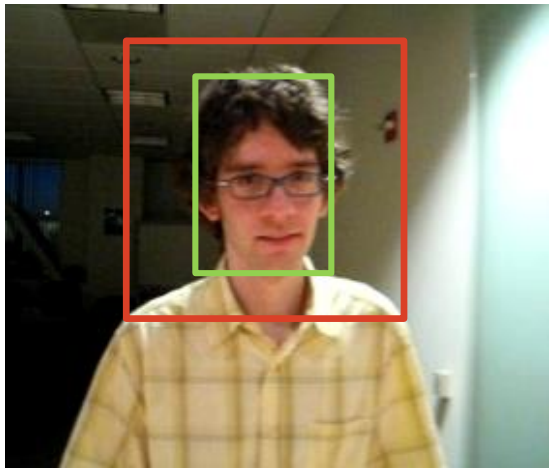
Discriminative Correlation Filter with Channel and Spatial Reliability (CSR-DCF)

Alan Lukežič[1], Andrej Muhič[1], Tomáš Vojíř[2],
Luka Čehovin[1], Jiří Matas[2], Matej Kristan[1]

1 Faculty of Computer and Information Science, University of Ljubljana, Slovenia
2 Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic

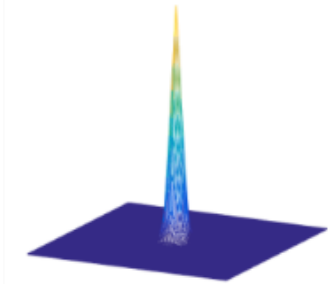
https://github.com/alanlukezic/csr-dcf

Single-Channel CFs: Filter Learning



Training example: P

Filter: H



Desired response: G

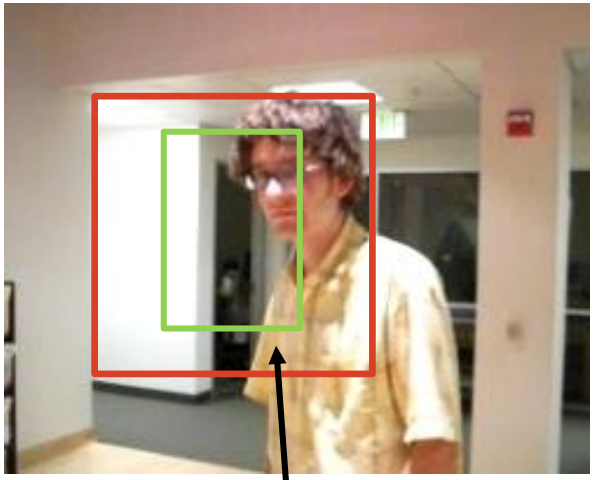
$$\arg\min_{\tilde{\mathbf{H}}} |\mathbf{P} * \tilde{\mathbf{H}} - \mathbf{G}|^2 = \arg\min_{\hat{\mathbf{H}}^\dagger} |\hat{\mathbf{P}} \odot \hat{\mathbf{H}}^\dagger - \hat{\mathbf{G}}|^2$$

Closed-form solution: $\hat{\mathbf{H}}^\dagger = \dfrac{\hat{\mathbf{G}} \odot \hat{\mathbf{P}}^\dagger}{\hat{\mathbf{P}} \odot \hat{\mathbf{P}}^\dagger}$

the hat symbol ($\widehat{\ }$) denotes the Fourier domain image

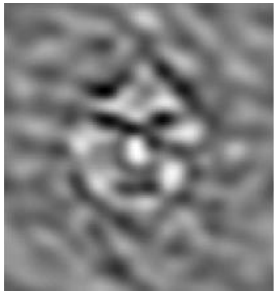
green bbox: target region, red bbox: search region

Single-Channel CFs: Target Localization

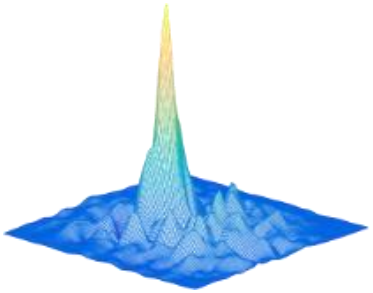


Position from previous frame

Localization patch: P



Filter: H



Correlation response: R

Efficiently in Fourier domain:

$$\mathbf{R} = \mathcal{F}^{-1}\left(\hat{\mathbf{P}} \odot \hat{\mathbf{H}}^{\dagger}\right)$$

