VOT
visual object tracking

The Visual Object Tracking VOT2016: Challenge and results

Matej Kristan, Aleš Leonardis, Jiri Matas, Michael Felsberg, Roman Pflugfelder, Luka Čehovin, Gustavo Fernandez, Tomaš Vojir, Gustav Hager, Alan Lukežič, et al.

University *of Ljubljana*
Faculty *of Computer and Information Science*

UNIVERSITY OF BIRMINGHAM

LINKÖPING UNIVERSITY

AUSTRIAN INSTITUTE OF TECHNOLOGY

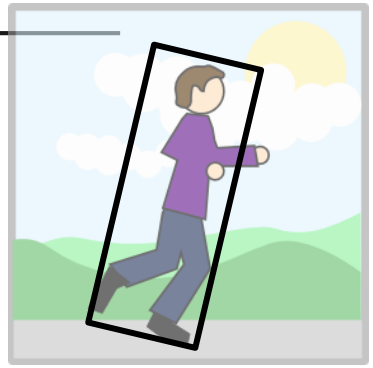
Outline

1. Scope of the VOT challenge
2. VOT2016 challenge overview
   - Evaluation system
   - Dataset
   - Performance evaluation measures
3. VOT2016 results overview
4. Summary and outlook

VOT2016

SCOPE OF THE VOT2016 CHALLENGE

Selected class of trackers



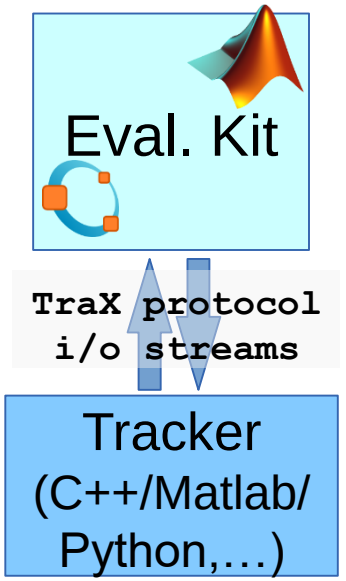
- *Single-object, single-camera, model-free, short-term, causal trackers*
- Model-free:
  - Nothing but a single training example is provided by the BBox in the first frame
- Short-term:
  - Tracker does not perform re-detection
  - Once it drifts off the target we consider that a failure
- Causality:
  - Tracker does not use any future frames for pose estimation
- Object state defined as a rotated bounding box (rectangle)

VOT2016

VOT2016 EVALUATION SYSTEM

VOT2016 Challenge evaluation kit

- Matlab-based kit to automatically perform a battery of standard experiments
- Plug and play!
  - Supports multiple platforms and programming languages (C/C++/Matlab/Python, etc.)



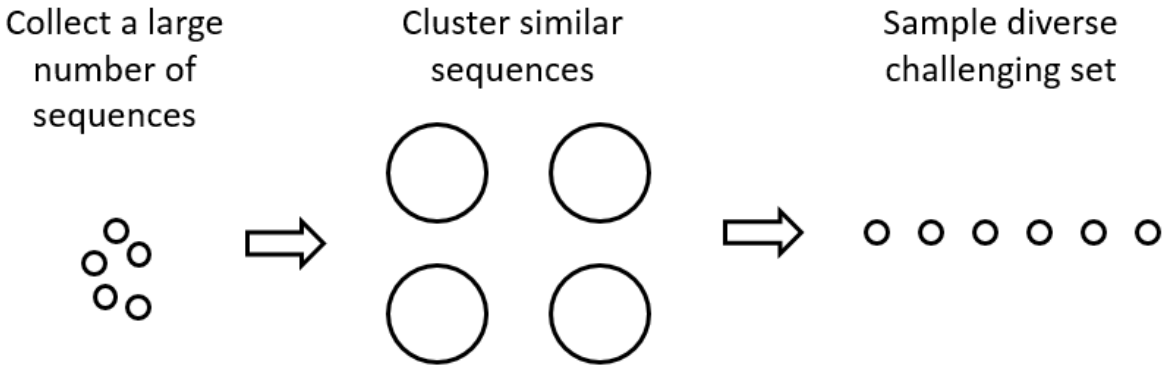
- Easy to evaluate your tracker on all our benchmarks
- Backward compatibility with VOT2013/VOT2014/VOT2015

- Download from our homepage
  https://github.com/vicoslab/vot-toolkit

VOT2016 DATASET

Dataset construction approach

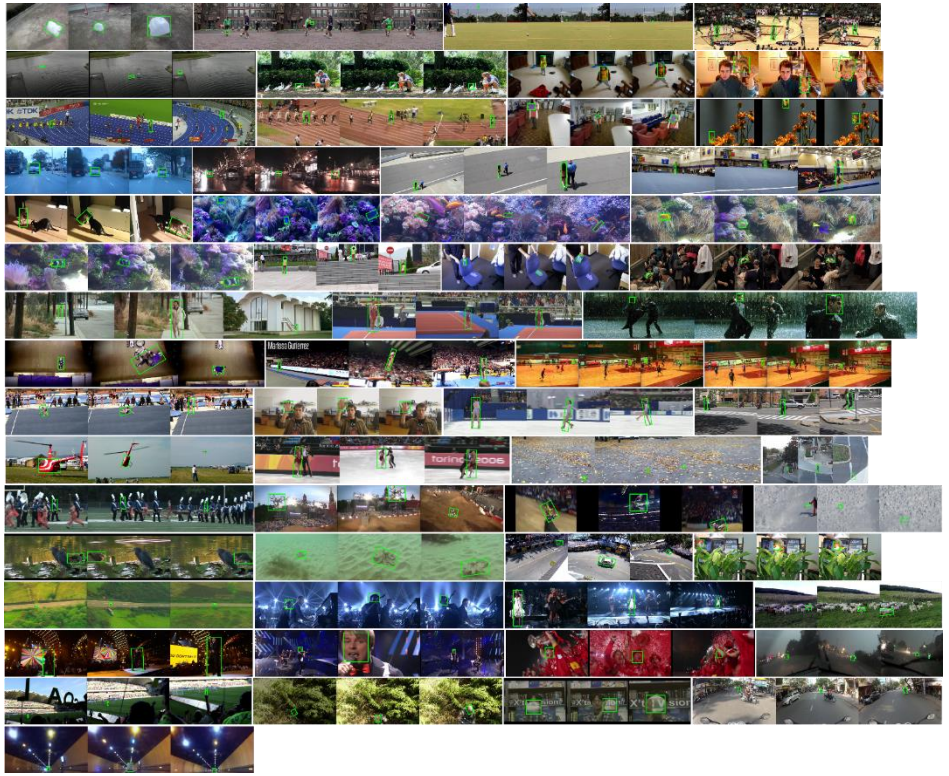
- Current trend [Wu et al. CVPR2013, Smeulders et al. PAMI2013, Wang et al. arXiv2015, Wu et al. PAMI2015]:
  - Large datasets by collecting many sequences from internet
  - Large dataset ≠ diverse or useful
- VOT2013/2014/2015 approach:
  - Keep it sufficiently small, well annotated and diverse
  - Developed the VOT dataset construction methodology



The VOT2016 dataset

- The performance on VOT2015 dataset did not saturate in 2015 challenge
- Kept all 60 sequences from VOT2015 challenge
- NEW:

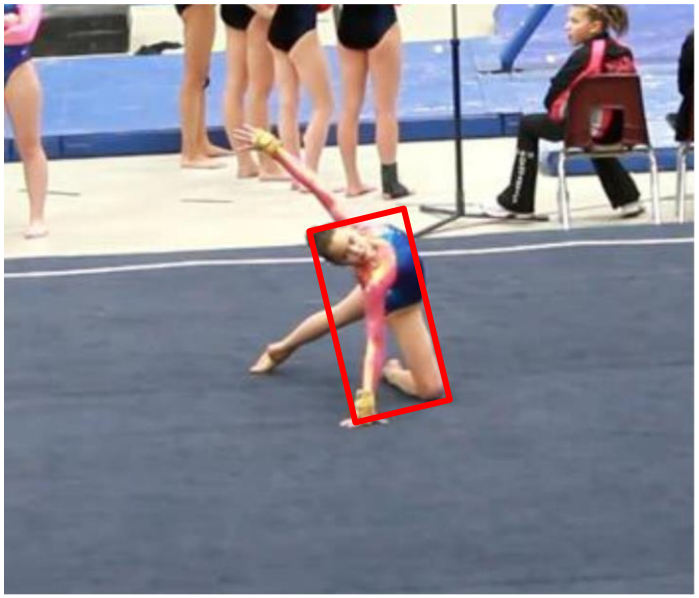
  Objects re-annotated!

