Learning Spatially Regularized Correlation Filters for Visual Tracking

Martin Danelljan  Gustav Häger  Fahad Khan  Michael Felsberg 

Computer Vision Laboratory, Linköping University, Sweden

- **Winner** of OpenCV Challenge in tracking [1].
- Among top 3 in **VOT 2015** and **VOT-TIR 2015.**
- **Best** results on OTB-2013 published in **ICCV.**

Introduction

Application: Visual tracking
Goal: Address problems caused by the periodic assumption in Discriminative Correlation Filters (DCF).

- **Limits** training data
- **Restricts** search region
- **Inaccurate** negative training samples
- Increased sample size **degrades** discriminative power

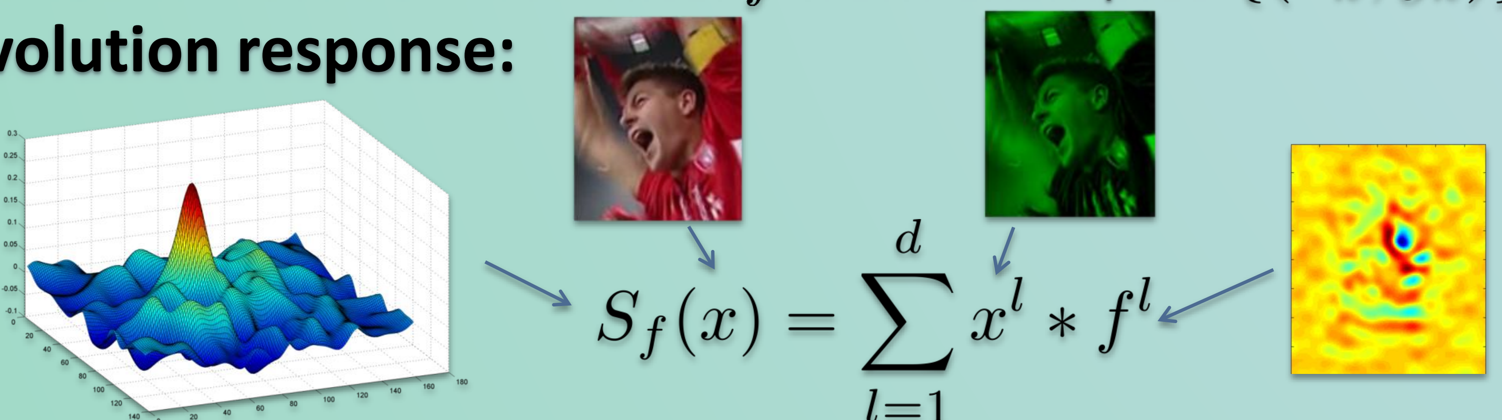
Contributions

- **Spatial regularization** component in the DCF learning.
  - **Penalizes** filter coefficients based on the target size.
  - **Alleviates** the negative effects of the periodic assumption.
  - Training and detection on **larger image regions.**
  - Increases **discriminative** power => more **robust** tracking.
- Efficient **optimization** method.
  - Unlike previous DCF trackers, **no** approximate learning.
- Fast target detection with **sub-grid accuracy.**
- **Best results on 4 datasets:** OTB-13, OTB-15, ALOV, VOT2014.

The Standard DCF Tracking

Idea: Learn a convolution filter f from examples $\{(x_k, y_k)\}_{k=1}^t$

Convolution response:

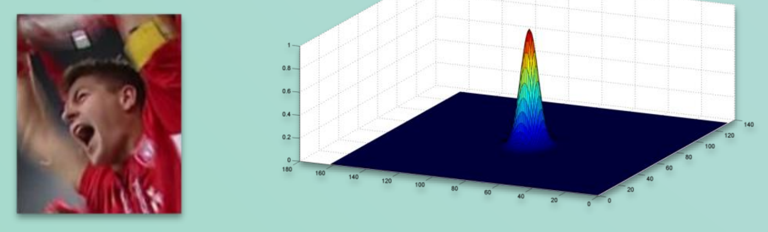


$$S_f(x) = \sum_{l=1}^{d} x^l * f^l$$

Learning loss:

$$\varepsilon_t(f) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k) - y_k \right\|^2 + \lambda \sum_{l=1}^{d} \left\| f^l \right\|^2$$

Optimization:
Solved using approximate update rules, leading to suboptimal filters f.