Position of the maximum: new target location

Issues With Standard CFs: Search Region

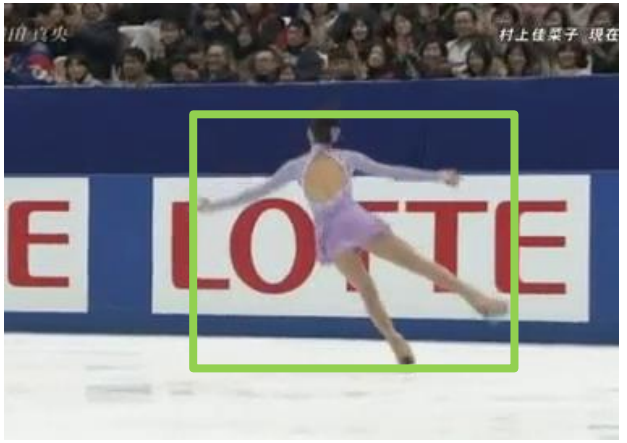
Filter learned from cyclic shifts

Unrealistic training examples

Search region size equal to template

Difficult to address large displacements

Poor approximation with bbox

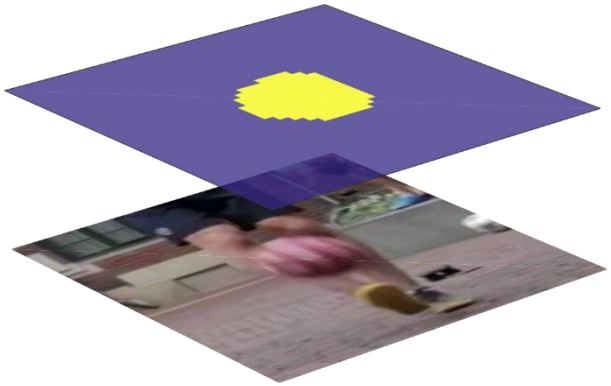


Background enters filter

Contribution 1: Spatial Relibility Map

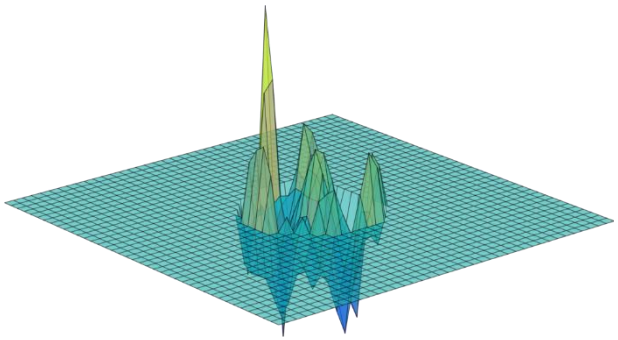
- Assume an approximate binary target segmentation M

Spatial reliability map M



Training patch

Learned filter



Spatial constraint:

$$\mathbf{H}_{\mathrm{M}} = \mathbf{H} \odot \mathbf{M}$$

Cost function:

$$\arg\min_{\hat{\mathbf{H}}^{\dagger}} \left(|\hat{\mathbf{P}} \odot \hat{\mathbf{H}}^{\dagger} - \hat{\mathbf{G}}|^2 + |\hat{\mathbf{H}} - \hat{\mathbf{H}}_{\mathrm{M}}|^2 \right)$$

Closed-form solution does not exist!

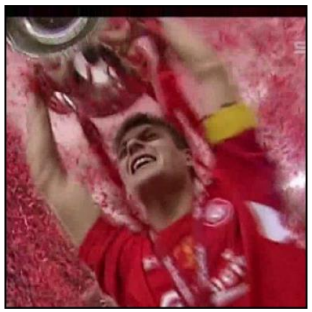
→ Efficient iterative ADMM [1] optimization

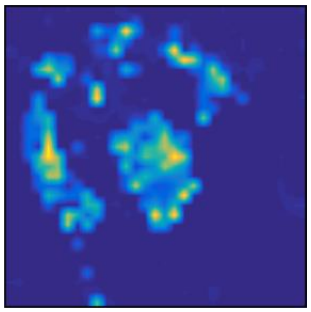
[1] Boyd et al., Distributed optimization and statistical learning via the alternating direction method of multipliers, FTML 2011

