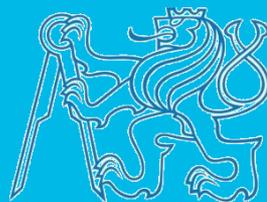


The Visual Object Tracking VOT-TIR2015 Challenge Results

Michael Felsberg, Amanda Berg, Jörgen Ahlberg,
Gustav Häger, Matej Kristan, Jiří Matas, Aleš
Leonardis, Luka Čehovin, Gustavo Fernández, Tomáš
Vojtř, Georg Nebhay, Roman Pflugfelder, et al.



University of Ljubljana
Faculty of Computer and
Information Science

UNIVERSITY OF
BIRMINGHAM

Outline

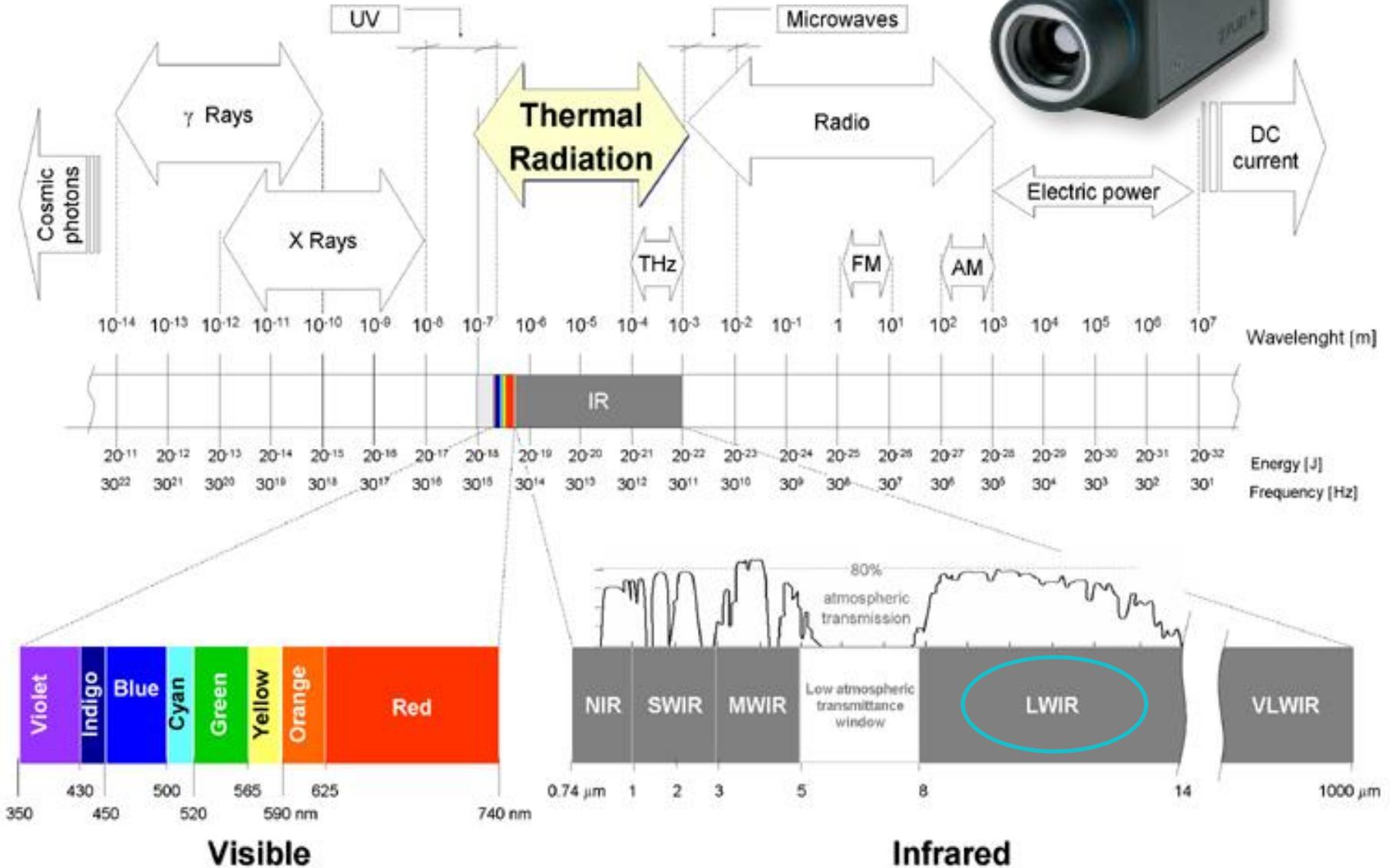
1. Scope of the VOT-TIR challenge
 - Thermal infrared imaging
2. VOT-TIR2015 challenge overview
 - Evaluation system
 - **Dataset**
 - Performance evaluation measures
3. VOT-TIR2015 results overview
4. Summary and outlook

Scope of the VOT-TIR challenge

- Single-object, single thermal infrared (TIR) 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 an upright bounding box



Thermal Infrared



Applications of TIR

- Scientific research
- Security
- Fire monitoring
- Search and rescue
- Automotive safety
- Personal use
- ~~Military~~



Why a separate challenge?

Tracking in TIR different from tracking in low resolution grayscale visual?

