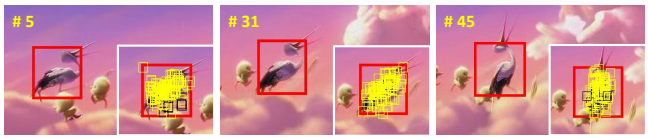
In our framework, each particle represents a DCF tracker with robust tracking capability. Thus, different from previous particle filter method [17], our tracker does not need to resample all the particles in each frame and just resamples the corrupted ones. In the following, we introduce the proposed reliability estimation with geometric and appearance consideration.

Geometry Reliability. Geometric reliability includes motion angle and motion range estimation.

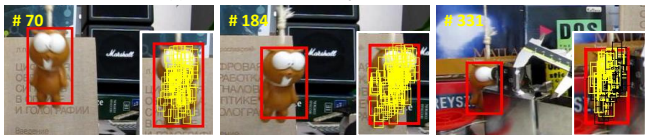
(1) The motion angle of each local part in successive frames can be computed as follows.

$$G_t^i = cos\left(\frac{\Delta_{t-1}^i \cdot \Delta_t^i}{|\Delta_{t-1}^i| \cdot |\Delta_t^i|}\right), \tag{9}$$

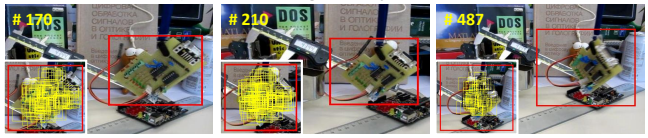
where Δ_t^i represents the deformation vector and $G_t^i \in [-1, 1]$. If $G_t^i < 0$, the corresponding particle moves in the opposite direction between two continuous frames. Even for the target with extreme heavy deformation, the local parts are impossible to change such a large angle. So the particle satisfied with $G_t^i < 0$ is regarded as unreliable.



(a) Video ***Bird2*** with object deformation



(b) Video ***Lemming*** with object occlusion



(c) Video ***Board*** with object rotation

Fig. 3. Through the proposed particle resample strategy by considering geometry and appearance reliability, our tracker gradually resamples the unstable particles (black boxes in figure) and alleviates the model corruption by the background clutters or occlusion objects.

(2) Besides, we consider the motion range of the local parts. If the local part is far away from the target center, we consider the corresponding local part as unreliable, *i*.e., $\|\Delta_t^i\|_2 > \alpha \cdot \frac{1}{2}(W+H), (\alpha > 1)$, where W and H represent the width and height of the global target, and α is the distance reliability threshold.

Appearance Reliability. The PSR value defined in Eq. (7) reflects the correlation between the current frame and previous frames. As discussed in [8], the PSR value will decrease when facing challenging factors such as illumination variation and deformation, which can be quantified to estimate the appearance agreement. If the current PSR value is significantly lower than the previous average PSR, we consider the corresponding particle as unreliable, *i*.e., $PSR(R_t^i) < \beta \cdot Average\left(PSR(R_{1:t-1}^i)\right)$, where $\beta < 1$ is the PSR reliability threshold.

The particle is determined as unreliable when any one of the above conditions (both geometric and appearance estimation) is satisfied. Further, if the particle is continuously regarded as unreliable in a period (5 frames in our experiments), the particle will be resampled and initialized by the DCF using the current frame. As shown in Fig. 3, this particle resample strategy makes our tracker pay more attention to stable and trackable local parts. As a result, our method is more adept than global models at handing challenging factors like heavy deformation, partial occlusion and rotation.