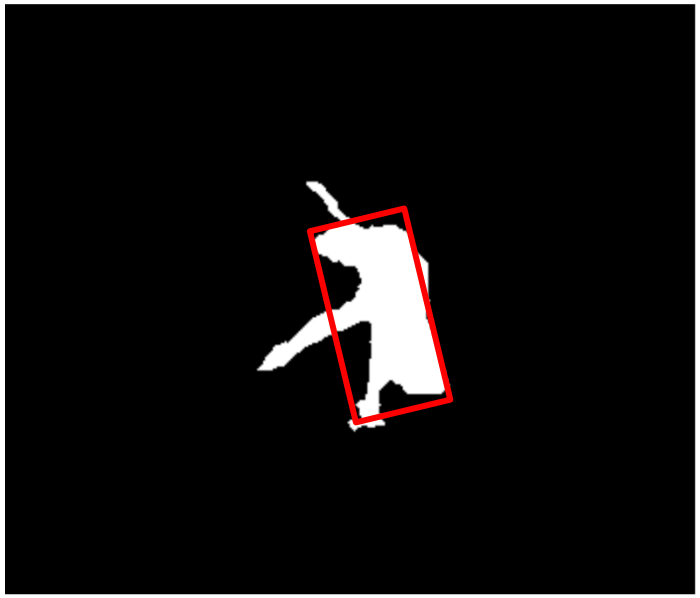


Object annotation

Automatic bounding box placement

1. Segment the target (semi-automatic)
2. Automatically fit a bounding box by optimizing a cost function

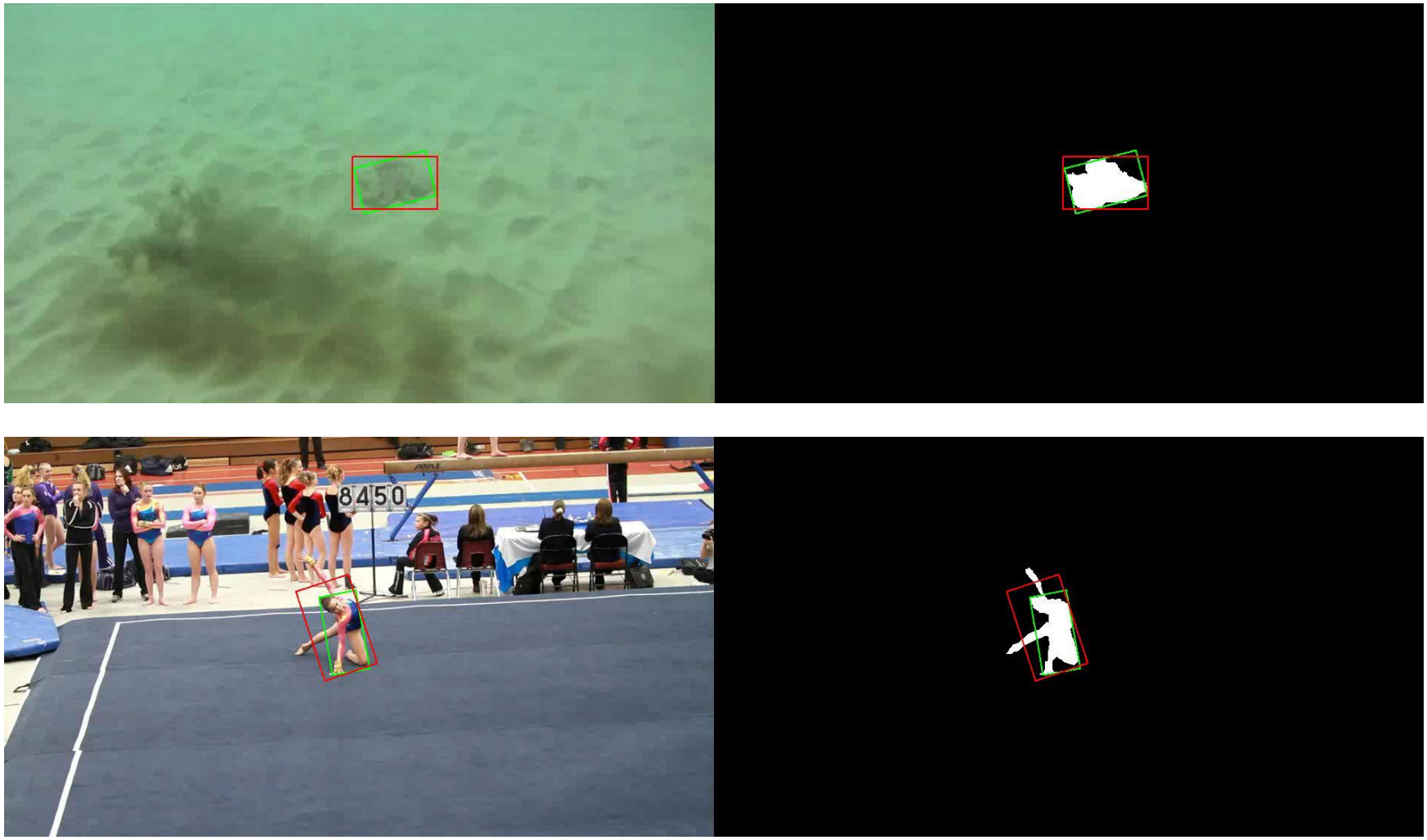




- Visual verification of the results
  - 12% reverted to the VOT2015 annotation

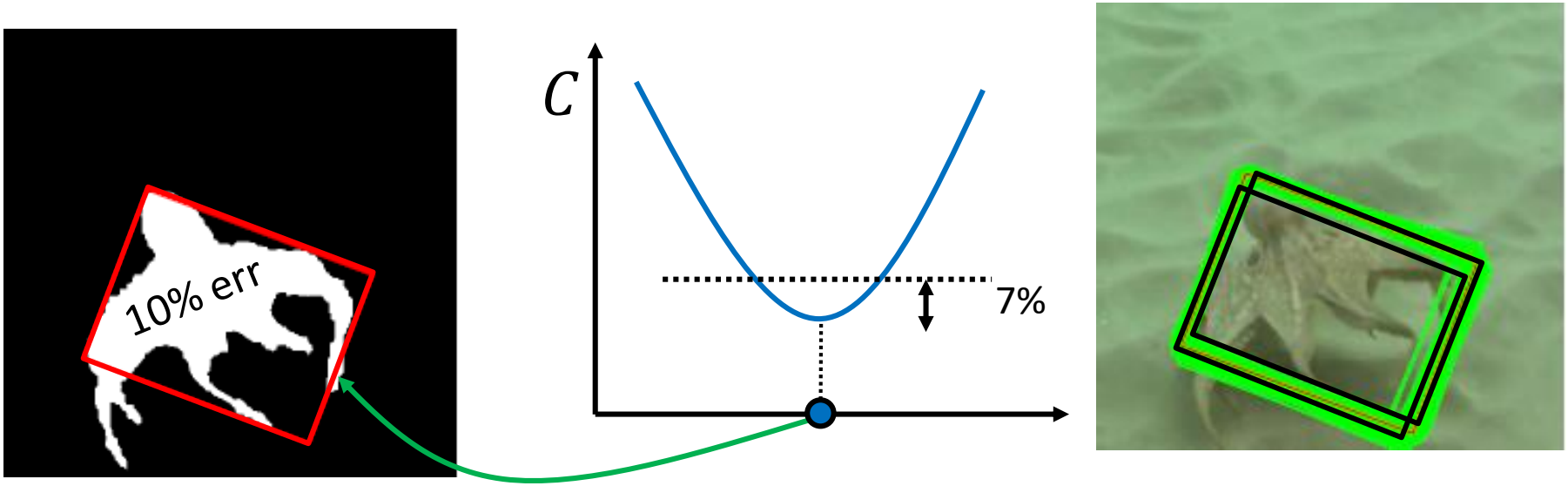
VOT2016 dataset – object annotation

- Average overlap between VOT2015 and VOT2016 BB: 0.74



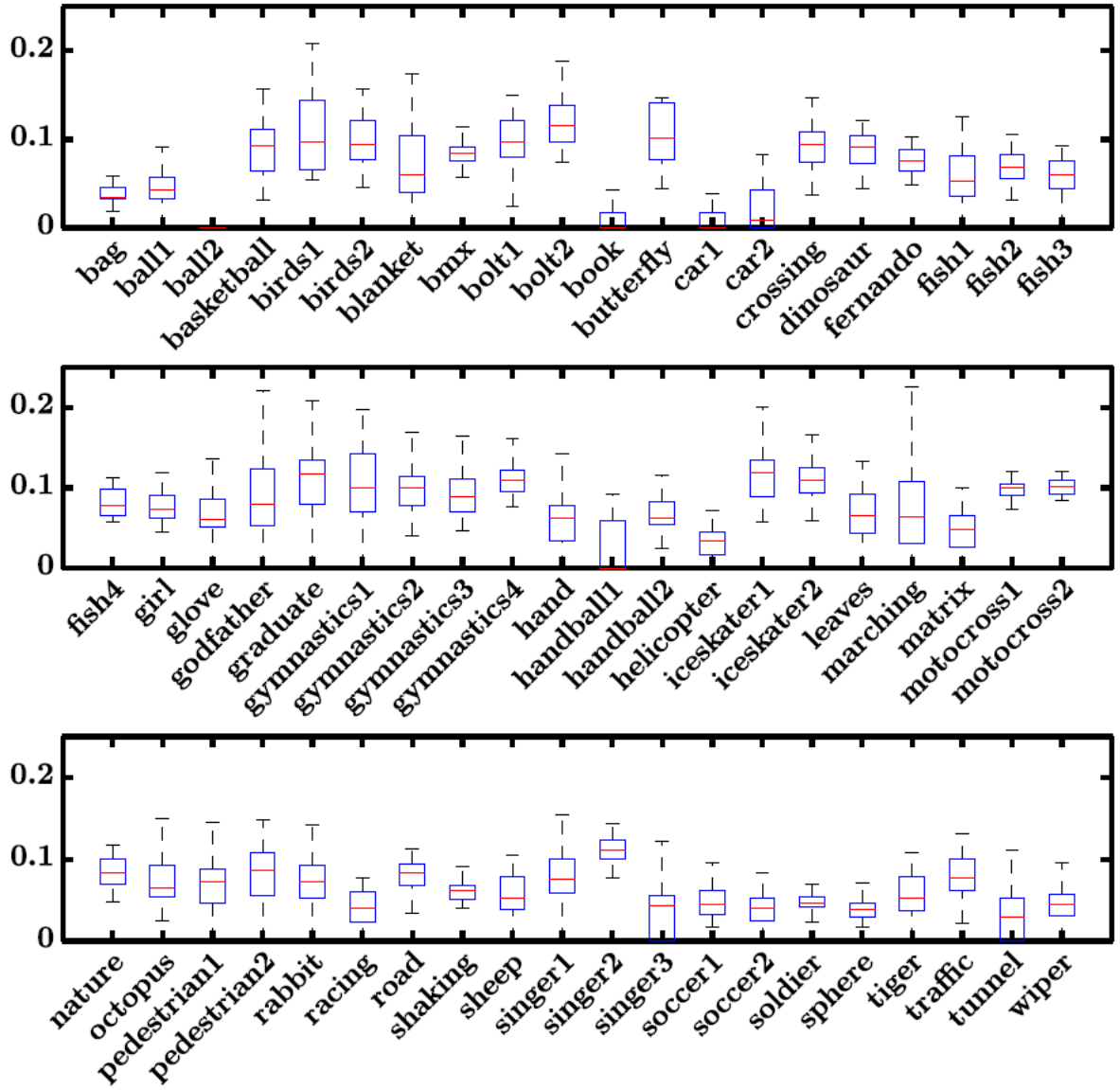
Annotation uncertainty

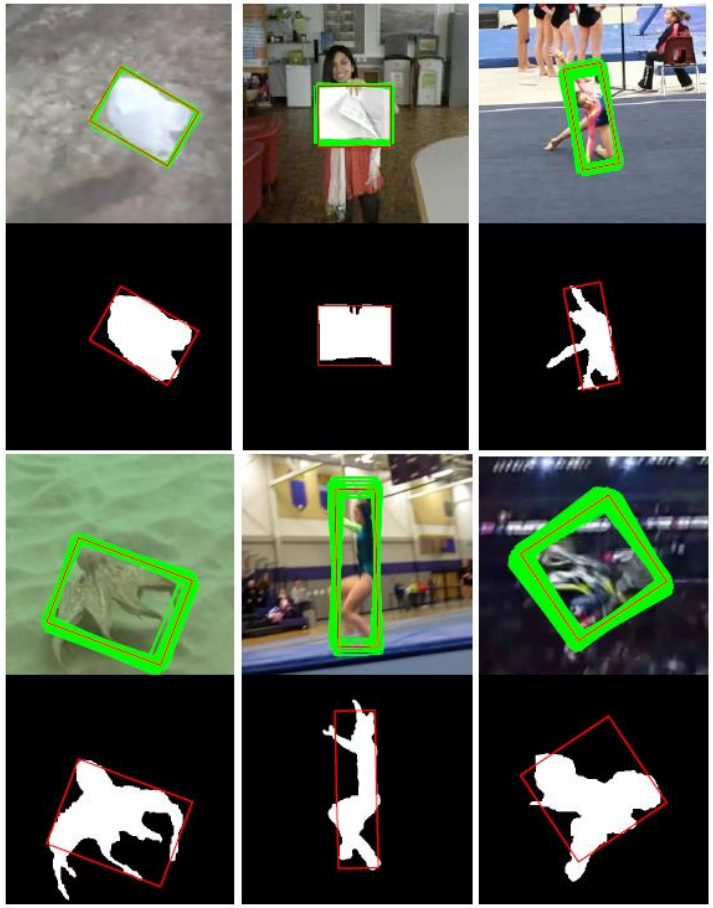
- Segmentation uncertainty results in bounding box uncertainty



- Uncertainty: Average of overlaps between optimal bounding box and those within 7% C increase.

Practical differences





Reduced by half compared to VOT2015

VOT2016 dataset – frame annotation

- Manually and automatically labeled each frame with VOT2013 visual attributes (same as VOT2015):
  - i. Occlusion (M)
  - ii. Illumination change (M)
  - iii. Object motion (A)