Many similarities but also interesting differences

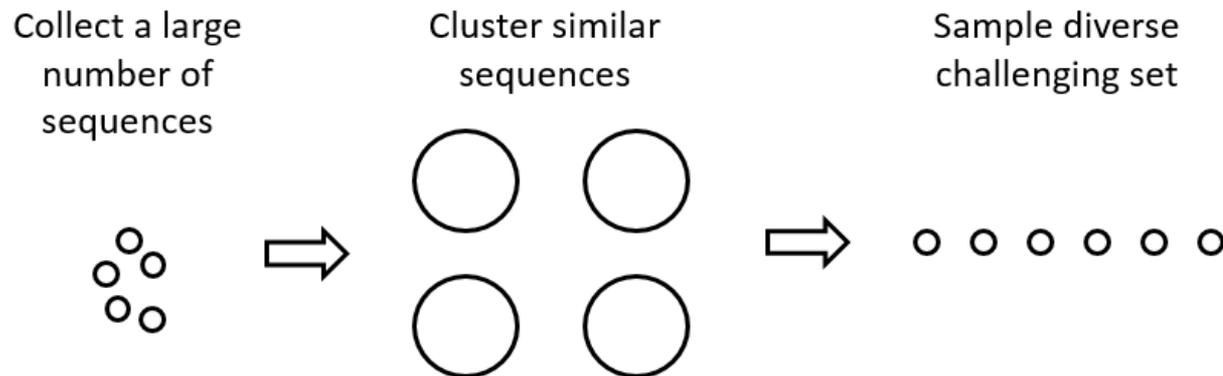
- 16-bit
- Constant values if radiometric
- Less structure/edges/texture
- No shadows
- Noise: blooming, resolution, dead pixels

Evaluation system from VOT 2015

- Matlab-based kit to automatically perform a battery of standard experiments
- Download from our homepage
 - <https://github.com/votchallenge/vot-toolkit>
 - select the `vottir2015` experiment stack
- Plug and play!
 - Supports multiple platforms and programming languages (C/C++/Matlab/Python, etc.)
- Easy to evaluate your tracker on our benchmarks
- Deep integration with tracker - Fast execution of experiments

VOT-TIR2015 Dataset: LTIR

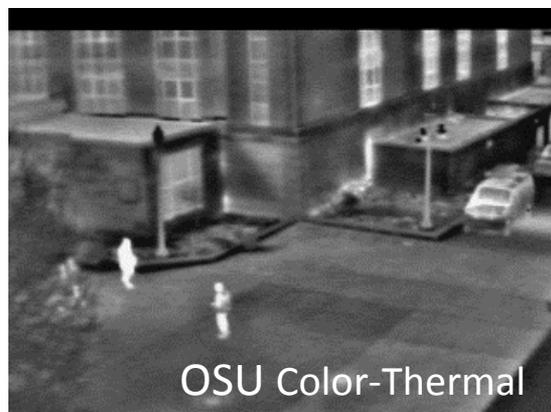
- Follows VOT approach:
 - Keep it sufficiently small, diverse and well annotated
 - Follow the VOT dataset construction methodology



- Linköping Thermal InfraRed (LTIR) dataset
A. Berg, J. Ahlberg, M. Felsberg, *A Thermal Object Tracking Benchmark*. AVSS 2015.

Existing datasets

Name	Purpose	Resolution	#Bits	Stat/Mov
OSU Pedestrian [5]	Pedestrian detection and tracking.	360×240	8	Y/N
OSU Color-Thermal [6]	Pedestrian detection, tracking and thermal/visual fusion.	360×240	8	Y/N
Terravic Motion [7]	Detection and tracking	320×240	8	Y/N
LITIV [8]	Visible-infrared registration.	320×240	8	Y/N
ASL-TID [9]	Object (pedestrian, cat, horse) detection and tracking.	324×256	8/16	N/Y
BU-TIV [10]	Various visual analysis tasks. Single-object, multiple-object and multiple sensor tracking as well as motion patterns.	Up to 1024×1024	16	Y/N



- Different sources
- Different applications
- Different sensors
- Moving + stationary sensors
- Radiometric + non-radiometric
- 8/16 bits



- Different sources
- **Different applications**
- Different sensors
- Moving + stationary sensors
- Radiometric + non-radiometric
- 8/16 bits



- Different sources
- Different applications
- **Different sensors**
- Moving + stationary sensors
- Radiometric + non-radiometric
- 8/16 bits



- Different sources
- Different applications
- Different sensors
- **Moving + stationary sensors**
- **Radiometric + non-radiometric**
- **8/16 bits**



Properties

- 20 Sequences
- Average sequence length 563
- Annotations in accordance with VOT-standard
 - Bounding-box
 - 11 global attributes (per-sequence)
 - 6 local attributes (per-frame)

Occlusion, **dynamics change**, object motion, object size change, camera motion, neutral

Sequence details

	ID	Name	Sensor	Resolution	#Frames	#Bit	Object
	1	rhino behind tree	FLIR A35	320 × 256	619	8/16	Rhino
	2	running rhino	FLIR A35	320 × 256	763	8/16	Rhino
	3	garden	FLIR Tau 320	324 × 256	676	8/16	Human
	4	horse	FLIR Tau 320	324 × 256	348	8/16	Horse
	5	hiding	FLIR Photon 320	320 × 240	358	8	Human
	6	mixed distractors	FLIR Photon 320	320 × 240	270	8	Human
	7	saturated	AIM QWIP	640 × 480	218	8	Human
	8	street	AIM QWIP	640 × 480	172	8	Human
	9	car	FLIR A655SC	640 × 480	1420	8/16	Car
	10	crouching	FLIR A655SC	640 × 480	618	8/16	Human
	11	crowd	FLIR A65	640 × 512	71	8/16	Human
	12	soccer	3×AXIS Q-1922	1920 × 480	775	8	Human
	13	birds	FLIR T640	640 × 480	270	8	Human
	14	crossing	FLIR A655SC	640 × 480	301	8/16	Human
	15	depthwise crossing	FLIR A655SC	640 × 480	851	8/16	Human
	16	jacket	FLIR A655SC	640 × 480	1451	8/16	Human
	17	quadrocopter	FLIR T640	640 × 480	178	8	Quadrocopter
	18	quadrocopter2	FLIR A655SC	640 × 480	1010	8/16	Quadrocopter
	19	selma	FLIR A655SC	640 × 480	235	8/16	Dog
	20	trees	FLIR A655SC	640 × 480	665	8/16	Human