2.3. Bottom-Up Adaptive Update

Following the standard DCF methods [8, 2], the online update of correlation filter is defined as follows,

$$\begin{aligned} \hat{\mathbf{A}}_t &= (1-\eta)\,\hat{\mathbf{A}}_{t-1} + \eta\,\hat{\mathbf{y}} \odot \hat{\mathbf{x}}_t^*, \\ \hat{\mathbf{B}}_t &= (1-\eta)\,\hat{\mathbf{B}}_{t-1} + \eta \sum_{i=1}^{D} \hat{\mathbf{x}}_t^{i*} \odot \hat{\mathbf{x}}_t^i, \end{aligned} \tag{10}$$

where $\hat{\mathbf{A}}_t$ and $\hat{\mathbf{B}}_t$ represent the numerator and denominator of the filter $\hat{\mathbf{w}}$ in Eq. (2), η is the learning rate and t is the index of the current frame. We first estimate the reliability of multiple local parts and selectively update them, which can further guide the adaptive update of the global DCF model (Fig. 2).

If the local part is regarded as unreliable in Section 2.2, the corresponding local DCF does not conduct update to avoid the contamination of its model. After updating local models, we compute the failure ratio among all the particles: $r = N_u/N$, where N_u and N denote the number of unreliable and total particles, respectively. The sudden high failure ratio usually means the target is potentially under full occlusion or out-of-view. Then, the learning rate of global DCF is decreased by a designed Gaussian function. The adaptive learning rate of global and local DCF is given by Eq. (11).

$$\begin{cases} \eta_{local} = 0 \;/\; 1, & \text{for unreliable / reliable parts,} \\ \eta_{global} = exp(-\frac{r^2}{2\sigma^2}). \end{cases} \tag{11}$$

3. EXPERIMENTS

3.1. Experimental Setup

Implementation Details. In our experiments, we follow the parameters in standard DCF method [2] to construct experts. We draw N=100 particles from a Gaussian distribution, whose mean and covariance are $(x, y, 0.2w, 0.2h)$ and $0.2 \cdot (w, h, w, h)$, $(x, y, w, h$ are the positions, width and height of the global target). In the particle initialization stage, particles out of the target area will be re-initialized to ensure all the particles are in the target region. The reliability threshold in resample mechanism α and β are set to 1.2 and 0.7, respectively. The parameter σ in Eq. (6) is set to 0.1.

We use the same setting of parameters for all the experiments. Global DCF tracker is the basic model which provides the coarse target localization. To achieve satisfying performance, we construct global tracker using standard DCF with deep features [12] and color histogram-based features [22]. As for local DCFs, we just use HOG [23] features because small local patches are usually lack of semantic information, which also ensures the efficiency. On a computer with an Intel I7-4790K 4.00GHz CPU and a GeForce GTX 1080Ti GPU, our tracker runs at about 3 frames per second (FPS).

Evaluation Benchmarks and Metrics. Our method is evaluated on three benchmark datasets by a no-reset evaluation protocol: OTB-2013 [19] (50 videos), OTB-2015 [20] (100 videos) and Temple-Color [21] (128 videos). All the tracking methods are evaluated by the overlap precision (OP) using