Computing Spatial Reliability Map

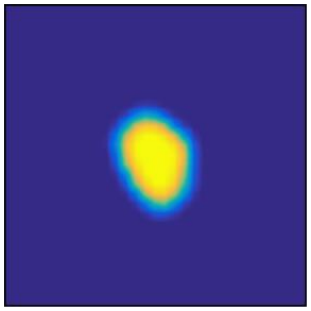
Color likelihood

Training patch

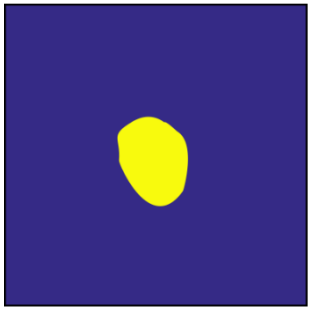




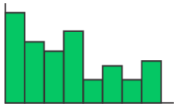
Posterior



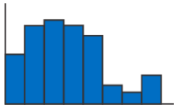
Spatial reliability map



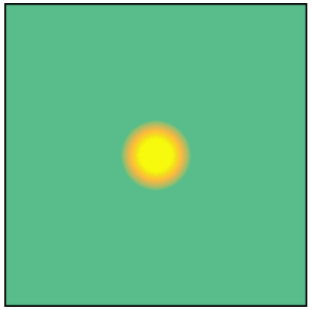
Histograms



fg



bg



Spatial prior

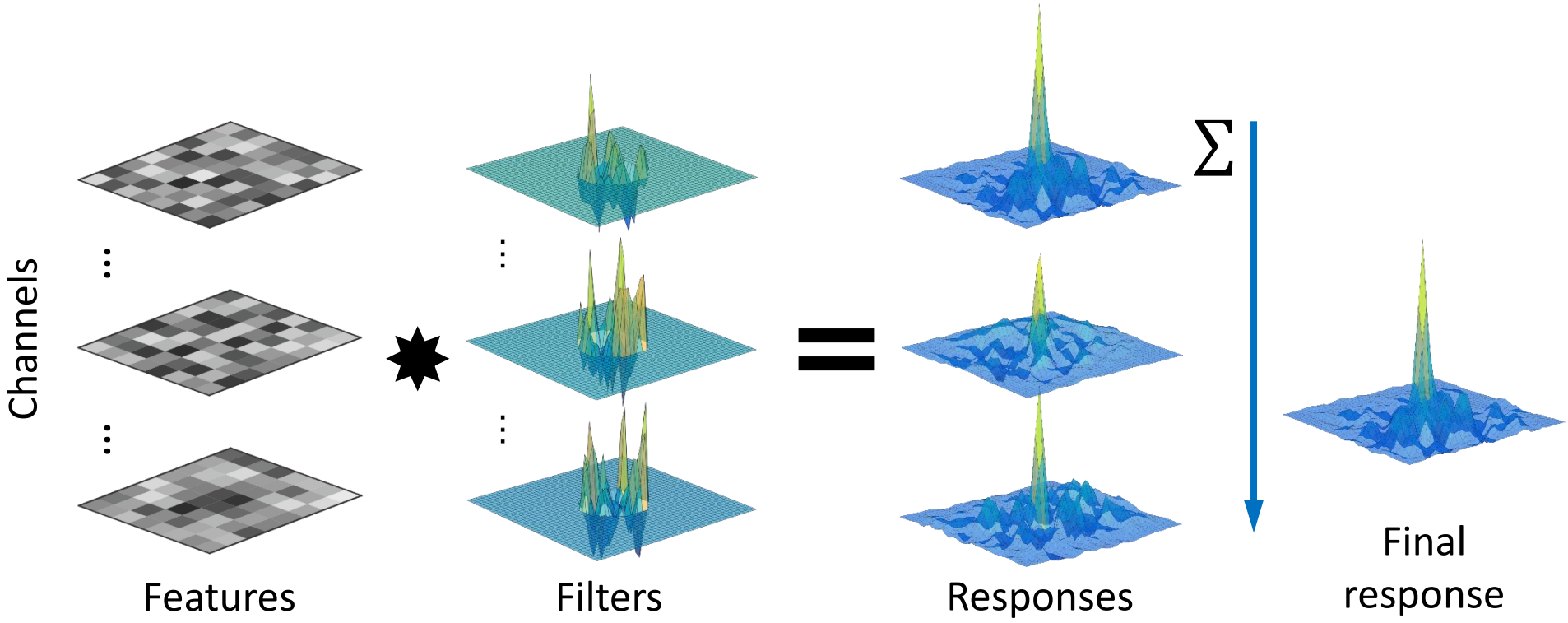
Efficient MRF optimization [1]

Learned during tracking

[1] Kristan et al., Fast image-based obstacle detection from unmanned surface vehicles, TCyB, 2015