(1) rhino behind tree (2) running rhino (3) garden (4) horse (5) hiding (6) mixed distractors



(7) saturated (8) street (9) car (10) crouching (11) crowd



(12) soccer (13) birds (14) crossing (15) depthwise crossing



(16) jacket (17) quadcopter (18) quadcopter2 (19) selma (20) trees



Ours

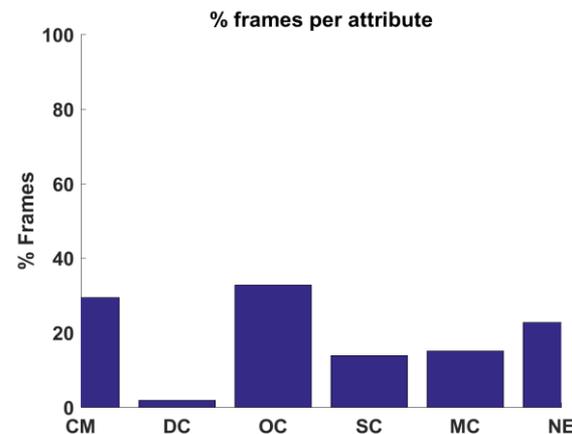
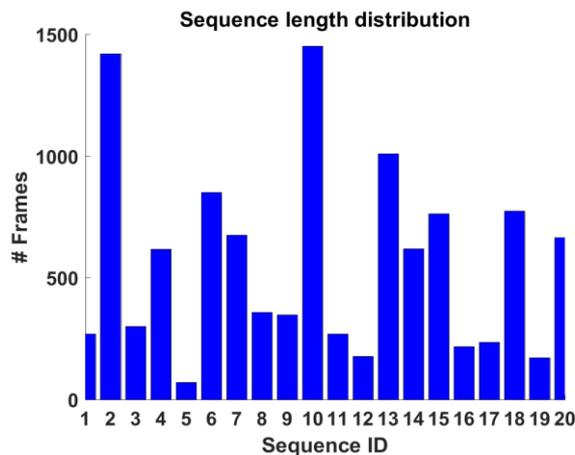
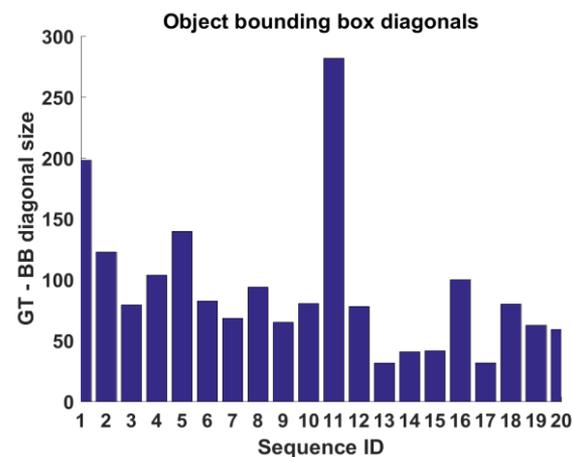
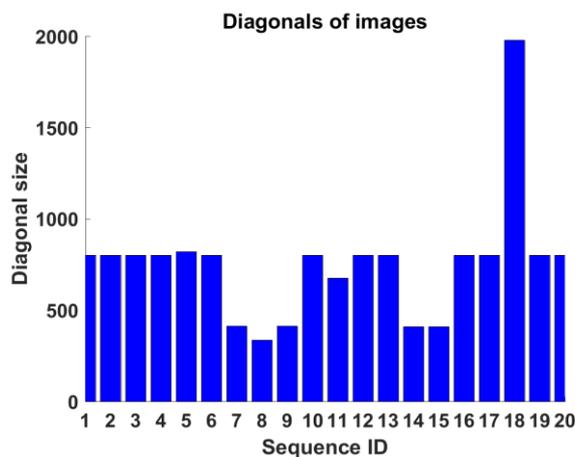