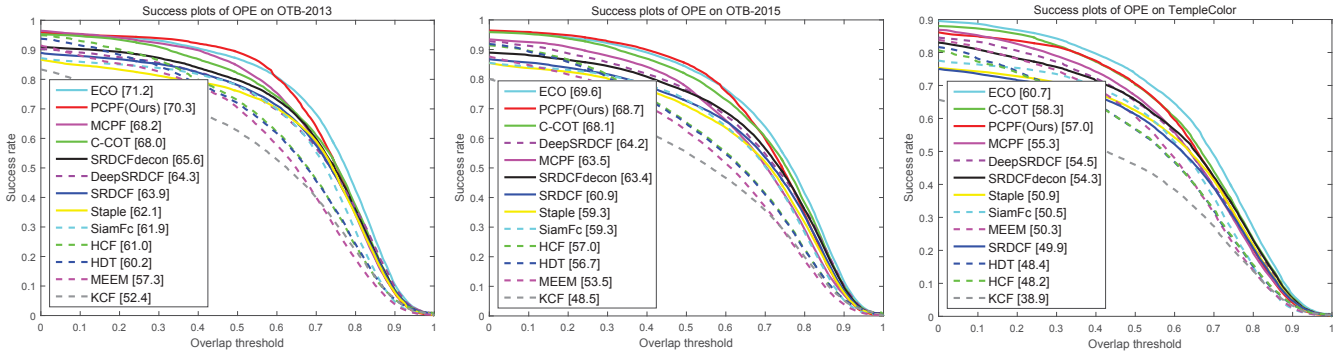


Fig. 4. Success plots on the OTB-2013 [19], OTB-2015 [20] and Temple-Color [21] datasets. In the legend, the area-under-curve (AUC) score is reported. The proposed PCPF tracker shows state-of-the-art performance.

Table 1. Effectiveness study of the proposed framework. The distance precision (DP) at 20 pixels threshold and overlap precision (OP) at an overlap threshold 0.5 are reported on OTB-2015 [20] datasets. Both coarse-to-fine localization and bottom-up update help improve the performance.

	Only Global	→	Global+Local	→	Global+Local+Update (Final)
DP	86.1%	→	89.0% (**3.9%** ↑)	→	90.2% (**1.2%** ↑)
OP	80.8%	→	84.5% (**3.7%** ↑)	→	85.3% (**0.8%** ↑)

one-pass evaluation (OPE) [19, 20]. We provide the success plots using OP metric over a range of thresholds, and the area-under-curve (AUC) score evaluates the overall performance.

3.2. Framework Effectiveness Study

In Table 1, we study the effectiveness of the proposed framework on the OTB-2015 benchmark. Due to the limited search range of particles, our local DCFs rely on the coarse localization of holistic model. The combination of global and local models in a coarse-to-fine fasion obviously improves the accuracy of only individual global model (3.9% in DP and 3.7% in OP on the OTB-2015), which means the part-based correlation particle filter refines the target location. Besides, the local reliability estimation helps adaptively update global DCF, which in return keeps the global tracker from corruption to some extent, and boosts the final performance further (1.2% in DP and 0.8% in OP on the OTB-2015).

3.3. State-of-the-art Comparison

We evaluate our PCPF algorithm with 12 recent state-of-the-art trackers including KCF [2], MEEM [24], HCF [12], SRDCF [15], DeepSRDCF [25], SRDCFdecon [26], HDT [13], Staple [22], MCPF [16], C-COT [14] and ECO [3].

Figure 4 shows the success plots of our approach and state-of-the-art trackers. The proposed PCPF tracker provides the AUC scores of (70.3 %, 68.7%, 57.0%) on the OTB-2013, OTB-2015 and Temple-Color datasets, while the recent C-COT and MCPF exhibit (68.0%, 68.1%, 58.3%) and (68.2%, 63.5%, 55.3%), respectively. Overall, our tracker shows compariable results compared to C-COT and outperforms the recent MCPF method on all three datasets. Specially, our tracker surpasses MCPF with a gain of 5.2% AUC on the challenging OTB-2015 benchmark. We do not show the results of MDNet [27] which utilizes various similar tracking videos for network training and needs re-trainging when evaluating on Temple-Color dataset. The MDNet method provides the AUC score of 67.8% on the OTB-2015, and our tracker shows better performance (68.7%). The enhanced version of C-COT is ECO [3], which improves both speed and performance by introducing several efficient strategies. However, it should be noted that our method only uses standard DCF, more sophisticated models (*e.g.,* SRDCF, ECO), more types of features [11, 28, 29, 30] and other novel techniques [31, 32, 3] can be integrated into our model to further boost the performance, which will be studied in our future work.

In Table 2, we analyze the common challenging factors in visual tracking task. From the results we can observe that out method is adept at handling deformation, rotation and background clutter (outperforms C-COT by a considerable margin), which can be attributed to our part-based model with resample mechanism. The C-COT tracker only uses global model and involves larger searching range (RoI) compared with traditional DCF due to the effectiveness of SRDCF, which is more suitable for tracking targets with large motion range. In contrast, due to the proposed part-based correlation particle filter framework, our tracker is more suitable for tracking nonrigid objects with drastic deformation and rotation.