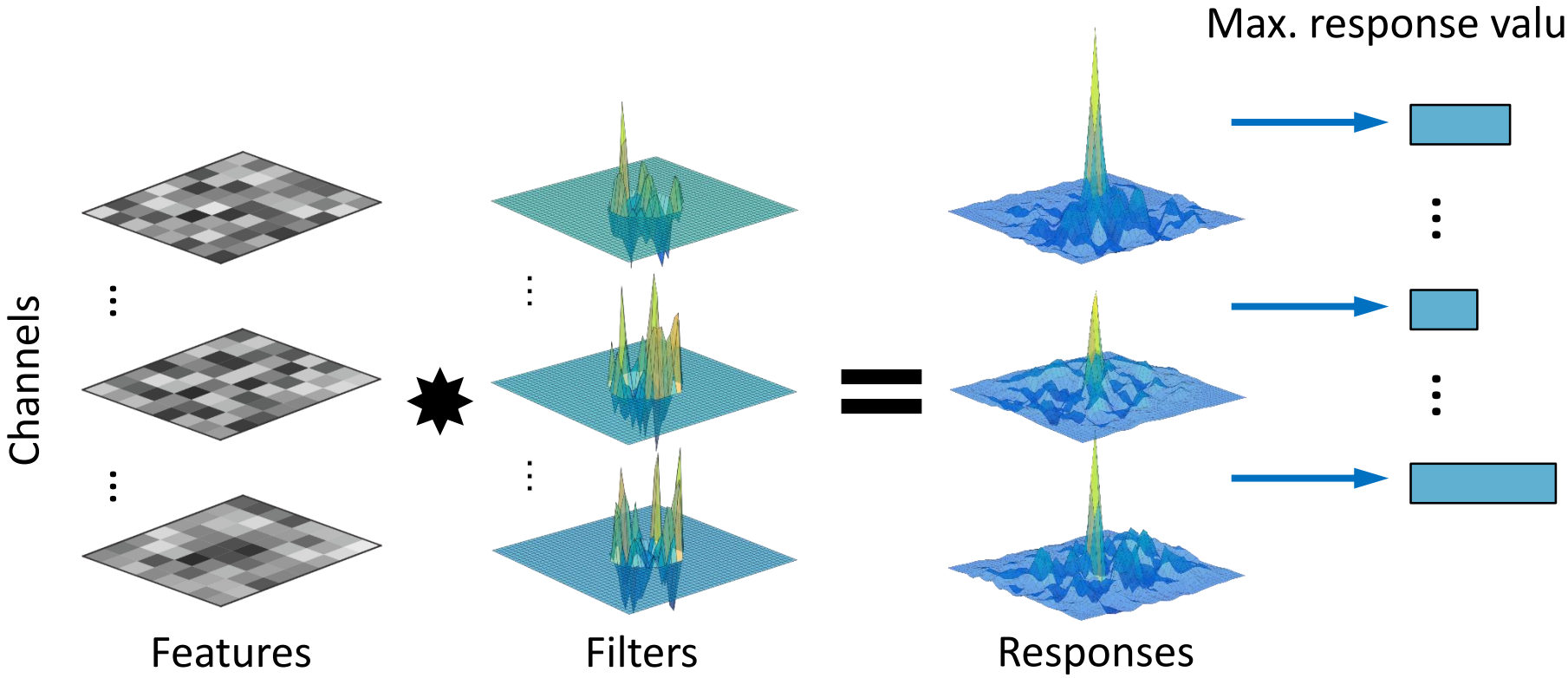
Multiple-Channel Correlation Filters

- Multiple channel representations
  - HoG (31 D), ColorNames (10 D), CNN (~100-1000 D)
- Combined by averaging



Contribution 2: Channel Reliability

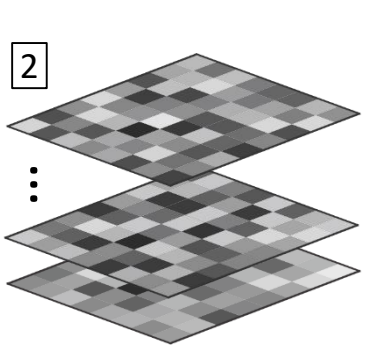
- Estimation in training step
- Observation: channel discriminativity is reflected in height of the response peak



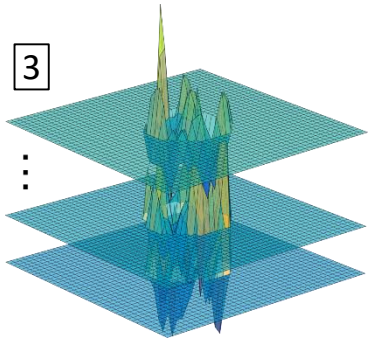
CSR-DCF Tracking Iteration

Position from previous frame

Image patch



Feature channels



Correlation filter

Correlation

EFFICIENT

8 Spatial reliability map estimation (color segmentation)

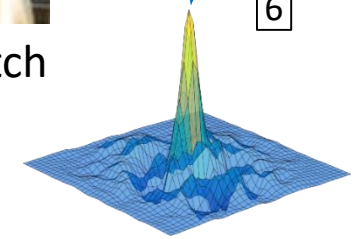
Training patch

EFFICIENT

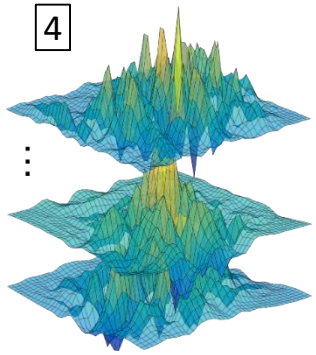
9 Filter learning (Iterative optimization)