Old

Global attributes

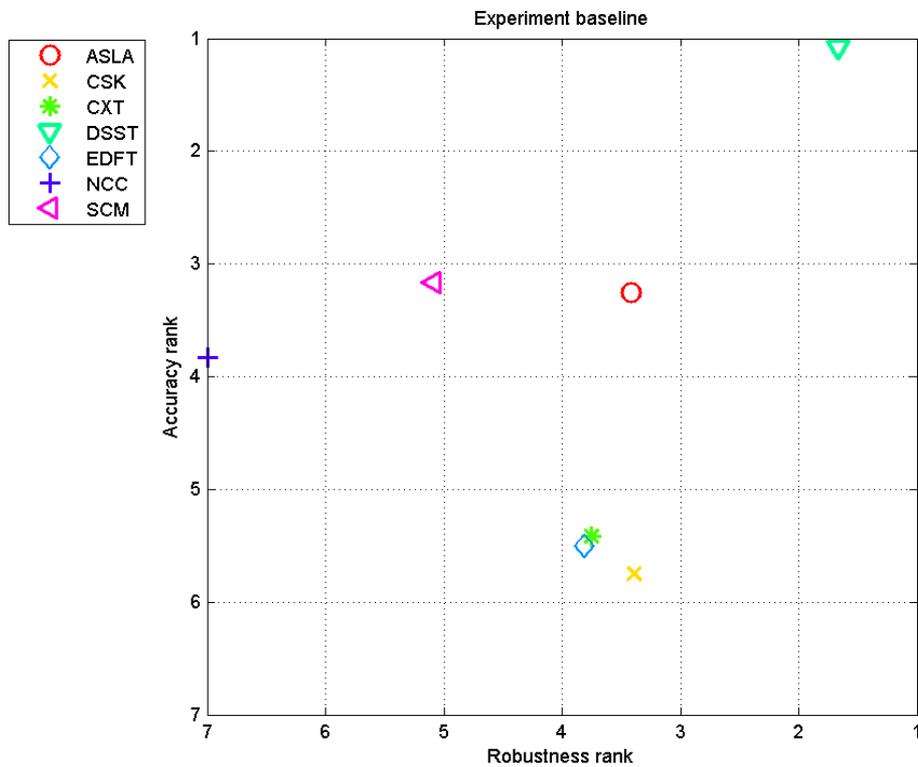
ID	Name	Cam mot	Dyn chg	Obj mot	Bg cltr	Size chg	Asp ratio	Blur	Tmp chg	Def	Occ	Mot chg
1	rhino behind tree	x		x						x	x	x
2	running rhino	x		x	x		x	x		x		
3	garden	x		x	x	x	x	x		x	x	x
4	horse	x		x		x	x			x	x	x
5	hiding	x	x	x	x	x	x	x		x	x	x
6	mixed distractors			x	x		x		x	x		x
7	saturated	x		x	x	x				x	x	x
8	street	x		x	x	x				x	x	x
9	car			x	x	x	x					x
10	crouching			x	x	x	x			x	x	x
11	crowd			x	x					x	x	
12	soccer			x	x	x	x			x	x	x
13	birds			x	x	x	x	x		x	x	x
14	crossing			x						x	x	
15	depthwise crossing			x	x	x				x		x
16	jacket			x	x	x			x	x	x	x
17	quadrocopter	x	x	x	x		x	x				x
18	quadrocopter2			x	x	x		x				x
19	selma			x	x					x		x
20	trees			x						x	x	