4. CONCLUSION

In this paper, we propose a part-based correlation particle filter for robust object tracking. With the proposed particle resample mechanism with reliabilty consideration, our method

Table 2. Per-attributed comparison between our approach and state-of-the-art trackers. The eleven attributes are deformation (DEF), scale variation (SV), illumination variation (IV), occlusion (OCC), motion blur (MB), fast motion (FM), in-plane rotation (IPR), out-of-plane rotation (OPR), out of view (OV), background clutter (BC) and low resolution (LR). The AUC score is reported on the OTB-2015 [20]. The first and second highest values are highlighted by bold and underline.

Attribute (Video Number)	SiamFc [33]	DeepSRDCF [25]	MCPF [16]	C-COT [14]	ECO [3]	**PCPF Ours**
DEF (44)	51.9	56.7	57.3	61.6	63.6	**65.4**
SV (64)	56.9	62.0	61.8	67.4	**68.5**	64.9
IV (38)	57.8	61.9	62.9	68.2	**71.4**	70.0
OCC (49)	55.3	60.1	62.2	67.6	**68.2**	65.8
MB (29)	56.1	63.7	59.7	70.4	**70.6**	67.7
FM (39)	57.8	62.6	59.7	67.7	**68.4**	66.5
IPR (51)	56.7	58.8	62.3	62.8	65.6	**66.4**
OPR (63)	56.7	60.9	62.3	65.6	**67.7**	66.9
OV (14)	51.6	56.4	56.1	66.1	**67.4**	63.4
BC (31)	53.2	63.9	61.2	66.4	**71.4**	69.4
LR (9)	63.4	57.4	59.5	64.3	60.3	**69.2**
Overall (100)	59.3	64.2	63.5	68.1	**69.6**	68.7

preserves object structure and pays more attention to the discriminative and stable target parts. To cope with the limited search range of local particles and inadequate update issues, coarse-to-fine localization and adaptive update strategies are introduced. Experiments on several datasets demonstrate the state-of-the-art performance of our tracker, especially on tracking nonrigid targets.

5. REFERENCES

[1] Kaihua Zhang, Lei Zhang, and Ming-Hsuan Yang, "Real-time compressive tracking," in *ECCV*, 2012.

[2] João F Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista, "High-speed tracking with kernelized correlation filters," *IEEE TPAMI*, vol. 37, no. 3, pp. 583–596, 2015.

[3] Martin Danelljan, Goutam Bhat, Fahad Shahbaz Khan, and Michael Felsberg, "Eco: Efficient convolution operators for tracking," in *CVPR*, 2017.

[4] Si Liu, Tianzhu Zhang, Xiaochun Cao, and Changsheng Xu, "Structural correlation filter for robust visual tracking," in *CVPR*, 2016.

[5] Ting Liu, Gang Wang, and Qingxiong Yang, "Real-time part-based visual tracking via adaptive correlation filters," in *CVPR*, 2015.

[6] Heng Fan and Jinhai Xiang, "Robust visual tracking via local-global correlation filter," in *AAAI*, 2017.

[7] A Lukezic, L. C. Zajc, and M Kristan, "Deformable parts correlation filters for robust visual tracking," *IEEE TCyb*, vol. PP, no. 99, pp. 1–13, 2017.

[8] David S Bolme, J Ross Beveridge, Bruce A Draper, and Yui Man Lui, "Visual object tracking using adaptive correlation filters," in *CVPR*, 2010.

[9] Kaihua Zhang, Lei Zhang, Ming-Hsuan Yang, and David Zhang, "Fast visual tracking via dense spatio-temporal context learning," in *ECCV*, 2013.

[10] Martin Danelljan, Gustav Häger, Fahad Khan, and Michael Felsberg, "Accurate scale estimation for robust visual tracking," in *BMVC*, 2014.