10 Channel reliability weights estimation

New object location

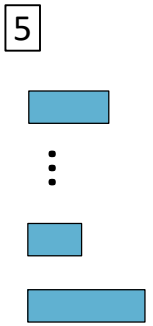


Final response



Response channels

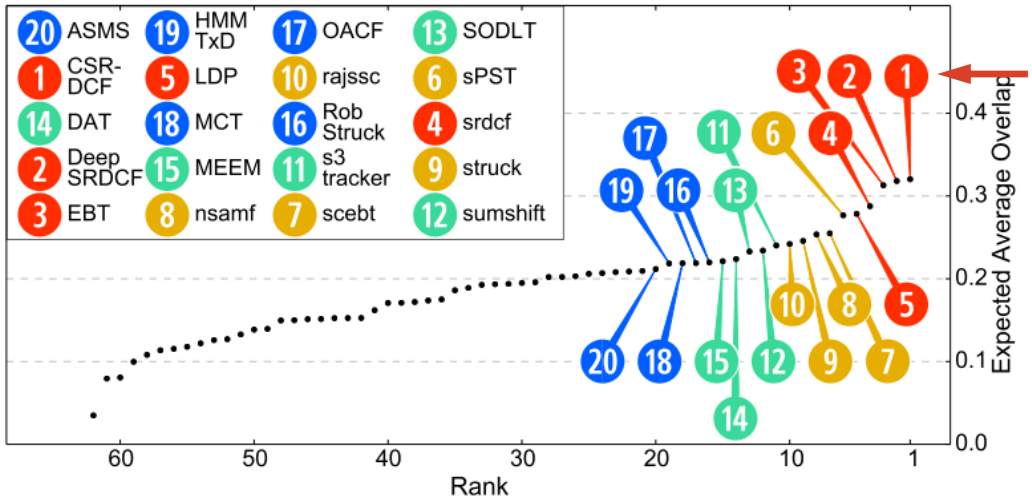
Weighted average



Reliability weights

Results: Expected Average Overlap

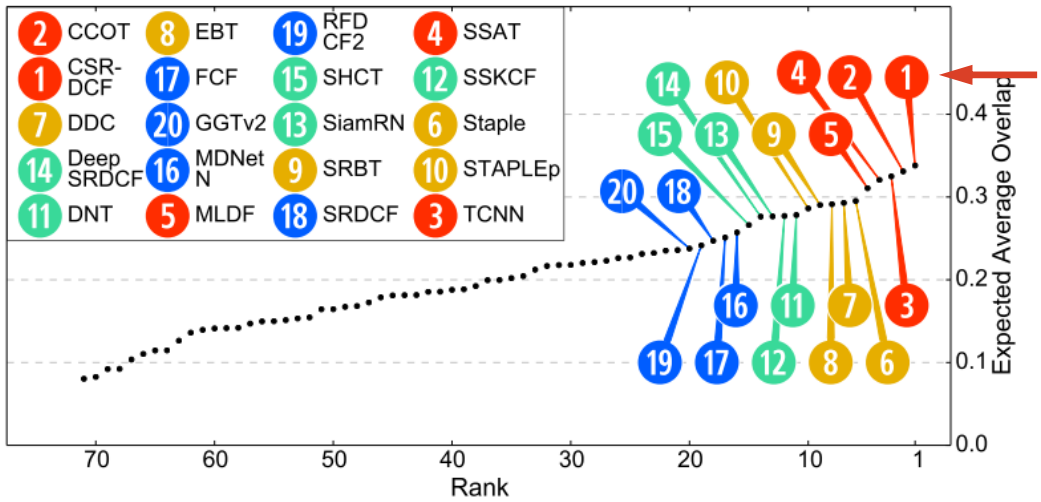
VOT 2015



CSR-DCF [1]