  - iv. Object size change (A)
  - v. Camera motion (M)
  - vi. Unassigned (A)

M ... manual annotation, A ... automatic annotation

(i)	0	1	1	0
(ii)	0	0	0	0
(iii)	0	0	0	0
(iv)	1	1	1	0
(v)	0	0	0	0
(vi)	0	0	0	1

VOT2016

EVALUATION METHODOLOGY

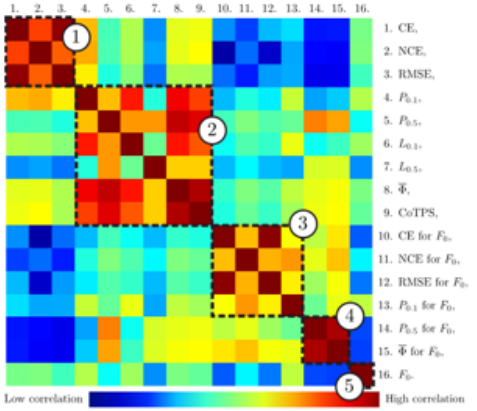
Performance measures

- Target localization properties measured using the VOT2013/VOT2014/VOT2015 methodology.

- Approach in VOT2013/VOT2014:
  - Interpretability of performance measures
  - Select as few as possible to provide clear comparison

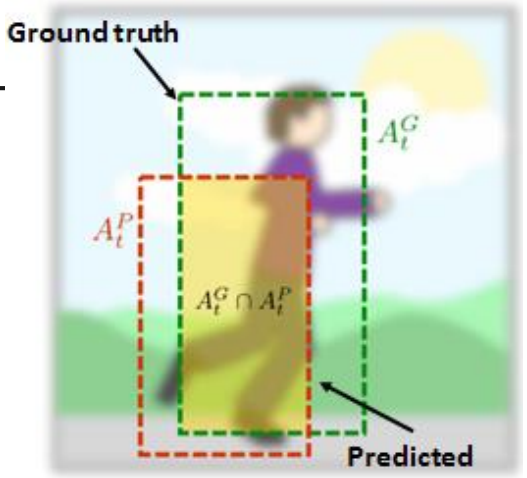
- Based on a recent study[1] two basic weakly-correlated measures are chosen:
  - Robustness
  - Accuracy

Correlation analysis of performance measures[1]



[1]Čehovin, Leonardis, Kristan. *Visual object tracking performance measures revisited*, IEEETIP 2016

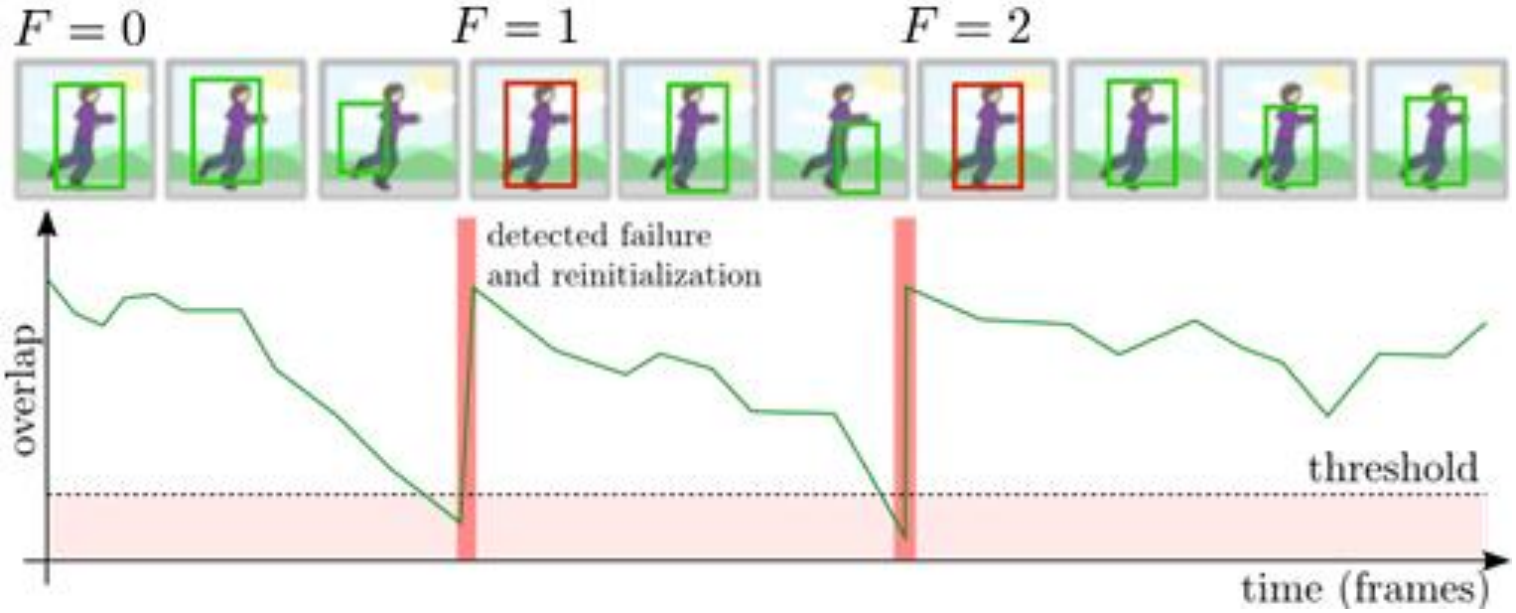
VOT performance measures



- Robustness:

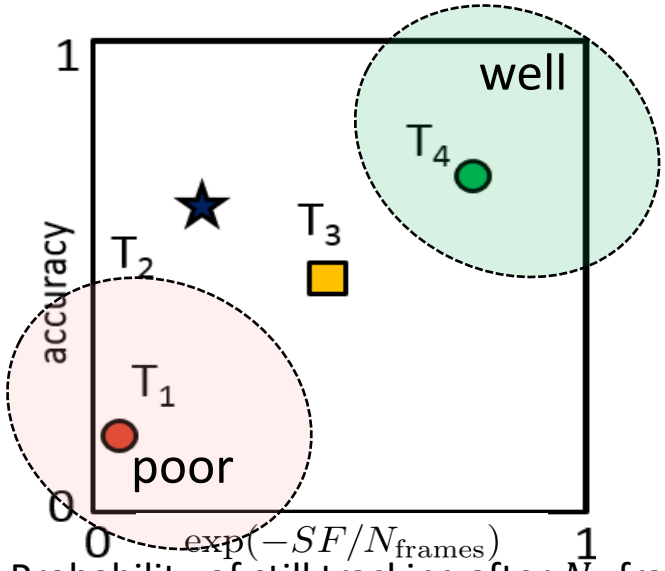
  Number of times a tracker drifts off the target.