LTIR general stats

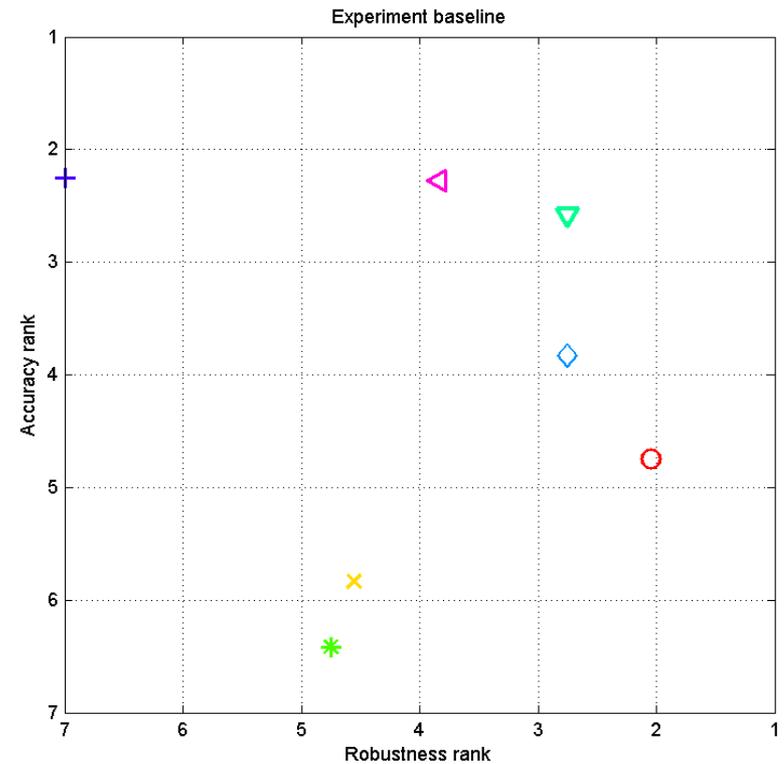




Will it be different? Test against VOT2014



VOT2014



LTIR

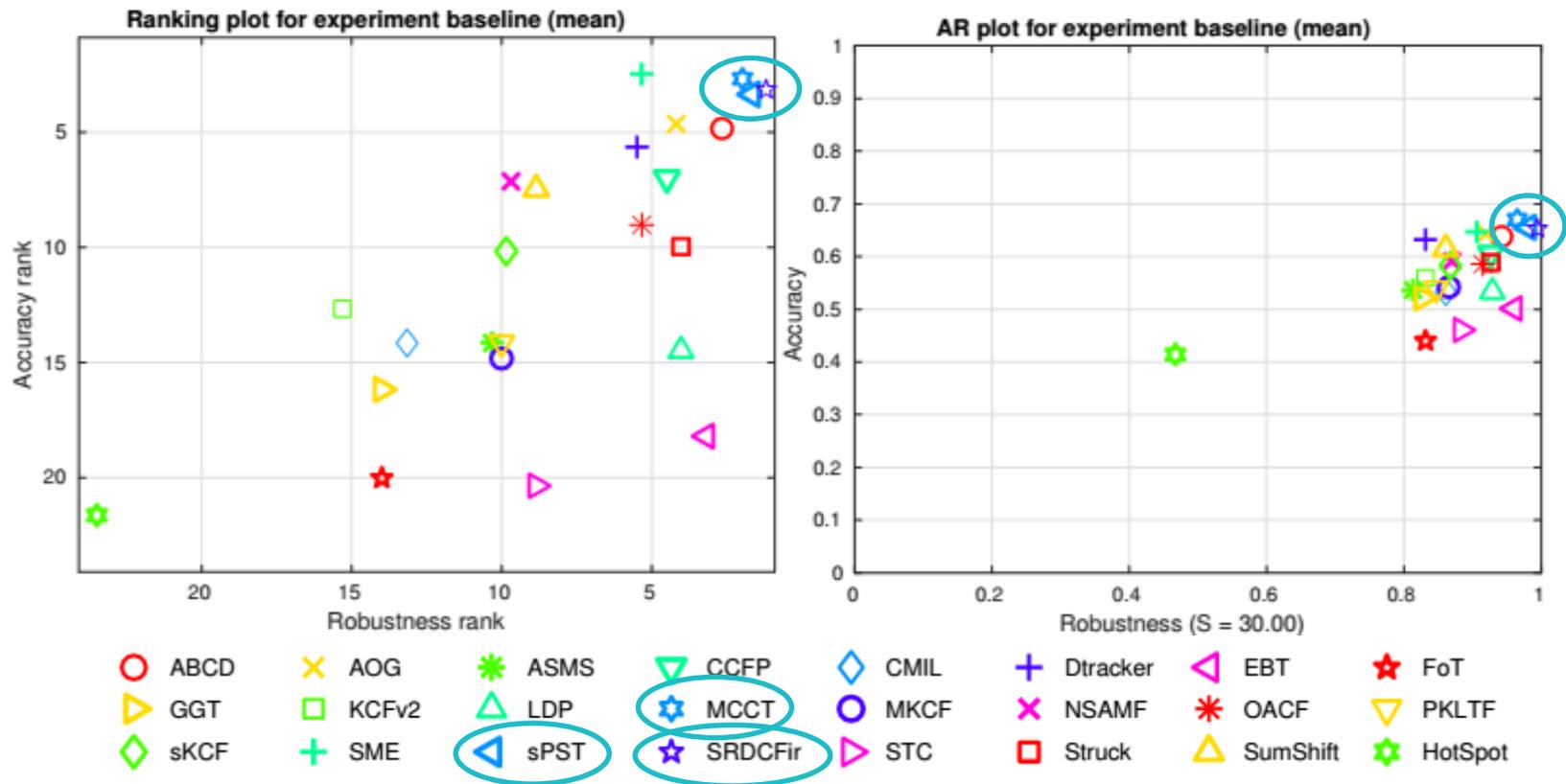
Performance evaluation measures

- Basically the same as VOT2015 (based on 8-bit)
 - accuracy
 - robustness
- Evaluated globally and per-attribute
 - raw value
 - rank
- Overall: expected average overlap
- Speed in EFO units

