Joint Group Feature Selection and Discriminative Filter Learning for Robust Visual Object Tracking

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

JIANGNAN UNIVERSITY

UNIVERSITY OF SURREY

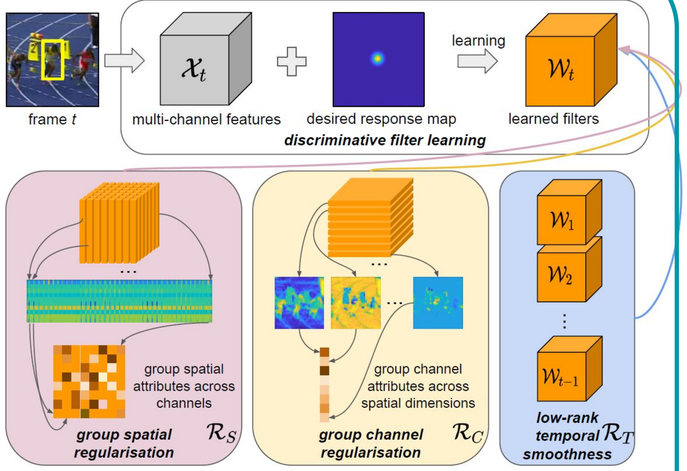
ICCV 2019 Seoul, Korea

Introduction

We propose a new Group Feature Selection method for Discriminative Correlation Filters (GFS-DCF) based visual object tracking. The key innovation of the proposed method is to perform group feature selection across both channel and spatial dimensions, thus to pinpoint the structural relevance of multi-channel features to the filtering system. In contrast to the widely used spatial regularisation or feature selection methods, to the best of our knowledge, this is the first time that channel selection has been advocated for DCF-based tracking. We demonstrate that our GFS-DCF method is able to significantly improve the performance of a DCF tracker equipped with deep neural network features. In addition, our GFS-DCF enables joint feature selection and filter learning, achieving enhanced discrimination and interpretability of the learned filters.

To further improve the performance, we adaptively integrate historical information by constraining filters to be smooth across temporal frames, using an efficient low-rank approximation. By design, specific temporal-spatial-channel configurations are dynamically learned in the tracking process, highlighting the relevant features, and alleviating the performance degrading impact of less discriminative representations and reducing information redundancy.

Methodology



↑ *In contrast to the classical DCF paradigm, our GFS-DCF performs channel and spatial group feature selection for the learning of correlation filters. Group sparsity is enforced in the channel and spatial dimensions to highlight relevant features with enhanced discrimination and interpretability. Additionally, a low-rank temporal smoothness constraint is employed across temporal frames to improve the stability of the learned filters.*

Challenges

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

Solutions

- Employ circulant matrix to generate augmented training samples
- Powerful deep concolutional feature maps, *e.g.*, VGG, ResNet
- Spatial regularisations

Issues

- Spatial boundary effect caused by circulant shift
- Representation redundancy and noise caused by deepfeature maps
- Temporal filter degeneration caused by online learning-detection

Our GFS-DCF - a new group feature selection based tracker

$$\widetilde{\mathcal{W}} = \arg\min_{\mathcal{W}} \left\| \sum_{k=1}^{C} \mathbf{W}^k * \mathbf{X}^k - \mathbf{Y} \right\|_F^2 + \lambda_1 \sum_{i=1}^{N} \sum_{j=1}^{N} \|\mathbf{w}_{ij:}\|_2 + \lambda_2 \sum_{k=1}^{C} \left\|\mathbf{W}^k\right\|_F + \lambda_3 \sum_{k=1}^{C} \left\|\mathbf{W}_t^k - \mathbf{W}_{t-1}^k\right\|_F^2$$

Optimisation - ADMM

Lagrange function:

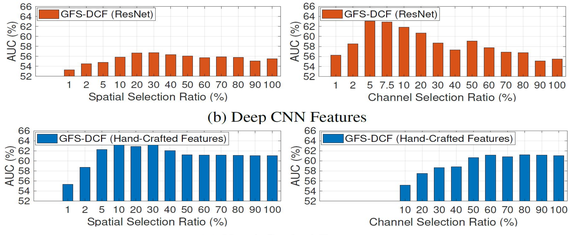
$$\begin{aligned}\mathcal{L} = & \left\| \sum_{k=1}^{C} \mathbf{W}_t^k * \mathbf{X}_t^k - \mathbf{Y} \right\|_F^2 + \lambda_1 \sum_{k=1}^{C} \left\|\mathbf{W}_t'^k\right\|_F \\ & + \lambda_2 \sum_{i=1}^{N} \sum_{j=1}^{N} \left\|\mathbf{w}'_{ij_t}\right\|_2 + \lambda_3 \sum_{k=1}^{C} \left\|\mathbf{W}_t^k - \mathbf{W}_{t-1}^k\right\|_F^2 \\ & + \frac{\mu}{2} \sum_{k=1}^{C} \left\| \mathbf{W}_t^k - \mathbf{W}_t'^k + \frac{\mathbf{\Gamma}^k}{\mu} \right\|_F^2,\end{aligned}$$