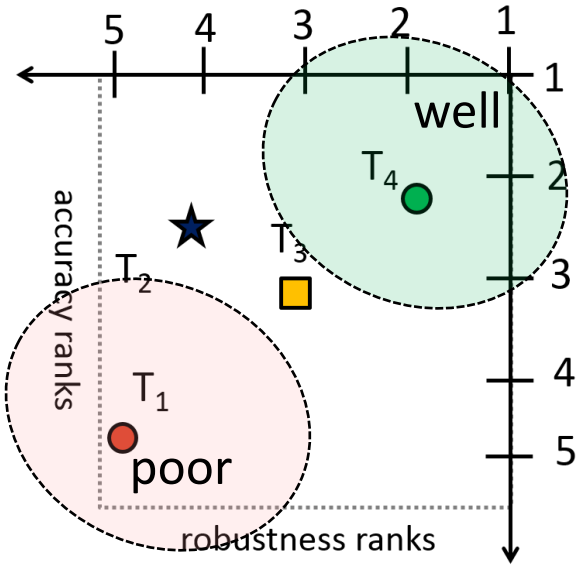
- Accuracy: *Average overlap during successful tracking.*



VOT performance evaluation

- Ranking methodology w.r.t. Accuracy and Robustness
- Assign equal rank to “equally” performing trackers:
  - Statistical significance of results and practical difference

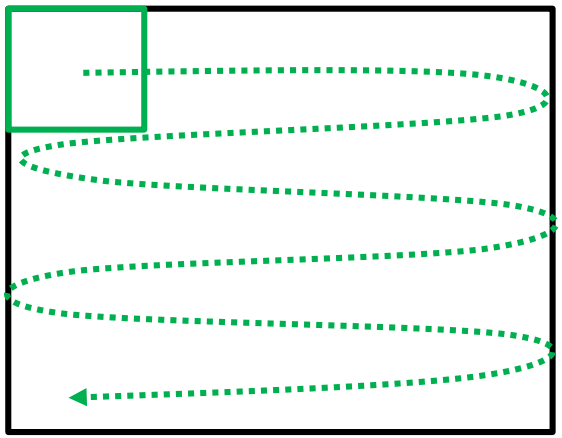




- A principled way to merge Accuracy and Robustness:
  - Expected average overlap (EAO)

VOT2016 Speed measurement

- Reduce the hardware bias in reporting tracking speed.
- Approach: The VOT2014 speed benchmark



600x600 image
Max operation in 30x30 window
Apply this filter to all pixels
Measure the time for filtering

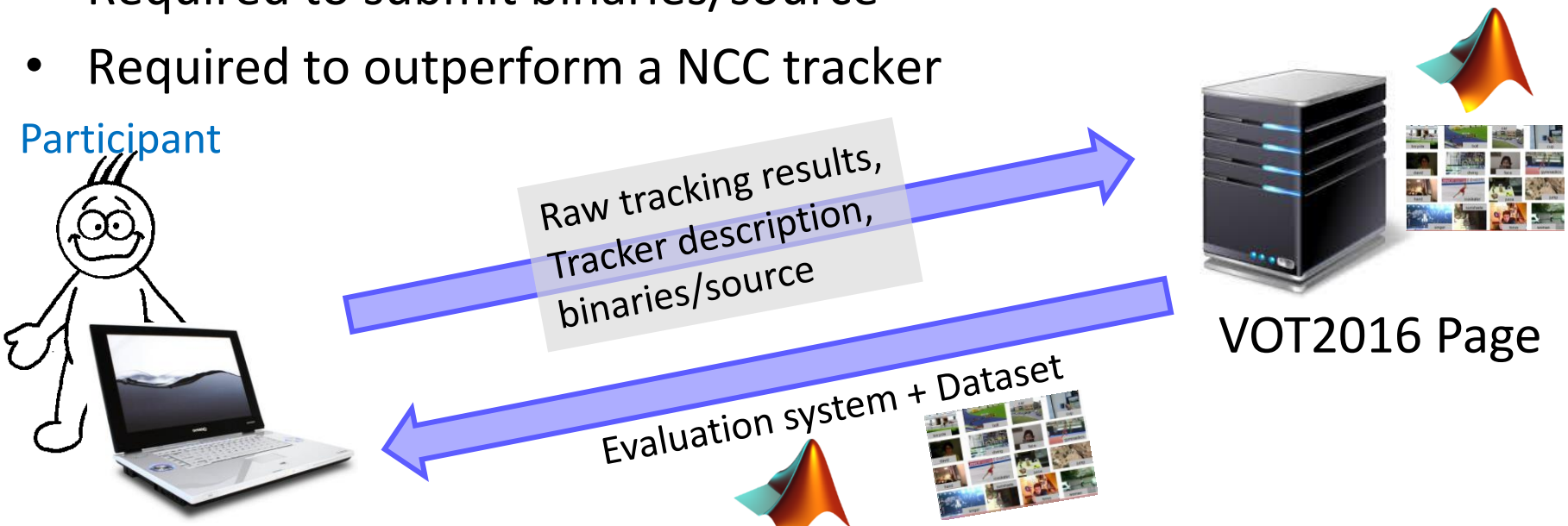
- Divide tracking time with time required to perform the filtering operation
- Equivalent Filter Operations (EFO)

VOT2016

CHALLENGE PARTICIPATION AND SUBMITTED TRACKERS

VOT2016 Challenge: participation

- Participants would download the evaluation kit:
  - Evaluation system + Dataset
- Integrate their tracker into the evaluation system
- Predefined set of experiments automatically performed – submit the results back
- Required to submit binaries/source
- Required to outperform a NCC tracker



70 trackers tested!

Diverse set of entries: 70 = 48 submissions + 22 existing

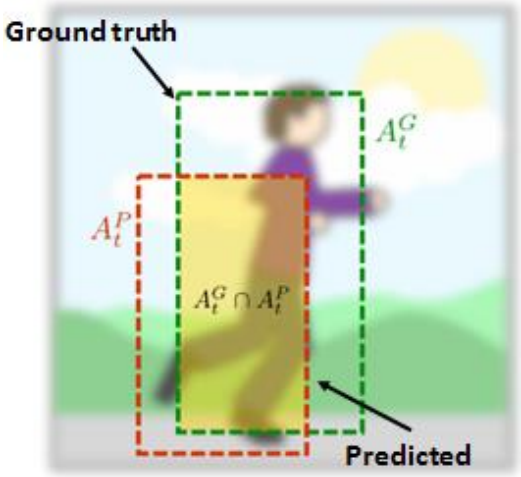
- Deep convolutional neural networks
  (MLDF, SiamFC-R, SiamFC-A, TCNN, DNT, SO-DLT, MDNet-N, SSAT)
- Correlation filters
  (SRDCF, SWCF, FCF, GCF, ART-DSST, DSST2014, SMACF, STC, DFCT, KCF2014, SAMF2014, OEST, sKCF, Staple, Staple+, MvCFT, NSAMF, SSKCF, ACT, ColorKCF, deepMKCF, HCF, DDC, DeepSRDCF, C-COT, RFD-CF2, NCC)
- Discriminative models – single part
  (MIL, Struck2011, EBT, TGPR)
- Global generative-model-based
  (DAT, SRBT, ASMS, LoFT-Lite, IVT, CCCT, DFT)
- Part-based trackers
  (LT-FLO, SHCT, GGTv2, MatFlow, Matrioska, CDTT, BST, TRIC-track, DPT, SMPR, CMT, HT, LGT, ANT, FoT, FCT, FT, BDF)
- Combinations of multiple trackers
  (PKLTF, MAD, CTF, SCT, HMMTxD)

VOT2016

EXPERIMENTS AND RESULTS

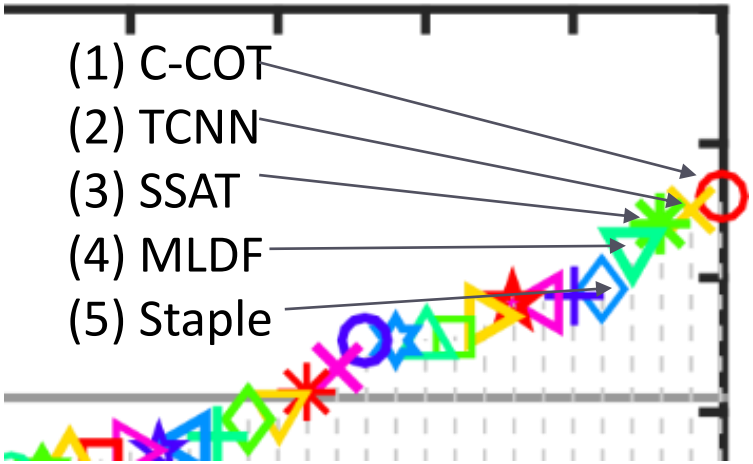
VOT2016 Experiment

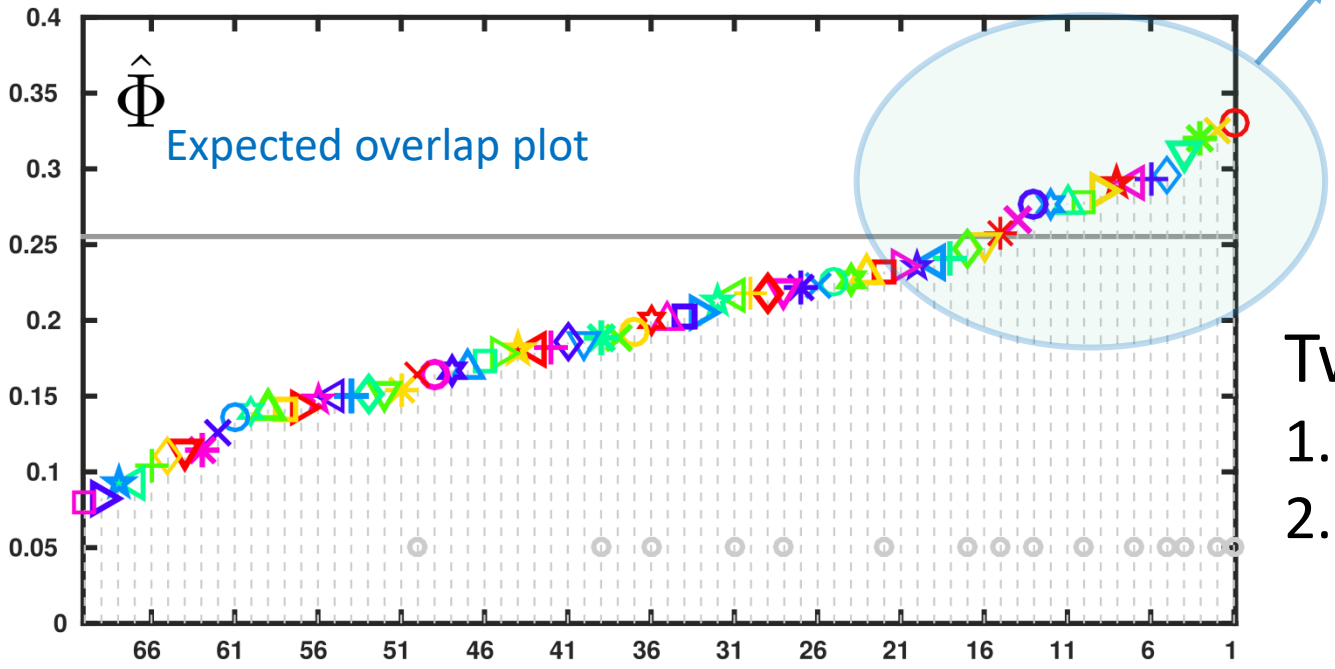
- Initialization on ground truth BBs
- Each tracker run 15 times on each sequence to obtain a better statistic on its performance.
- Reinitialization at overlap 0.