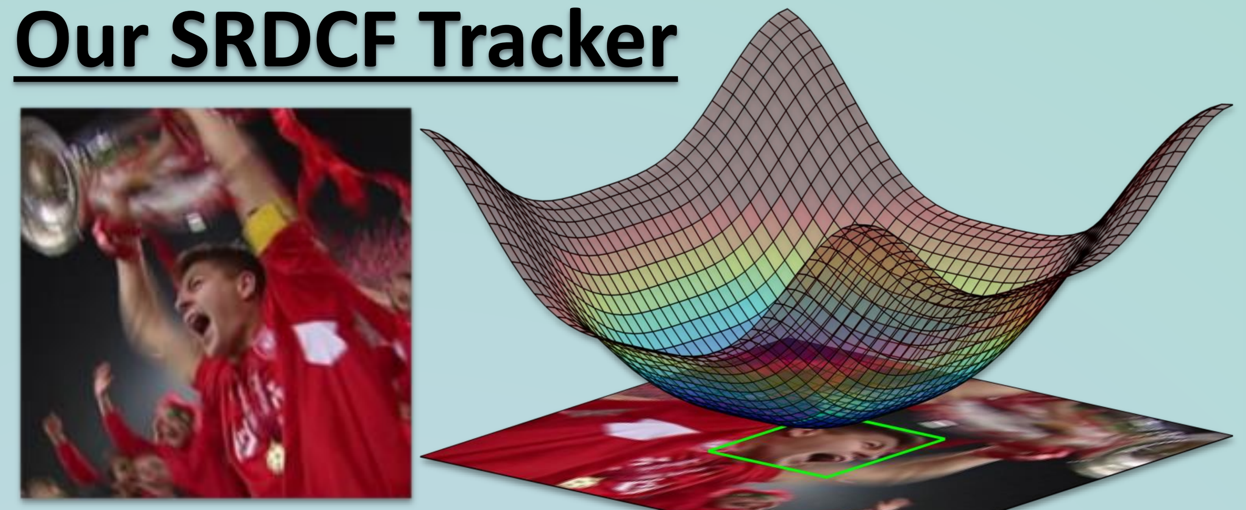


Detection:
Maximize classification scores in the image sample z.

$$S_f(z) = \mathscr{F}^{-1}\left\{ \sum_{l=1}^{d} \hat{z}^l \cdot \hat{f}^l \right\}$$

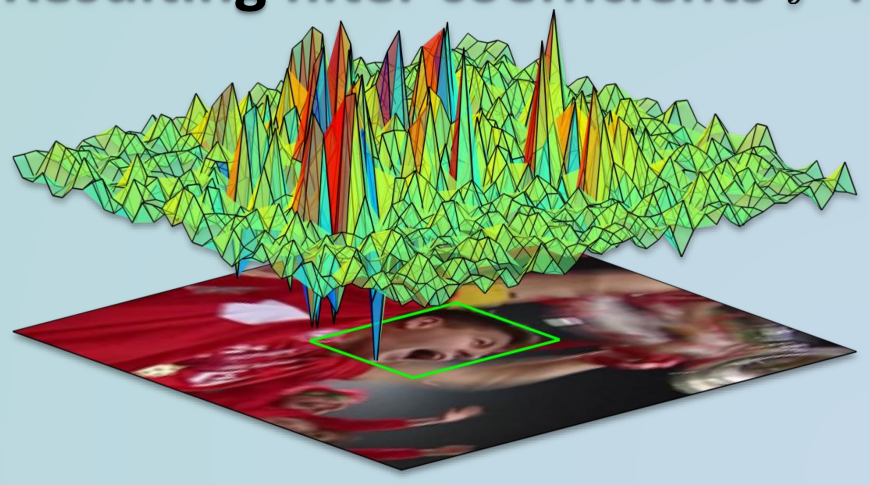
Our SRDCF Tracker

Our loss:
The regularization weights w penalize filter coefficients in the background.