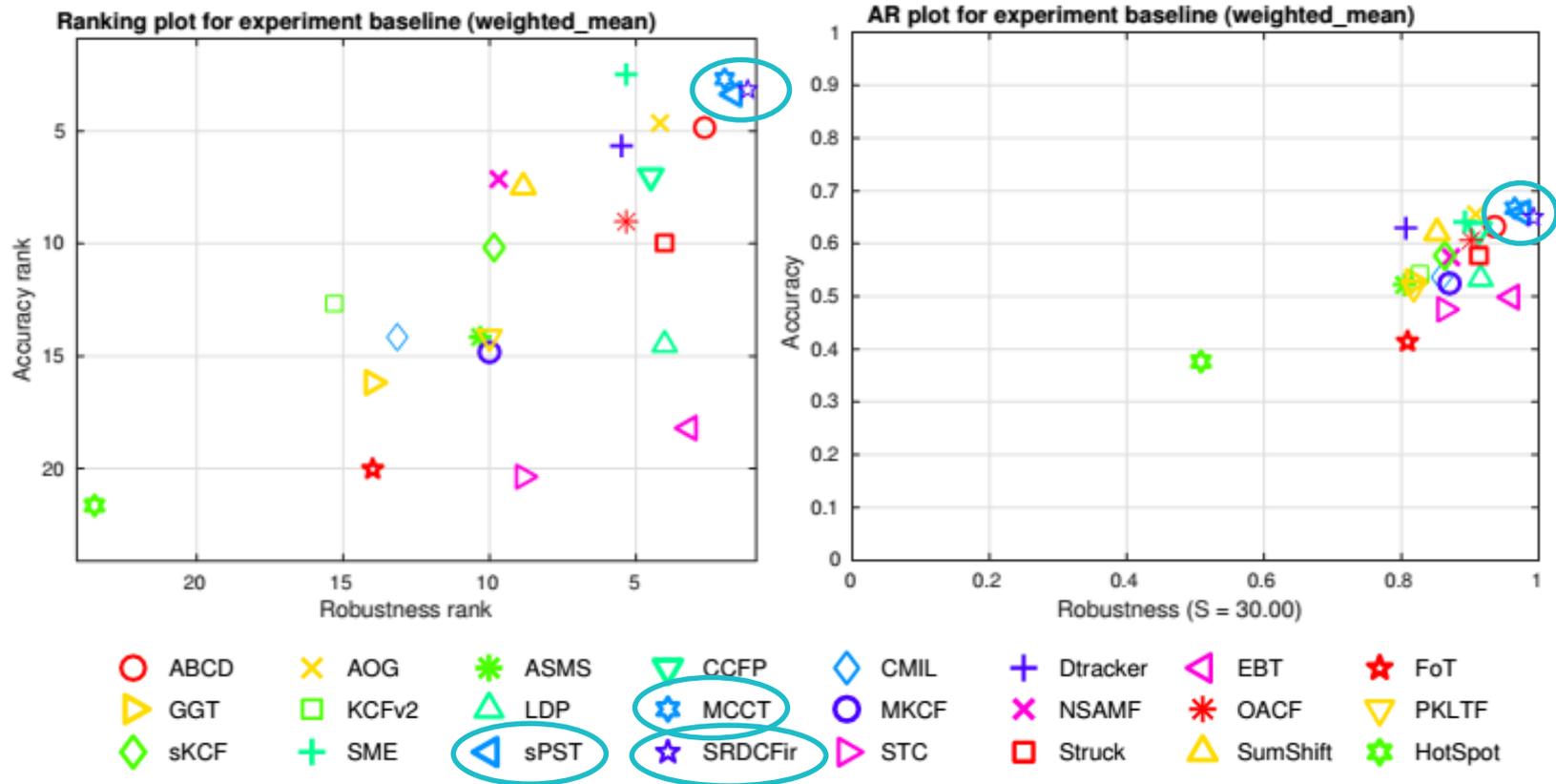
Results

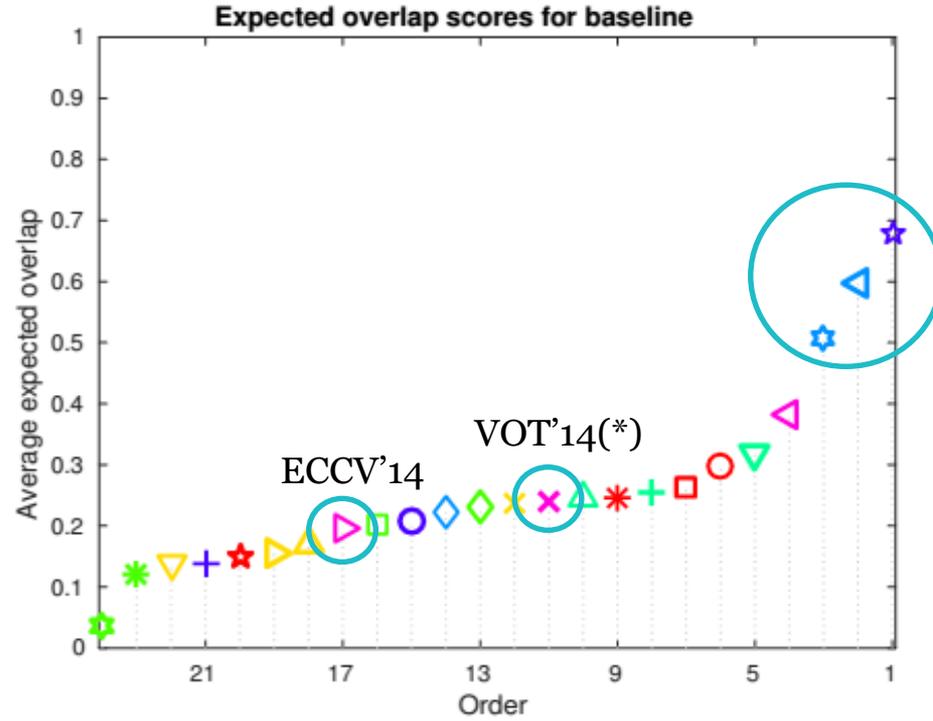
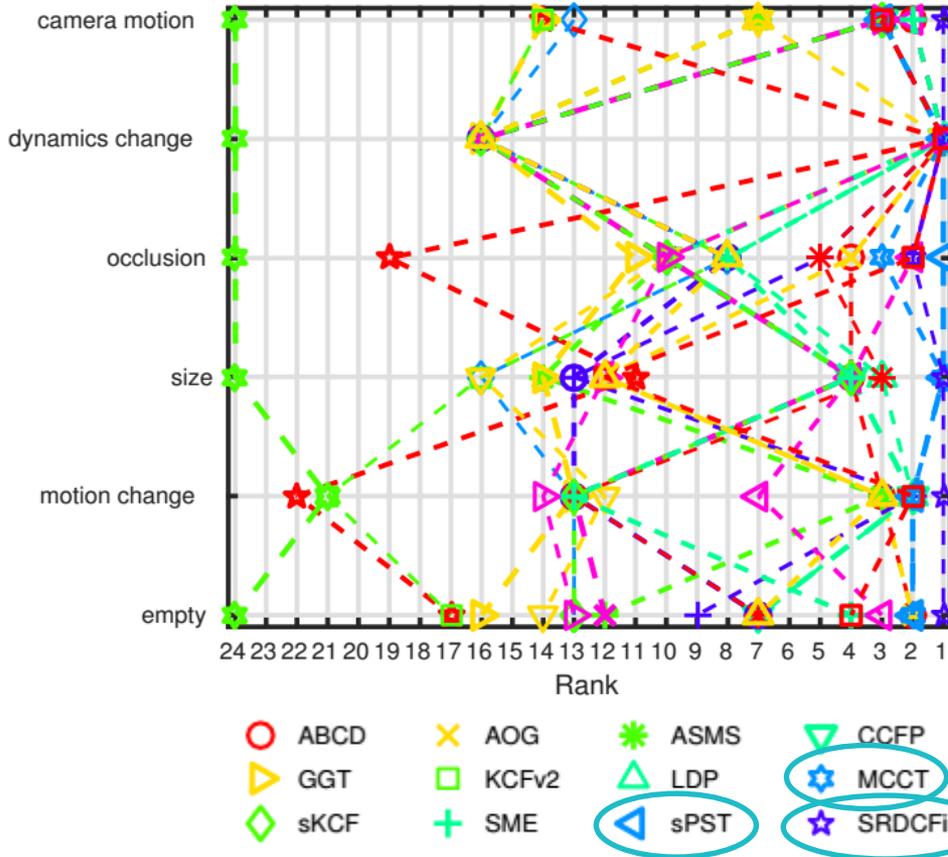
- 20 submitted trackers
- 4 added by VOT committee
- 20 of 24 trackers in both challenges
- Various classes of trackers
 - mean shift extensions (ASMS, PKLTF, SumShift, DTracker)
 - part-based trackers (LDP, G2T, AOGTracker, MCCT, FoT)
 - correlation filter based (NSAMF, OACF, SRDCFir, sKCF, STC, MKCF+, CCFP, SME, KCFv2)
 - others (EBT, CMIL, sPST, Struck, ABCD, HotSpot)