Expected average overlap

Tracker	Type
C-COT	Corr. Filter + CNN feats
TCNN	Multiple parallel CNNs
SSAT	CNN (extension of VOT2015 winner).
MLDF	CNN for position + CNN for scale



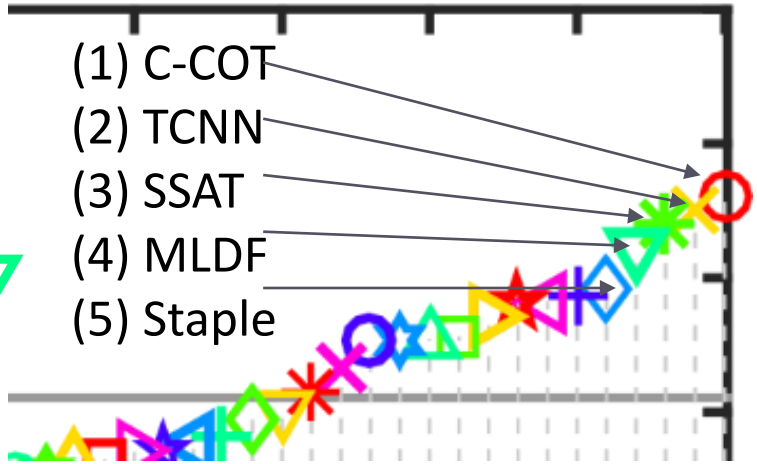


Two classes:

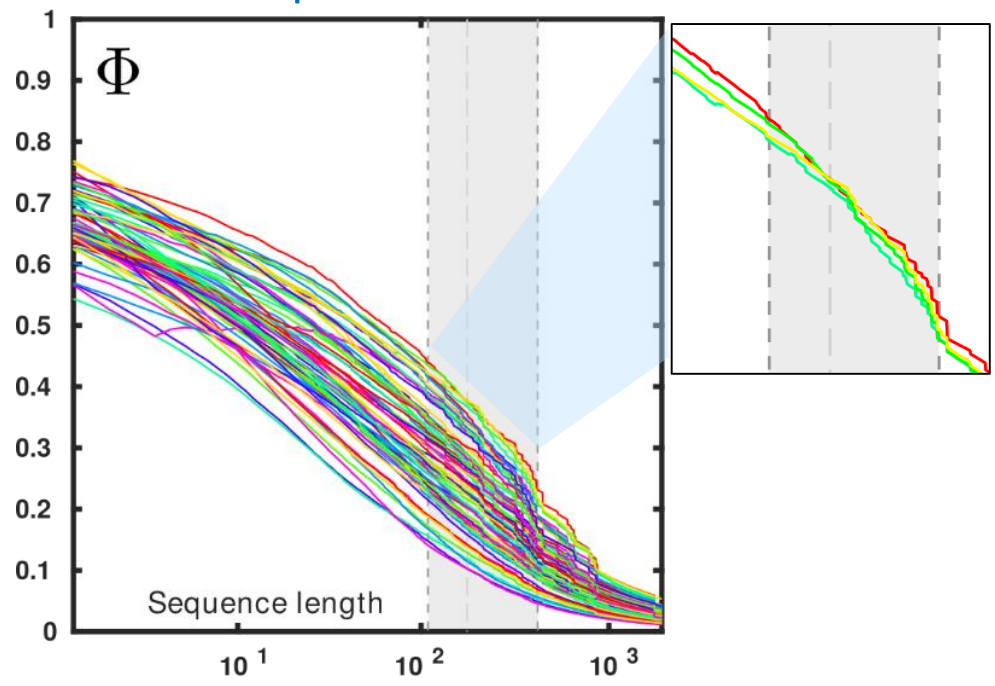
1. CNN-based
2. Correlation filters

Detailed analysis

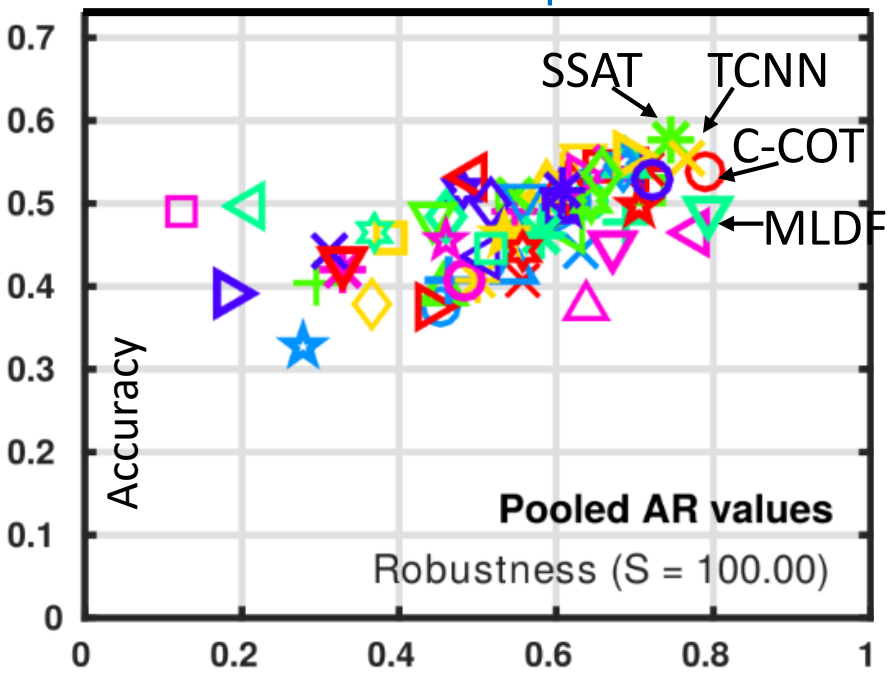
- C-COT slightly ahead of TCNN
- Most accurate: SSAT
- Most robust: C-COT and MLDF



Overlap curves



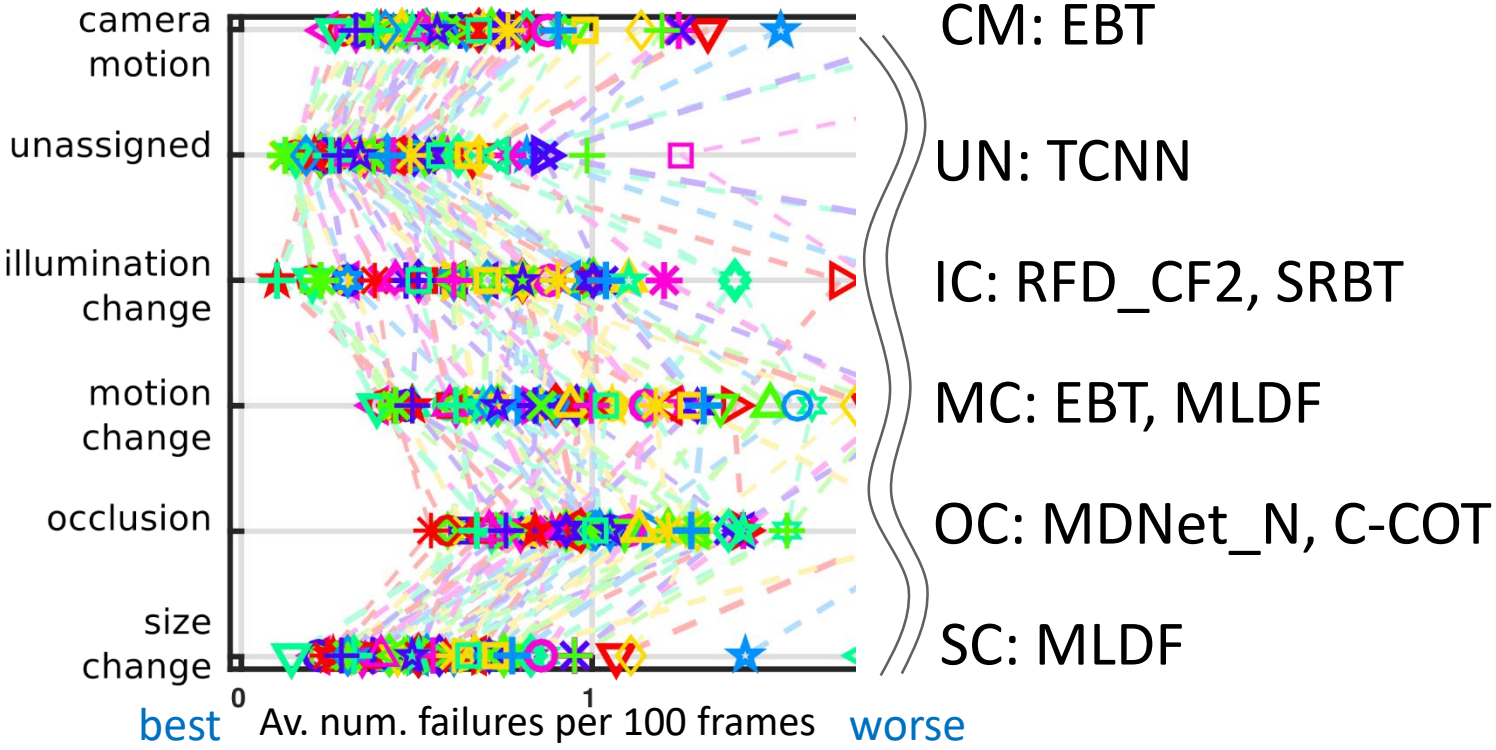
AR-raw plot



Detailed analysis: attributes

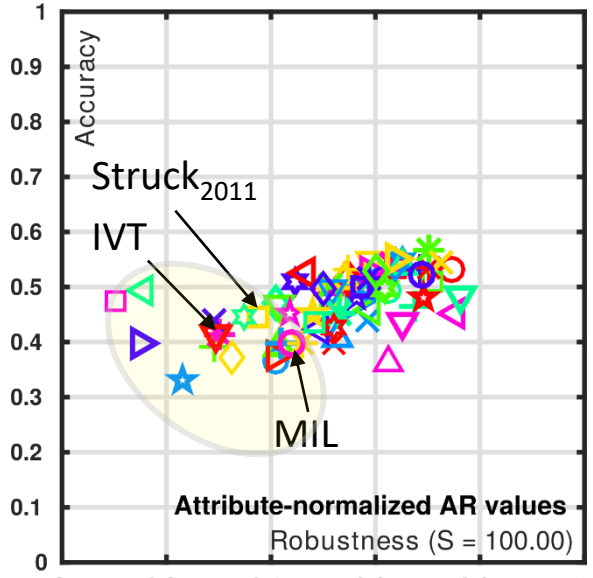
- Top EAO trackers mostly at top per attributes