Features used:
HoG + ColorNames

VOT 2016



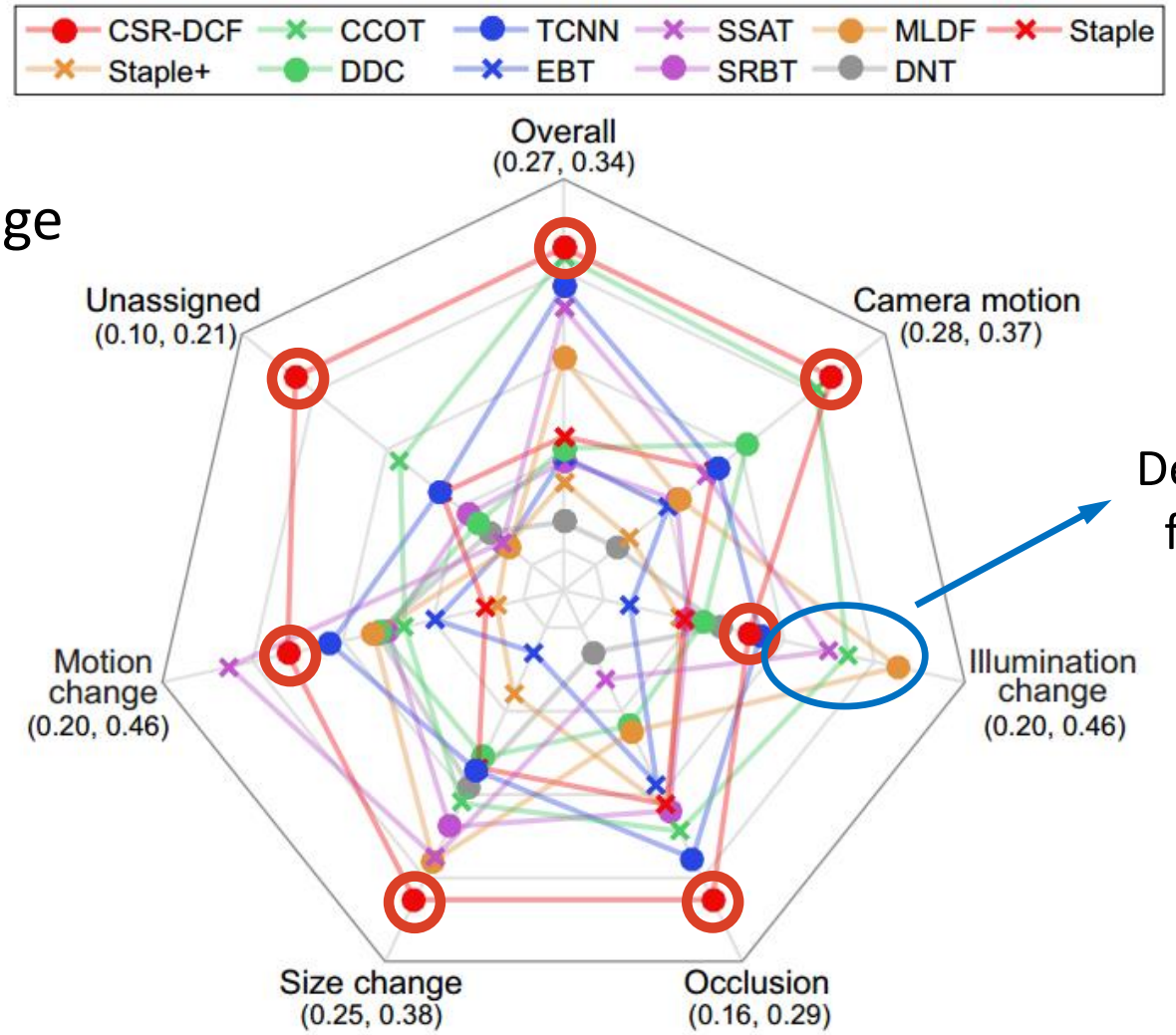
CSR-DCF [1]

[1] A. Lukežič, T. Vojíř, L. Z. Čehovin, J. Matas, M. Kristan. Discriminative Correlation Filter with Channel and Spatial Reliability, CVPR 2017

Visual Attribute Analysis (VOT2016)

Visual attributes:

