Adaptive Context-Aware Discriminative Correlation Filters for Robust Visual Object Tracking

Tianyang Xu, Zhen-Hua Feng, Xiao-Jun Wu, Josef Kittler

ICPR 2020 25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION Milan, Italy 10 | 15 January 2021

UNIVERSITY OF SURREY
JIANGNAN UNIVERSITY 江南大学

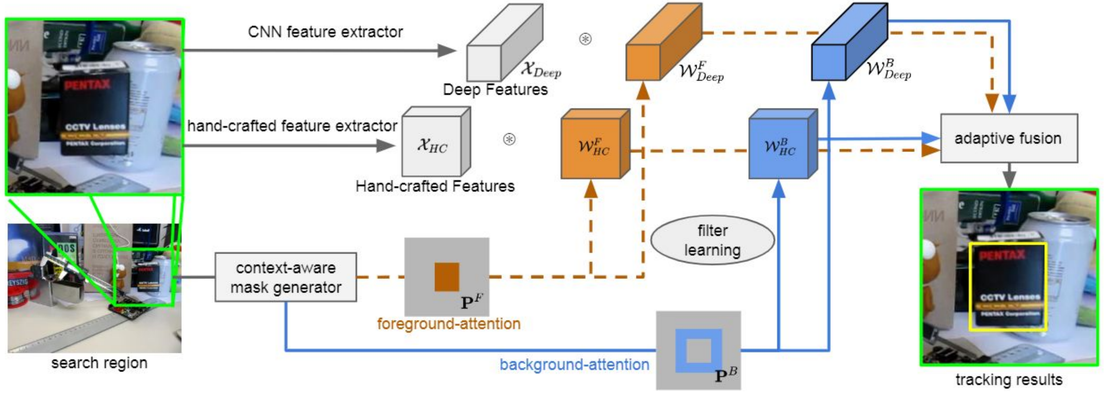
Introduction

In recent years, Discriminative Correlation Filters (DCFs) have gained popularity due to their superior performance in visual object tracking. However, existing DCF trackers usually learn filters using fixed attention mechanisms that focus on the centre of an image and suppresses filter amplitudes in surroundings. In this paper, we propose an Adaptive Context-Aware Discriminative Correlation Filter (ACA-DCF) that is able to improve the existing DCF formulation with complementary attention mechanisms. Our ACA-DCF integrates foreground attention and background attention for complementary context-aware filter learning. More importantly, we ameliorate the design using an adaptive weighting strategy that takes complex appearance variations into account. The experimental results obtained on several well-known benchmarks demonstrate the effectiveness and superiority of the proposed method over the state-of-the-art approaches.

Challenges

- Online learning
- Limited training samples
- Challenging appearance variations
- Unpredictable background clutters and occlusions

Methodology



↑ ***The proposed adaptive context-aware DCF method using both hand-crafted and deep features. The proposed context-aware mask generator is instrumental in applying foreground-background attention to the learned filters. The corresponding responses are adaptively fused to generate the final result..***

OurACA-DCF - a new foreground background context-aware tracker

$$\min \left\| \sum_{k=1}^{C} \mathbf{X}^k * \mathbf{W}^k - \mathbf{Y} \right\|_F^2 + \lambda \sum_{k=1}^{C} \left\| \mathbf{P} \odot \mathbf{W}^k \right\|_F^2$$

Optimisation - ADMM

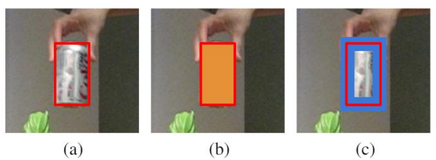
Lagrange function:

$$\mathcal{L} = \left\| \sum_{k=1}^{C} \mathbf{X}^k * \mathbf{W}^k - \mathbf{Y} \right\|_F^2 + \lambda \sum_{k=1}^{C} \left\| \mathbf{P} \odot \mathbf{W}'^k \right\|_F^2 + \frac{\mu}{2} \sum_{k=1}^{C} \left\| \mathbf{W}^k - \mathbf{W}'^k + \frac{\mathbf{\Gamma}^k}{\mu} \right\|_F^2,$$