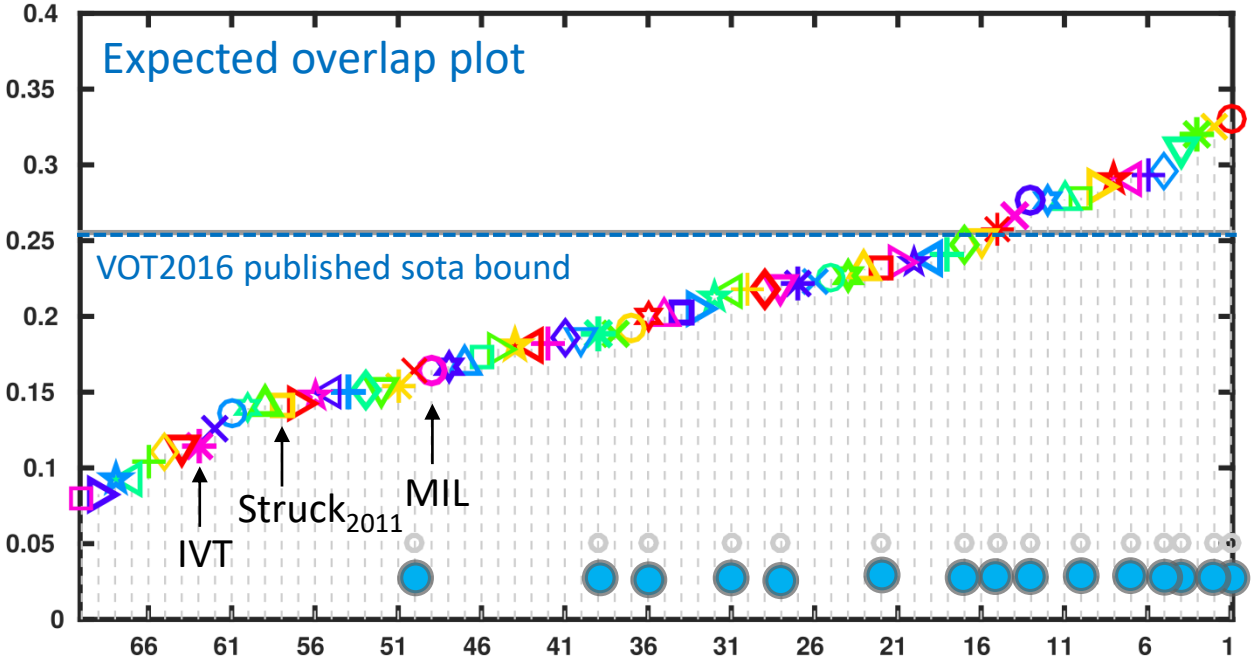
	cam. mot.	ill. ch.	mot. ch.	occl.	scal. ch.
Accuracy	0.49	0.53	0.44	**0.41**	*0.42*
Robustness	0.71	0.81	*1.02*	**1.11**	0.61

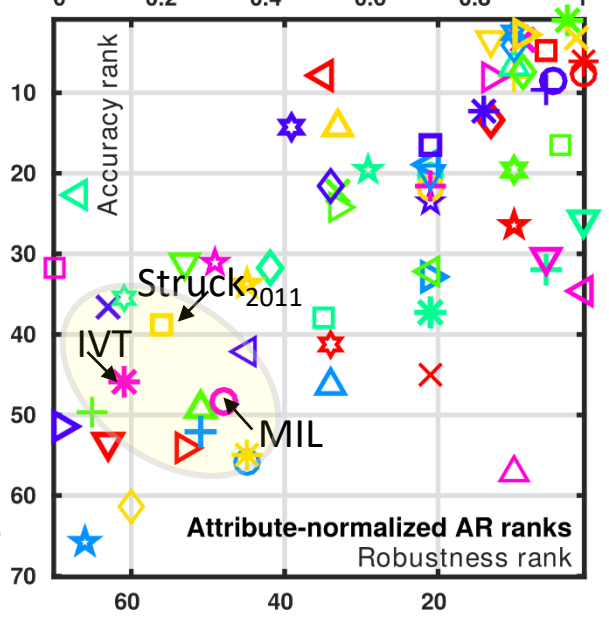


Detailed analysis: baselines + sota

- Baselines: IVT, MIL, Struck
- 15 trackers: (2015-2016) ICCV,ECCV,CVPR,PAMI…
  - Over 20% of tracker exceed their average EAO
- VOT2015: This value was over 40%