- Camera motion
- Illumination change
- Occlusion
- Size change
- Motion change



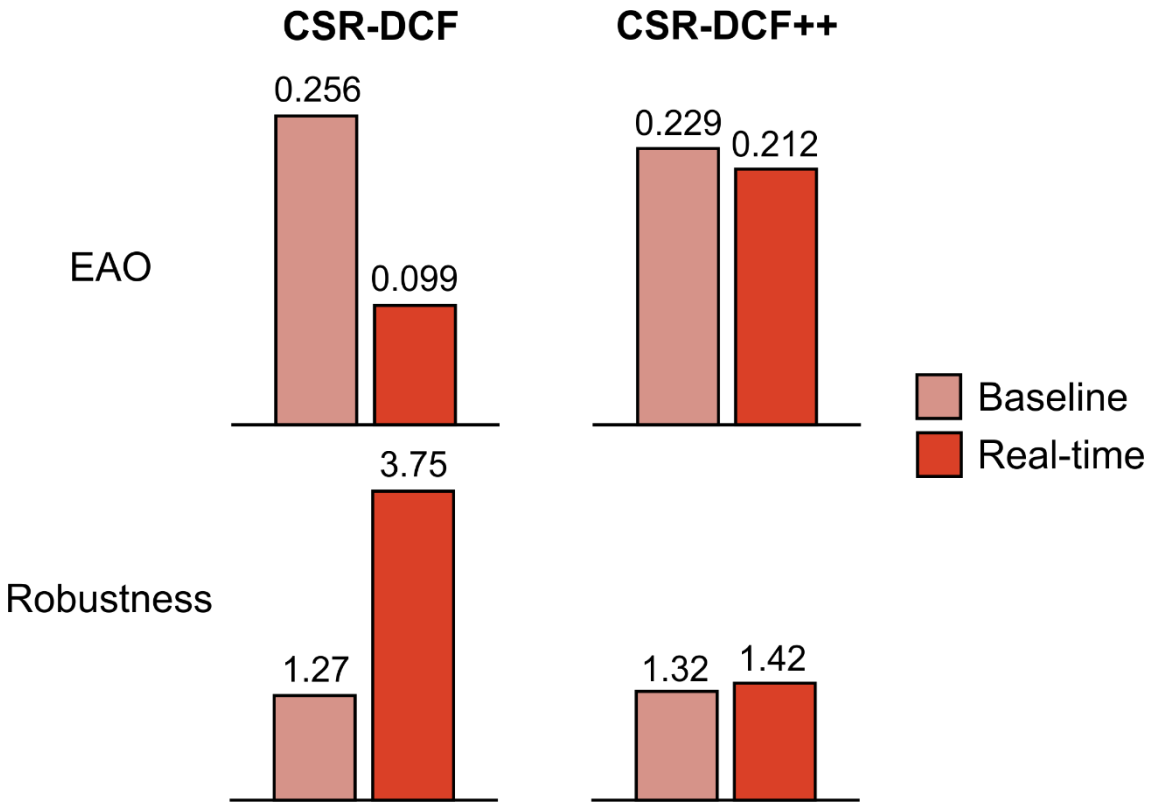
Results on VOT 17: Baseline vs. Realtime

	Baseline			Real time			
Tracker	EAO	A_{av}	R_{av}	EAO	A_{av}	R_{av}	fps
CSR-DCF	0.256	0.49	1.27	0.099	0.48	3.75	13
CSR-DCFf	0.227	0.48	1.37	0.158	0.48	2.30	20
CSR-DCF++	0.229	0.45	1.32	0.212	0.46	1.42	27

- Matlab version (CSR-DCF) fast, but a lot of overhead
  - Image resize, calling mex functions (HoG, segmentation)
- Optimized Matlab version (CSR-DCFf) faster, but still performance drop
- C++ version (CSR-DCF++): real-time speed
  - Almost the same tracking performance