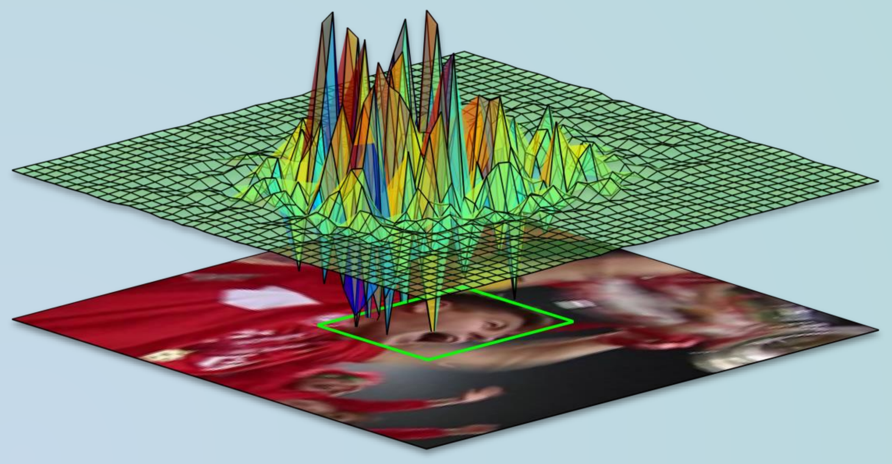


$$\varepsilon(f) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k) - y_k \right\|^2 + \sum_{l=1}^{d} \left\| w \cdot f^l \right\|^2$$

Resulting filter coefficients f^l**:**



Standard regularization



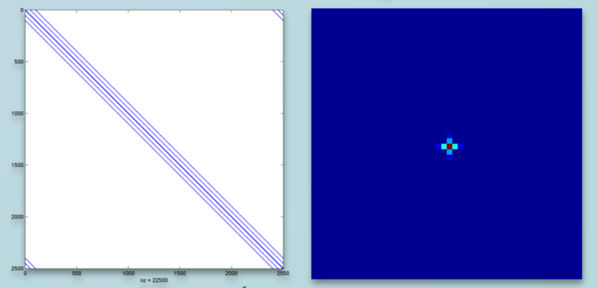
Our spatial regularization

Efficient optimization:
Parseval:

Discrete Fourier Transform (DFT)

$$\breve{\varepsilon}(\hat{f}) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} \hat{x}_k^l \cdot \hat{f}^l - \hat{y}_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{\hat{w}}{MN} * \hat{f}^l \right\|^2$$

The regularization weights w have sparse DFT coefficients $\hat{w}$, which gives a sparse convolution matrix $\mathcal{C}(\hat{\mathbf{w}})$.



Vectorization:

$$\breve{\varepsilon}(\hat{f}) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} \mathcal{D}(\hat{\mathbf{x}}_k^l) \hat{\mathbf{f}}^l - \hat{\mathbf{y}}_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{\mathcal{C}(\hat{\mathbf{w}})}{MN} \hat{\mathbf{f}}^l \right\|^2$$

Normal equations are solved in a **real Fourier basis** $A_t \tilde{\mathbf{f}} = \tilde{\mathbf{b}}$.
Efficient **incremental update** of the system:

$$A_t = \sum_{k=1}^{t} \alpha_k D_k^{\mathrm{T}} D_k + W^{\mathrm{T}} W \qquad \tilde{\mathbf{b}}_t = \sum_{k=1}^{t} \alpha_k D_k^{\mathrm{T}} \tilde{\mathbf{y}}_k$$

- Iterative solution using Gauss-Seidel by **exploiting sparsity**.
- Initialization with previous solution ensures **fast convergence**.