Solution:

$$\begin{cases} \hat{\mathbf{w}}_{ij_t} = \left(\mathbf{I} - \dfrac{\hat{\mathbf{x}}_{ij_t} \hat{\mathbf{x}}_{ij_t}^H}{(\lambda_3 + \mu/2)\, N^2 + \hat{\mathbf{x}}_{ij_t}^H \hat{\mathbf{x}}_{ij_t}} \right) \mathbf{q}, \\ w'^k_{ij_t} = \max\left(0, 1 - \dfrac{\lambda_1}{\mu \left\|\mathbf{P}^k\right\|_F} - \dfrac{\lambda_2}{\mu \left\|\mathbf{p}_{ij}\right\|_2}\right) p_{ij}^k, \\ \mathbf{\Gamma} \quad = \mathbf{\Gamma} + \mu \left(\mathcal{W}_t - \mathcal{W}'_t\right), \end{cases}$$

$$\mathbf{q} = \left(\hat{\mathbf{x}}_{ij_t} \hat{y}_{ij} / N^2 + \mu \hat{\mathbf{w}}'_{ij_t} - \mu \hat{\gamma}_{ij} + \lambda_3 \hat{\mathbf{w}}_{ij_{t-1}}\right) / (\lambda_3 + \mu)$$

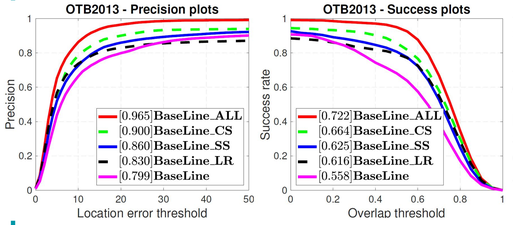
$$p_{ij}^k = w_{ij}^k + \gamma_{ij}^k / \mu$$



↑ *A comparison of spatial and channel group feature selection on OTB2015 using either (a) hand-crafted or (b) deep CNN features, parameterised by selection ratio.*

↑ *Visualisation of filters using David3 in OTB2015. We visualise the corresponding filters in frame #50 and #200. To better visualise the sparsity, we present the heat-maps of the obtained filters by gathering the energy across all the channels.*

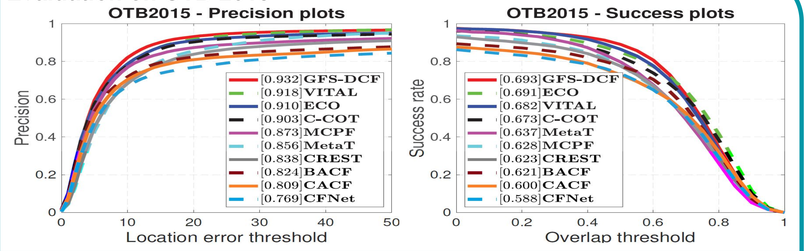
Ablation study



← ***A comparison of different regularisation terms in GFSDCF, evaluated on OTB2013. The precision plots with DP and the success plots with AUC in the legends are presented.***

LR : low-rank temporal smoothness term
SS : spatial feature selection term
CS : channel feature selection term
ALL : the proposed GFS-DCF

Evaluation on OTB-2015



Evaluation on VOT-2018

Tracking results on VOT2017/VOT2018.

	ECO	CFCF	CFWCR	LSART	UPDT	SiamRPN	MFT	LADCF	GFS-DCF
EAO	0.280	0.286	0.303	0.323	0.378	0.383	0.385	0.389	**0.397**
Accuracy	0.483	0.509	0.484	0.493	0.536	**0.586**	0.505	0.503	0.511
Robustness	0.276	0.281	0.267	0.218	0.184	0.276	**0.140**	0.159	0.143

Evaluation on TrackingNet

Tracking performance on the TrackingNet test set.

Method	Success	Precision	Normalised Precision
CACF	53.59%	46.72%	60.84%
ECO	56.13%	48.86%	62.14%
MDNet	**61.35%**	55.53%	71.00%
GFS-DCF	60.90%	**56.57%**	**71.79%**

Conclusion

- An effective appearance model with outstanding performance by learning spatial-channel group-sparse discriminative correlation filters, constrained by low-rank approximation across successive frames.

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.

Acknowledgements

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

More information

← Github
WeChat →