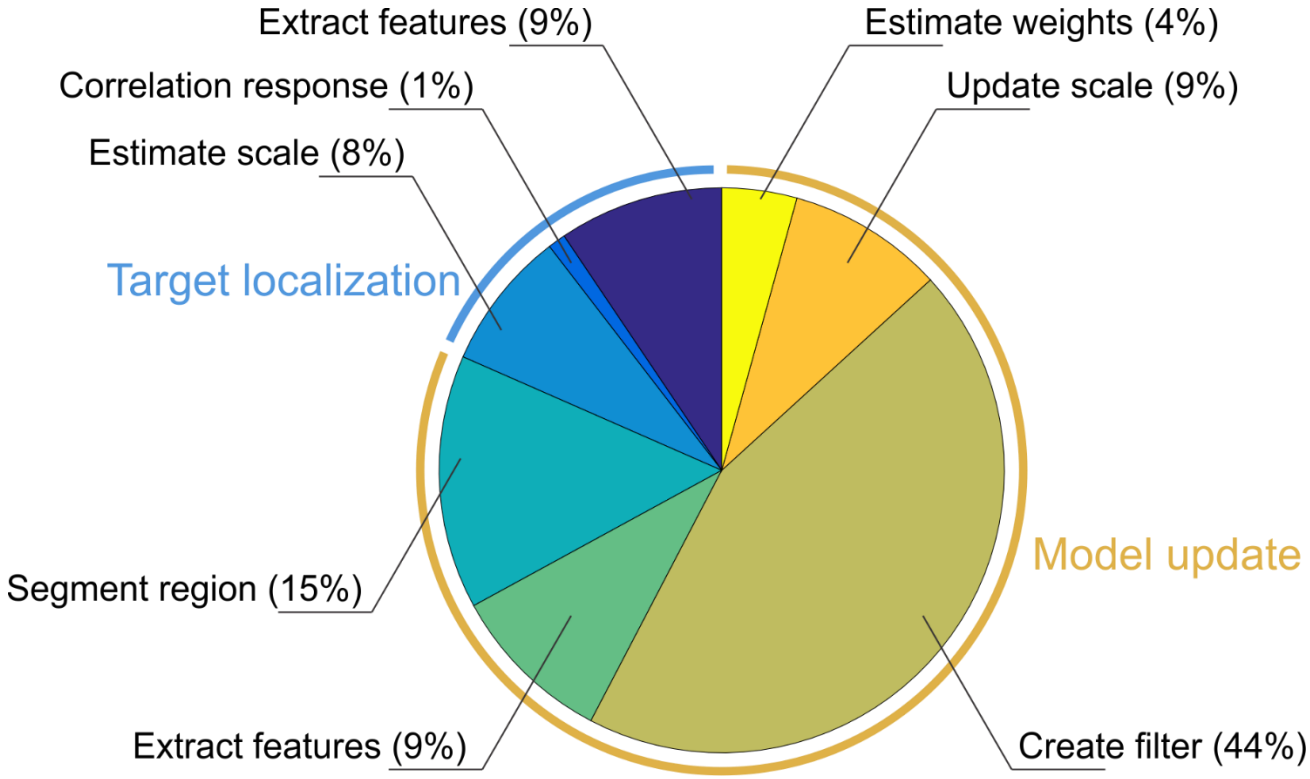
Results on VOT 17: Baseline vs. Realtime

	Baseline			Real time			
Tracker	EAO	A_{av}	R_{av}	EAO	A_{av}	R_{av}	fps
CSR-DCF	0.256	0.49	1.27	0.099	0.48	3.75	13
CSR-DCF++	0.229	0.45	1.32	0.212	0.46	1.42	27



Tracking Speed: C++

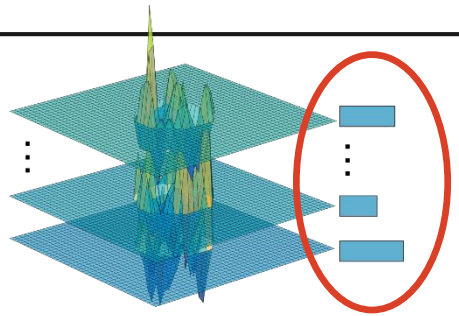
- Average speed on VOT 2016: 27 FPS
  - Desktop PC: CPU Intel i7 (3.40 GHz), 16 GB RAM
  - No GPU needed



Conclusion

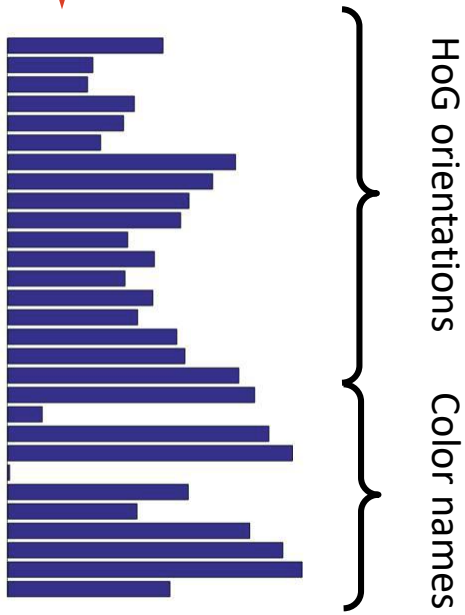
- CSR-DCF contributions in DCF:
  - Target modeling and a method for filter learning with spatial reliability map
  - Introducing channel reliability
- State-of-the-art results on the recent benchmarks
- Real-time tracking performance – C++ (approx. 2-times faster than Matlab verstion)
- Planning to publish it open source (in OpenCV contrib module)

Qualitative evaluation



Octopus sequence

Tracking result



Channel reliability weights

Thank you!

CVPR paper: A. Lukežič, T. Vojíř, L. Z. Čehovin, J. Matas, M. Kristan.
Discriminative Correlation Filter with Channel and Spatial Reliability

Matlab source: https://github.com/alanlukezic/csr-dcf