Solution:

$$\begin{cases} \hat{\mathbf{w}}_{i,j} = \left(\mathbf{I} - \dfrac{\hat{\mathbf{x}}_{i,j} \hat{\mathbf{x}}_{i,j}^\top}{\mu/2 + \hat{\mathbf{x}}_{i,j}^\top \hat{\mathbf{x}}_{i,j}} \right) \mathbf{g} \\ \mathbf{W}'^k = (\mathbf{1} - \mathbf{P}) \odot \dfrac{\mu \mathbf{W}^k + \mathbf{\Gamma}^k}{2\lambda + \mu} \\ \mathbf{\Gamma} = \mathbf{\Gamma} + \mu (\mathcal{W} - \mathcal{W}') \end{cases}$$

$$\mathbf{g} = \left(\hat{\mathbf{x}}_{i,j} \hat{y}_{i,j} + \mu \hat{\mathbf{w}}'_{i,j} - \mu \hat{\gamma}_{i,j} \right) / \mu$$

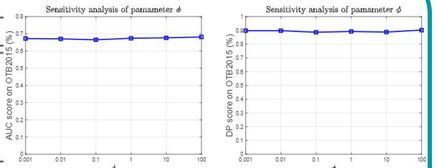


↑ ***Illustration of the complementary attention mechanism based context-aware mask generator: (a) target bounding box; (b) foreground-attention mask; and (c) background-attention mask.***

Ablation study

Methods	DP(%)	AUC(%)
Baseline	82.2	61.2
Baseline+F	87.3	65.1
Baseline+F+B	89.2	67.2
Baseline+F+B+AF	92.4	69.7

↑ ***Ablation studies on OTB2015.***
F: foreground-attention, B: background-attention
AF: adaptive fusion strategy

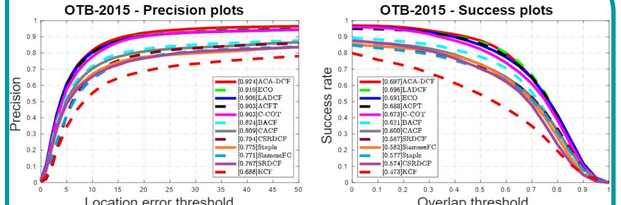


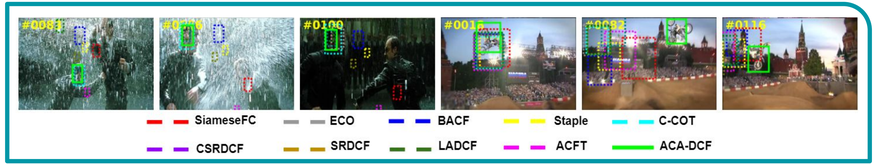
↑ ***Sensitivity analysis of parameter ϕ.***

Evaluation on VOT-2016

	STAPLE+	EBT	DDC	Staple	MLDF	SSAT	TCNN	C-COT	ACA-DCF
EAO	0.286	0.291	0.293	0.295	0.311	0.329	**0.327**	**0.331**	**0.385**
Accuracy	**0.559**	0.465	0.542	0.547	0.492	**0.579**	0.555	0.541	**0.581**
Robustness	0.37	0.25	0.34	0.38	**0.23**	0.29	0.27	**0.24**	**0.21**

Evaluation on OTB-2015





Acknowledgements

EPSRC
NSFC 国家自然科学基金委员会 National Natural Science Foundation of China

References

1. Henriques, João F., et al. "High-speed tracking with kernelized correlation filters." TPAMI. 2015.
2. Xu, Tianyang, et al. "Learning Adaptive Discriminative Correlation Filters via Temporal Consistency preserving Spatial Feature Selection for Robust Visual Object Tracking." TIP. 2019.