Fast sub-grid detection
Fourier domain interpolation

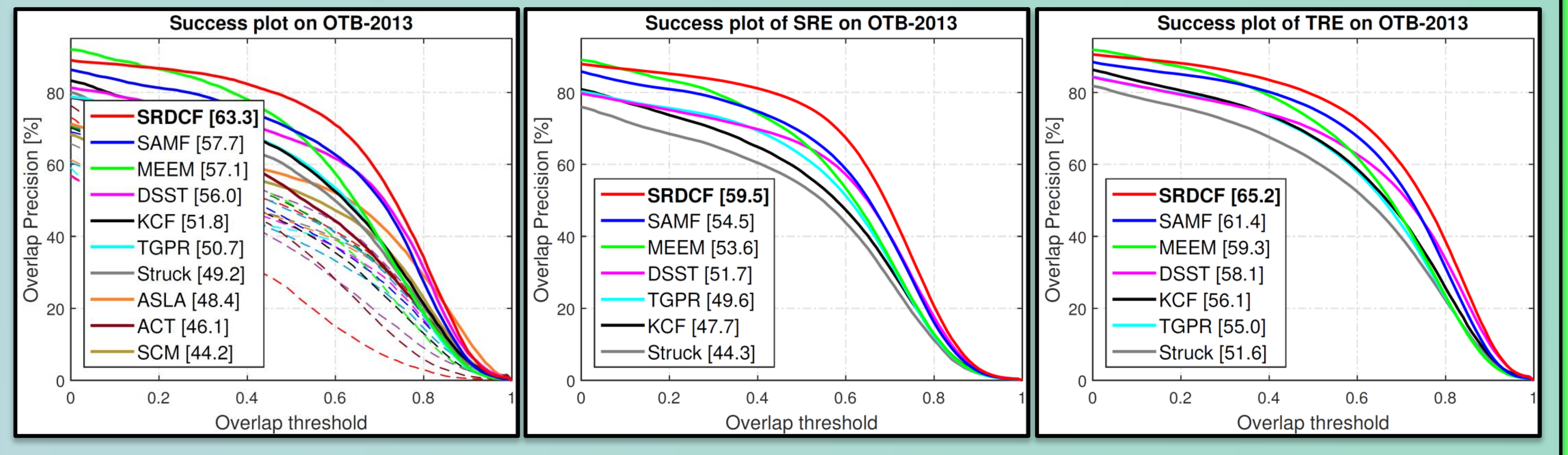
$$s(u,v) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \hat{s}(m,n) e^{i2\pi\left(\frac{m}{M}u + \frac{n}{N}v\right)}$$

Experiments

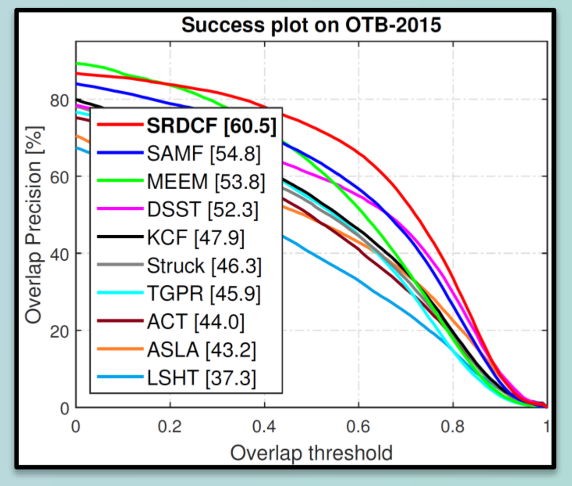
Baseline evaluation on OTB-2013 [2].