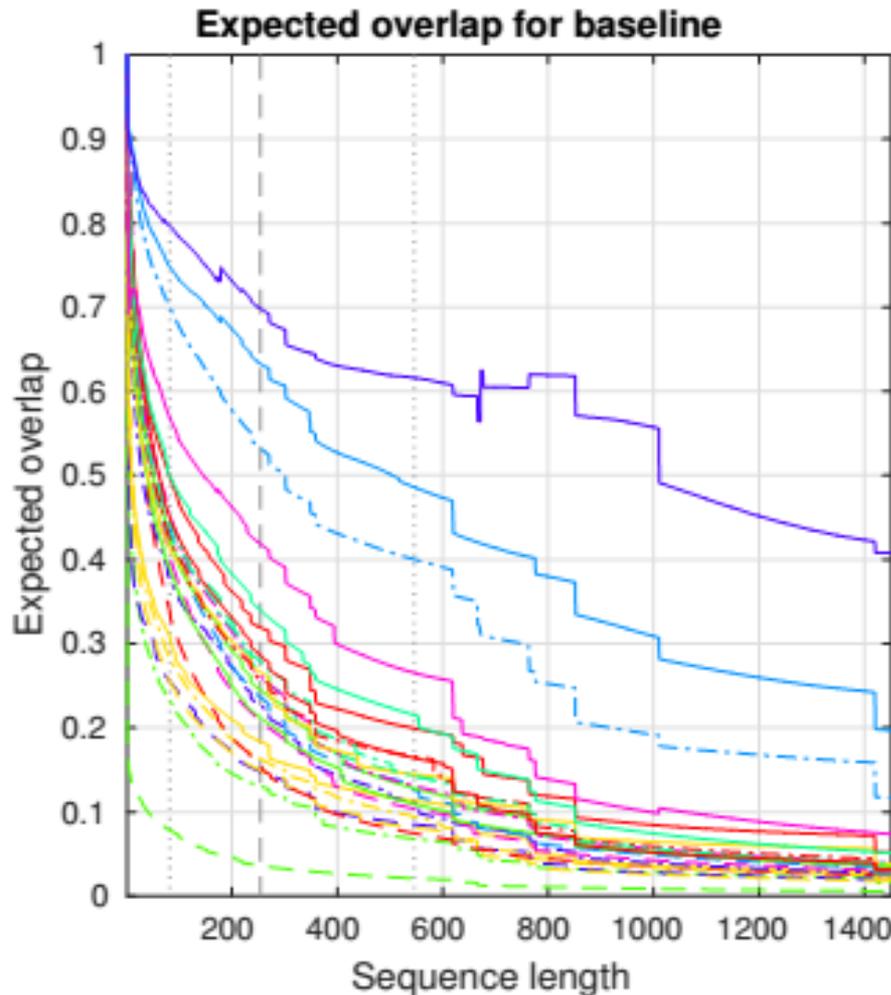
Results (sequence pooling)