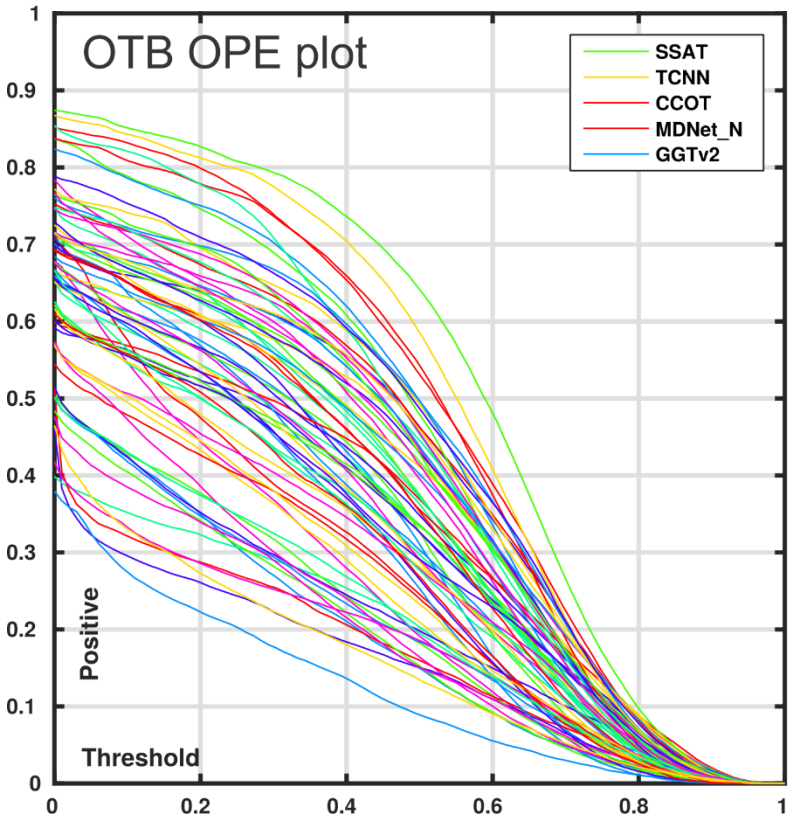


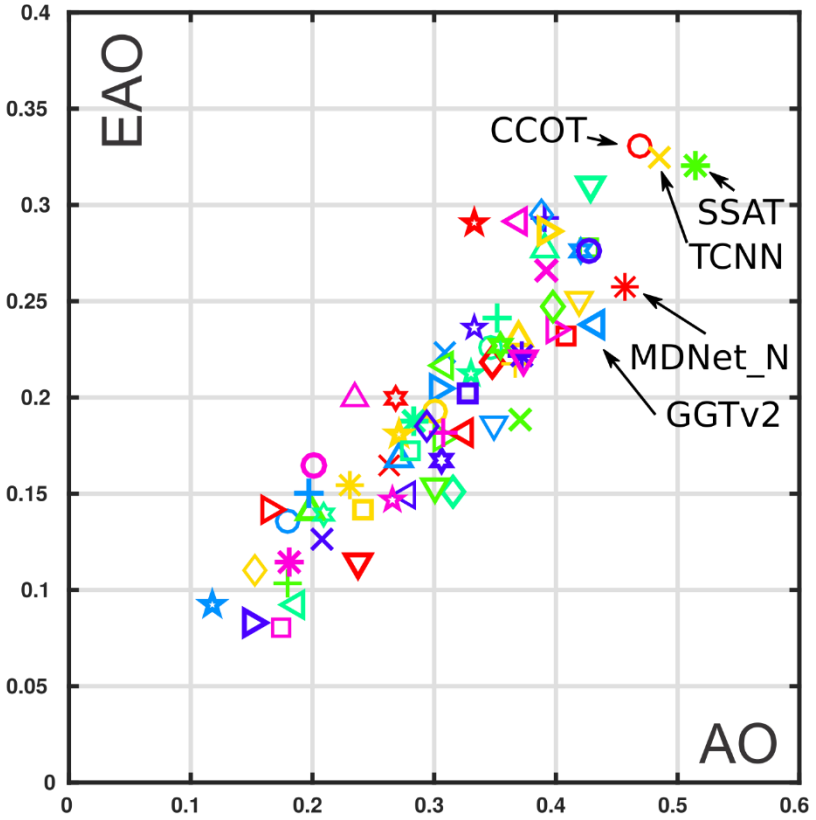




VOT unsupervised experiment

- OTB [Wu et al CVPR2013]: No reset at fail

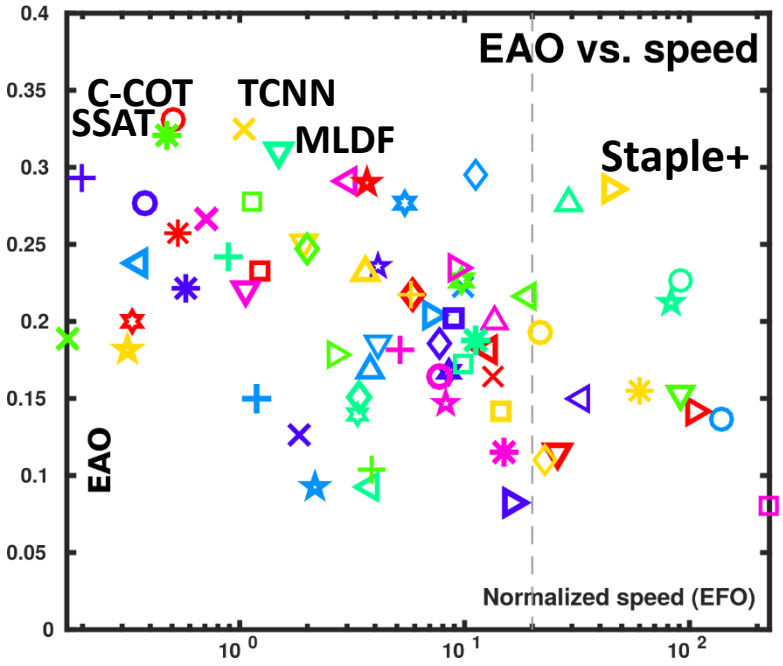


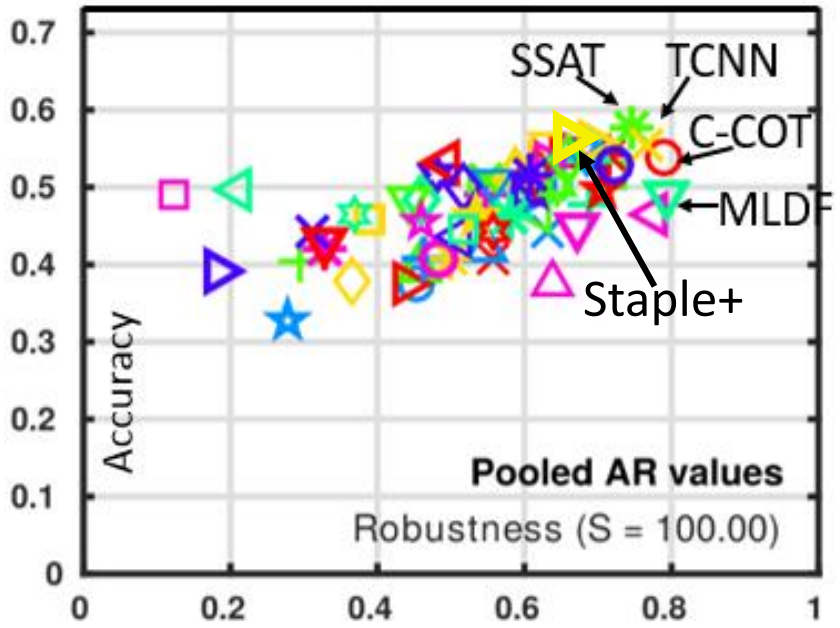


Tracking speed

- Top-performers slowest
  - Plausible cause: CNN
- Real-time bound: Staple+
  - Decent accuracy,
  - Decent robustness

Note: the speed in some Matlab trackers has been significantly underestimated by the toolkit since it was measuring also the Matlab restart time. The EFOs of Matlab trackers are in fact higher than stated in this figure.





Sequence ranking

- VOT2013 approach
  - Average number of trackers failed per frame (A_f)
  - Max. number of trackers failed at a single frame (M_f)

Sequence	Sequence	Sequence	Sequence
Leaves	Fish1	Crossing	Pedestrian2
Soccer2	Nature	Dinosaur	Fish4
Book	Handball2	Iceskater2	Godfather
Matrix	Fish2	Singer2	Bmx
Glove	Ball1	Blanket	Road
Ball2	Tiger	Bolt2	Sheep
Hand	Singer3	Iceskater1	Fish3
Pedestrian1	Gymnastics1	Gymnastics4	Birds2
Gymnastics3	Motocross2	Marching	Tunnel
Butterfly	Handball1	Wiper	Octopus
Rabbit	Soccer1	Helicopter	Singer1
Car1	Graduate	Sphere	Bag
Motocross1	Soldier	Basketball	Racing
Birds1	Bolt1	Shaking	Pedestrian2
Gymnastics2	Fernando	Traffic	Fish4

Challenging:

$A_f \sim [0.19, 0.41]$
$M_f \sim [56, 65]$

$A_f \sim [0.15, 0.17]$
$M_f \sim [45, 56]$

$A_f \sim [0.08, 0.11]$
$M_f \sim [36, 46]$

Intermediate:

$A_f \sim [0.05, 0.07]$
$M_f \sim [16, 30]$

Easiest:

$A_f \sim [0.01, 0.03]$
$M_f \sim [3, 18]$

Sequence ranking

- Among the most challenging sequences

Matrix (A_f = 0.33, M_f = 57)

Rabbit(A_f = 0.31, M_f = 43)

Butterfly ($A_f = 0.22, M_f = 45$)

- Among the easiest sequences

Singer1 ($A_f = 0.02, M_f = 4$)

Octopus ($A_f = 0.01, M_f = 5$)

Sheep ($A_f = 0.02, M_f = 15$)

VOT Summary

- Top-performing trackers C-COT & TCNN (in EAO)
  - AR analysis indicates high accuracy and rare failures
  - Computationally quite complex (EFO)
- All top-performing trackers applied CNN features different localization strategy
- Most submitted trackers outperform standard baselines
- 22% of submitted trackers outperform the published sota bound as defined in VOT2016.

The VOT2016 online resources

Available at: http://www.votchallenge.net/vot2016

- Presentations + papers + Dataset + Evaluation kit
- Guidelines on how to evaluate your trackers on VOT2016 and produce graphs for your papers (directly comparable to 70 trackers!)
- Resources to apply the OTB evaluation as well
- Publish the code/binaries of trackers of coauthors: 66!!
- VOT is open source !

VOT2016 summary

- Results published in a 44 pages joint paper ~ 141 coauthors!

Winners of the VOT2016 challenge:

T-CNN by: Hyeonseob Nam, Mooyeol Baek and Bohyung Han

Tree-structured Convolutional Neural Network Tracker
Presentation at VOT2016 next

visual object tracking challenge

Visual Object Tracking Challenge VOT

USE OF BENCHMARKS IN PAPERS

Current state of the field

- Overviewed tracking papers

  (ICCV2013, ICCV2014, ECCV2014, CVPR2014, CVPR2015, CVPR2016, AVSS2015).

- Most popular datasets:

  OTB [Wu et al.,CVPR2013], VOT [Kristan et al., TPAMI2016]

- Researchers seem to use benchmarks

  (reproducible research)

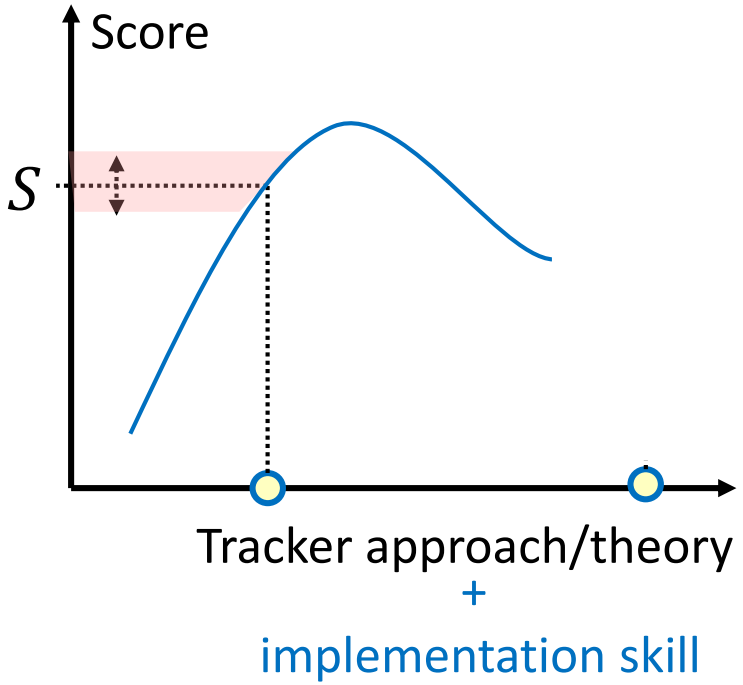
- The presented tracker is always “the best performing”

- BUT: (≤2015) Over 60% of papers did not use the entire benchmark, but only selected sequences!