Regularization	Conventional sample size: Standard	Conventional sample size: Ours	Expanded sample size: Standard	Expanded sample size: Ours
Mean OP (%)	71.1	72.2	50.1	78.1

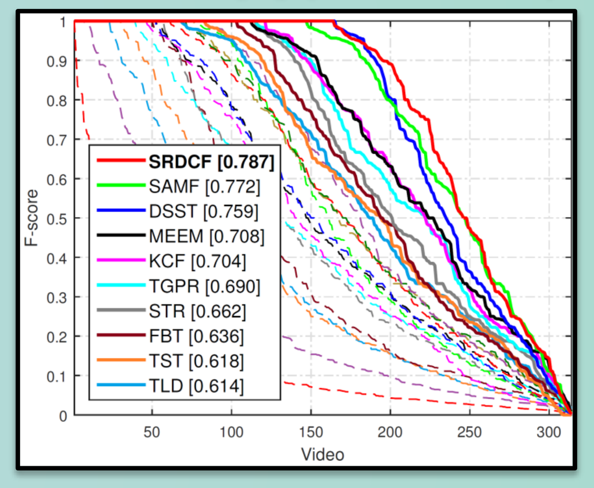
Online tracking benchmark (OTB) 2013 [2] (50 videos)



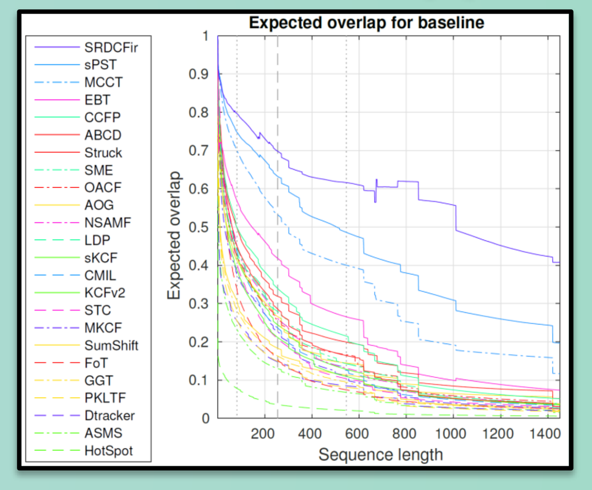
OTB-2015 [3] (100 videos)



ALOV++ [4] (314 videos)



VOT-TIR 2015 [5]



VOT2014 [6] (25 videos)

	Overlap	Failures	Acc. Rank	Rob. Rank	Final Rank
SRDCF	0.63	15.90	6.43	10.08	8.26
DSST	0.64	16.90	5.99	11.17	8.58
SAMF	0.64	19.23	5.87	14.14	10.00

Comparison with ICCV 2015 trackers on OTB-2013

Tracker	AUC (%)
SRDCF (ours)	**63.3**
SOWP (Han-Ul Kim, Dae-Youn Lee, Jae-Young Sim, Chang-Su Kim)	61.9
Understanding and Diagnosing (Naiyan Wang, Jianping Shi, Dit-Yan Yeung, Jiaya Jia)	61.8
HCF (Chao Ma, Jia-Bin Huang, Xiaokang Yang, Ming-Hsuan Yang)	60.5
FCN (Lijun Wang, Wanli Ouyang, Xiaogang Wang, Huchuan Lu)	59.9
Proposal Selection (Yang Hua, Karteek Alahari, Cordelia Schmid)	58.0
TRIC-track (Xiaomeng Wang, Michel Valstar, Brais Martinez, Muhammad Haris Khan, Tony Pridmore)	53.0
LNL (Bo Ma, Hongwei Hu, Jianbing Shen, Yuping Zhang, Fatih Porikli)	50.8

References

[1] http://code.opencv.org/projects/opencv/wiki/VisionChallenge.
[2] Y. Wu, J. Lim, and M.-H. Yang. Online object tracking: A benchmark. In CVPR, 2013.
[3] Y. Wu, J. Lim, and M.-H. Yang. Object tracking benchmark. PAMI, 2015.
[4] A. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah. Visual tracking: An experimental survey. PAMI 2014.
[5] M. Felsberg, A. Berg et al. The Thermal Infrared Visual Object Tracking VOT-TIR2015 Challenge Results. In ICCV Workshop, 2015.
[6] M. Kristan, R. Pflugfelder, A. Leonardis, J. Matas, and et al. The visual object racking vot2014 challenge results. In ECCV Workshop, 2014.

Matlab code available!