[11] Yang Li and Jianke Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *ECCV Workshop*, 2014.

[12] Chao Ma, Jia-Bin Huang, Xiaokang Yang, and Ming-Hsuan Yang, "Hierarchical convolutional features for visual tracking," in *ICCV*, 2015.

[13] Yuankai Qi, Shengping Zhang, Lei Qin, Hongxun Yao, Qingming Huang, and Jongwoo Lim Ming-Hsuan Yang, "Hedged deep tracking," in *CVPR*, 2016.

[14] Martin Danelljan, Andreas Robinson, Fahad Shahbaz Khan, and Michael Felsberg, "Beyond correlation filters: Learning continuous convolution operators for visual tracking," in *ECCV*, 2016.

[15] Martin Danelljan, Gustav Hager, Fahad Shahbaz Khan, and Michael Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *ICCV*, 2015.

[16] Tianzhu Zhang, Changsheng Xu, and Ming-Hsuan Yang, "Multi-task correlation particle filter for robust object tracking," in *CVPR*, 2017.

[17] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE TSP*, vol. 50, no. 2, pp. 174–188, 2002.

[18] Yang Li, Jianke Zhu, and Steven C. H. Hoi, "Reliable patch trackers: Robust visual tracking by exploiting reliable patches," in *CVPR*, 2015.

[19] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang, "Online object tracking: A benchmark," in *CVPR*, 2013.

[20] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang, "Object tracking benchmark," *IEEE TPAMI*, vol. 37, no. 9, pp. 1834–1848, 2015.

[21] Pengpeng Liang, Erik Blasch, and Haibin Ling, "Encoding color information for visual tracking: algorithms and benchmark," *IEEE TIP*, vol. 24, no. 12, pp. 5630–5644, 2015.

[22] Luca Bertinetto, Jack Valmadre, Stuart Golodetz, Ondrej Miksik, and Philip Torr, "Staple: Complementary learners for real-time tracking," in *CVPR*, 2016.

[23] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *CVPR*, 2005.

[24] Jianming Zhang, Shugao Ma, and Stan Sclaroff, "Meem: robust tracking via multiple experts using entropy minimization," in *ECCV*, 2014.

[25] Martin Danelljan, Gustav Hager, Fahad Shahbaz Khan, and Michael Felsberg, "Convolutional features for correlation filter based visual tracking," in *ICCV Workshop*, 2015.

[26] Martin Danelljan, Gustav Häger, Fahad Shahbaz Khan, and Michael Felsberg, "Adaptive decontamination of the training set: A unified formulation for discriminative visual tracking," in *CVPR*, 2016.

[27] Hyeonseob Nam and Bohyung Han, "Learning multi-domain convolutional neural networks for visual tracking," in *CVPR*, 2016.

[28] Weijian Kuan, Jun Chen, Chao Liang, Yi Wu, and Ruimin Hu, "Object tracking via online trajectory optimization with multi-feature fusion," in *ICME*, 2017.

[29] Kaihua Zhang, Lei Zhang, and Ming-Hsuan Yang, "Real-time object tracking via online discriminative feature selection," *IEEE TIP*, vol. 22, no. 12, pp. 4664–4677, 2013.

[30] Tianqi Zheng, Chao Xie, Wengang Zhou, and Houqiang Li, "Compressive tracking with adaptive color feature selection and foreground modeling," in *VCIP*, 2016.

[31] Ning Wang, Wengang Zhou, Qi Tian, Richang Hong, Meng Wang, and Houqiang Li, "Multi-cue correlation filters for robust visual tracking," in *CVPR*, 2018.

[32] Ning Wang, Wengang Zhou, and Houqiang Li, "Reliable re-detection for long-term tracking," *IEEE TCSVT*, 2018.

[33] Luca Bertinetto, Jack Valmadre, João F Henriques, Andrea Vedaldi, and Philip HS Torr, "Fully-convolutional siamese networks for object tracking," in *ECCV*, 2016.