  (≤2016) this number dropped to ~40%

Flaw of a single score obsession

- Idealized assumptions:
  - Single score ∝ Approach Quality
  - Score is "concave" in Approach
- Nonideal reality:
  - Imperfect/biased datasets:
    $\hat{S} = S + \text{noise}$



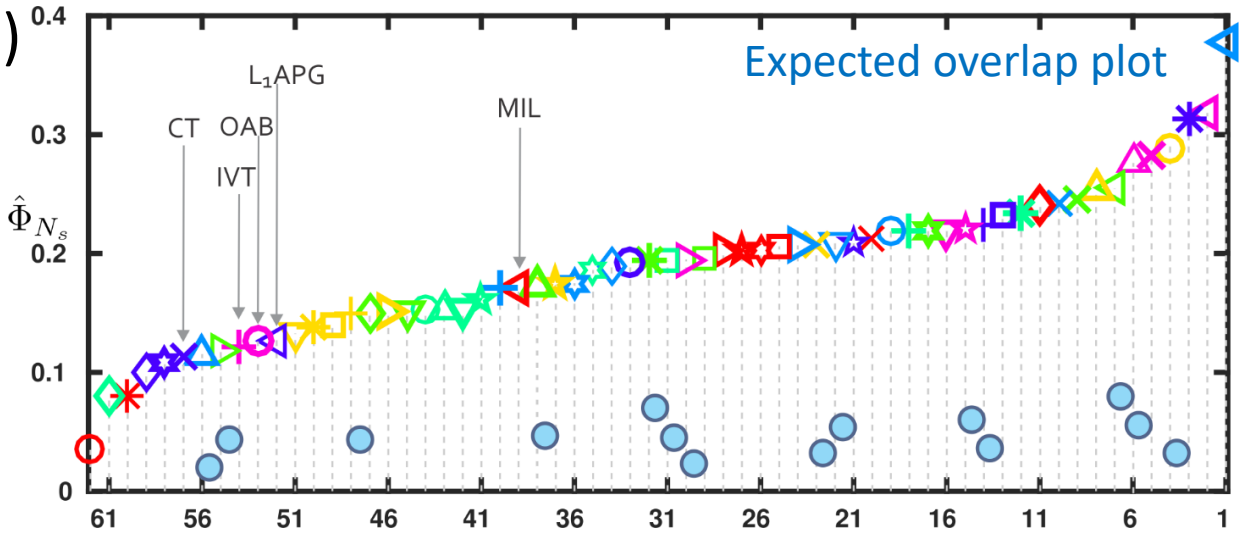
  - Scores also reflect implementation skill:

    Score $= f(\text{Implementation of the Approach})$

  - Score is NOT concave in approach (small increments)
  - Significant improvements may follow a change in paradigm

Flaw of a single score obsession

- VOT2015: 14 trackers published at ICCV, ECCV, CVPR, ICML, BMVC (2014-2015)



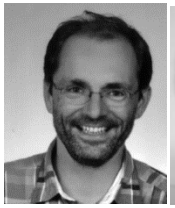
Guideline:

- Use a few non-correlated performance measures
- A tracker that scores reasonably high on a benchmark can be considered state-of-the-art
- Focus on a theory, not on maximizing a single performance measure

Thanks

- The VOT2016 committee


M. Kristan


J. Matas


A. Leonardis


M. Felsberg


R. Pflugfelder


G. Fernandez


L. Čehovin


T. Vojir


A. Lukežič


G. Häger

- Everyone who participated or contributed

Abhinav Gupta (Carnegie Mellon University, USA), Alfredo Petrosino (Parthenope University of Naples, Italy), Alireza Memarmoghadam (University of Isfahan, Iran), Alvaro Garcia-Martin (Universidad Autónoma de Madrid, Spain), Andrés Solís Montero (University of Ottawa, Canada), Andrea Vedaldi (University of Oxford, England), Andreas Robinson (Linköping University, Sweden), Andy J. Ma (Hong Kong Baptist University, China), Anton Varfolomieiev (Kyiv Polytechnic Institute, Ukraine), Aydin Alatan (Middle East Technical University, Turkey), Aykut Erdem (Hacettepe University, Turkey), Bernard Ghanem (KAUST, Saudi Arabia), Bin Liu (Moshanghua Tech Co., China), Bohyung Han (POSTECH, South Korea), Brais Martinez (University of Nottingham, England), Chang-Ming Chang (University at Albany, USA), Changsheng Xu (Chinese Academy of Sciences, China), Chong Sun (Dalian University of Technology, China), Chong Sun (Dalian University of Technology, China), Daijin Kim (POSTECH, South Korea), Dapeng Chen (Xi'an Jiaotong University, China), Dawei Du (University of Chinese Academy of Sciences, China), Dawei Du (University of Chinese Academy of Sciences, China), Deepak Mishra (Indian Institute of Space Science and Technology, India), Dit-Yan Yeung (Hong Kong University of Science and Technology, China), Erhan Gundogdu (Aselsan Research Center, Turkey), Erkut Erdem (Hacettepe University, Turkey), Fahad Khan (Linköping University, Sweden), Fahad Shahbaz Khan (Linköping University, Sweden), Fatih Porikli (ARC Centre of Excellence for Robotic Vision, Australia, Australian National University, Australia, Data61/CSIRO, Australia), Fei Zhao (Chinese Academy of Sciences, China), Filiz Bunyak (University of Missouri, USA), Francesco Battistone (Parthenope University of Naples, Italy), Gao Zhu (Australian National University, Australia), Giorgio Roffo (University of Verona, Italy), Gorthi R K Sai Subrahmanyam (Indian Institute of Space Science and Technology, India), Guilherme Bastos (Universidade Federal de Itajubá, Brazil), Guna Seetharaman (Naval Research Lab, USA), Henry Medeiros (Marquette University, USA), Hongdong Li (ARC Centre of Excellence for Robotic Vision, Australia), Honggang Qi (University of Chinese Academy of Sciences, China), Horst Bischof (Graz University of Technology, Austria), Horst Possegger (Graz University of Technology, Austria), Huchuan Lu (Dalian University of Technology, China), Huchuan Lu (Dalian University of Technology, China), Hyemin Lee (POSTECH, South Korea), Hyeonseob Nam (NAVER Corp., South Korea), Hyung Jin Chang (Imperial College London, England), Isabela Drummond (Universidade Federal de Itajubá, Brazil), Jack Valmadre (University of Oxford, England), Jae-chan Jeong (Electronics and Telecommunications Research Institute, South Korea), Jae-il Cho (Electronics and Telecommunications Research Institute, South Korea), Jae-Yeong Lee (Electronics and Telecommunications Research Institute, South Korea), Jianke Zhu (Zhejiang University, China), Jiayi Feng (Chinese Academy of Sciences, China), Jin Gao (Chinese Academy of Sciences, China), Jin Young Choi (ASRI, South Korea), Jingjing Xiao (University of Birmingham, England), Ji-Wan Kim (Electronics and Telecommunications Research Institute, South Korea), Jiyeoup Jeong (ASRI, South Korea), Joao F. Henriques (University of Oxford, England), Jochen Lang (University of Ottawa, Canada), Jongwon Choi (ASRI, South Korea), Jose M. Martinez (Universidad Autónoma de Madrid, Spain), Junliang Xing (Chinese Academy of Sciences, China), Junyu Gao (Chinese Academy of Sciences, China), Kannappan Palaniappan (University of Missouri, USA), Karel Lebeda (University of Surrey, England), Ke Gao (University of Missouri, USA), Krystian Mikolajczyk (Imperial College London, England), Lei Qin (Chinese Academy of Sciences, China), Lijun Wang (Dalian University of Technology, China), Lijun Wang (Dalian University of Technology, China), Longyin Wen (University at Albany, USA), Longyin Wen (University at Albany, USA), Luca Bertinetto (University of Oxford, England), Madan kumar Rapuru (Indian Institute of Space Science and Technology, India), Mahdieh Poostchi (University of Missouri, USA), Mario Maresca (Parthenope University of Naples, Italy), Martin Danelljan (Linköping University, Sweden), Matthias Mueller (KAUST, Saudi Arabia), Mengdan Zhang (Chinese Academy of Sciences, China), Michael Arens (Fraunhofer IOSB, Germany), Michel Valstar (University of Nottingham, England), Ming Tang (Chinese Academy of Sciences, China), Mooyeol Baek (POSTECH, South Korea), Muhammad Haris Khan (University of Nottingham, England), Naiyan Wang (Hong Kong University of Science and Technology, China), Nana Fan (Harbin Institute of Technology, China), Noor Al-Shakarji (University of Missouri, USA), Ondrej Miksik (University of Oxford, England), Osman Akin (Hacettepe University, Turkey), Payman Moallem (University of Isfahan, Iran), Pedro Senna (Universidade Federal de Itajubá, Brazil), Philip H. S. Torr (University of Oxford, England), Pong C. Yuen (Hong Kong Baptist University, China), Qingming Huang (Harbin Institute of Technology, China), Qingming Huang (University of Chinese Academy of Sciences, China), Rafael Martin-Nieto (Universidad Autónoma de Madrid, Spain), Rengarajan Pelapur (University of Missouri, USA), Richard Bowden (University of Surrey, England), Robert Laganière (University of Ottawa, Canada), Rustam Stolkin (University of Birmingham, England), Ryan Walsh (Marquette University, USA), Sebastian B. Krah (Fraunhofer IOSB, Germany), Shengkun Li (University at Albany, USA), Shengping Zhang (Harbin Institute of Technology, China), Shizeng Yao (University of Missouri, USA), Simon Hadfield (University of Surrey, England), Simone Melzi (University of Verona, Italy), Siwei Lyu (University at Albany, USA), Siwei Lyu (University at Albany, USA), Siyi Li (Hong Kong University of Science and Technology, China), Stefan Becker (Fraunhofer IOSB, Germany), Stuart Golodetz (University of Oxford, England), Sumithra Kakanuru (Indian Institute of Space Science and Technology, India), Sunglok Choi (Electronics and Telecommunications Research Institute, South Korea), Tao Hu (University of Chinese Academy of Sciences, China), Thomas Mauthner (Graz University of Technology, Austria), Tianzhu Zhang (Chinese Academy of Sciences, China), Tony Pridmore (University of Nottingham, England), Vincenzo Santopietro (Parthenope University of Naples, Italy), Weiming Hu (Chinese Academy of Sciences, China), Wenbo Li (Lehigh University, USA), Wolfgang Hübner (Fraunhofer IOSB, Germany), Xiangyuan Lan (Hong Kong Baptist University, China), Xiaomeng Wang (University of Nottingham, England), Xin Li (Harbin Institute of Technology, China), Yang Li (Zhejiang University, China), Yiannis Demiris (Imperial College London, England), Yifan Wang (Dalian University of Technology, China), Yuankai Qi (Harbin Institute of Technology, China), Zejian Yuan (Xi'an Jiaotong University, China), Zexiong Cai (Hong Kong Baptist University, China), Zhan Xu (Zhejiang University, China), Zhenyu He (Harbin Institute of Technology, China), Zhizhen Chi (Dalian University of Technology, China).

VOT2016 sponsor:

University *of Ljubljana*
Faculty *of Computer and Information Science*