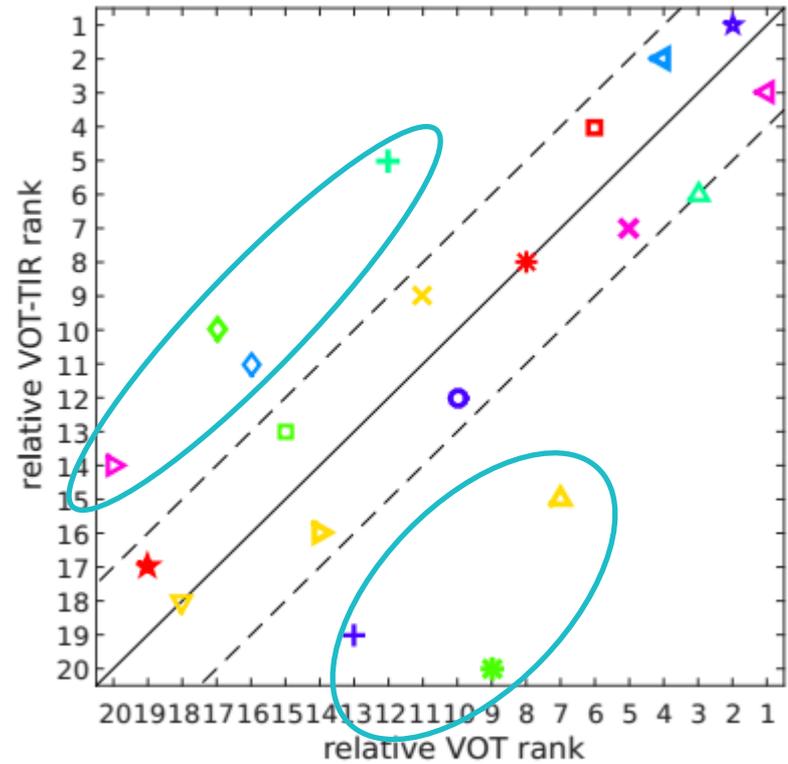
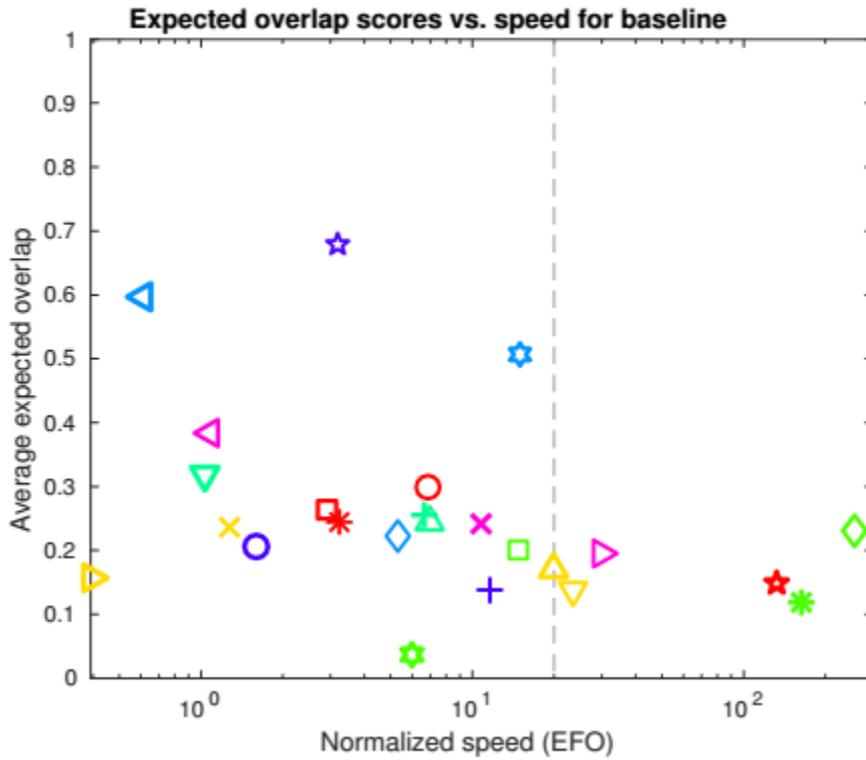
Results (attribute normalization)







Tracker	A	R	$\hat{\Phi}$	Speed
SRDCFir	0.65	0.58	0.70	3.17
sPST	0.66	2.18	0.64	0.61
MCCT	0.67	3.34	0.55	15.05
EBT	0.50	3.50	0.43	1.08
CCFP	0.63	8.55	0.36	1.03
ABCD	0.63	5.81	0.34	6.88
Struck	0.58	8.48	0.30	2.90
SME	0.64	9.97	0.30	6.67
LDP	0.53	8.33	0.29	6.96
NSAMF	0.57	12.63	0.28	10.69
OACF	0.61	9.57	0.28	3.22
AOG	0.65	8.76	0.27	1.27
sKCF	0.58	13.90	0.27	255.13
CMIL	0.54	14.04	0.25	5.31
MKCF+	0.52	12.61	0.24	1.60
KCFv2	0.54	17.81	0.23	14.78
STC	0.48	13.85	0.23	29.92
SumShift	0.62	15.67	0.19	19.78
G2T	0.53	18.59	0.18	0.39
FoT	0.41	19.40	0.17	131.57
PKLTF	0.52	19.30	0.16	23.65
Dtracker	0.63	19.69	0.16	11.55
ASMS	0.52	20.03	0.14	163.42
HotSpot	0.38	62.27	0.04	5.98



- | | | | | | | | |
|--------|---------|--------|-----------|--------|------------|------------|-----------|
| ○ ABCD | × AOG | * ASMS | ▽ CCFP | ◇ CMIL | + Dtracker | ◁ EBT | ★ FoT |
| ▷ GGT | □ KCFv2 | △ LDP | ☆ MCCT | ○ MKCF | × NSAMF | * OACF | ▽ PKLTF |
| ◇ sKCF | + SME | ▷ sPST | ☆ SRDCFir | ▷ STC | □ Struck | △ SumShift | * HotSpot |

Sequence ranking

- A_f : average number of trackers failed per frame
- M_f : max. number of trackers failed at a single frame

Sequence	Score
crowd	2
quadrocopter	2,5
quadrocopter2	2,5
garden	3
mixed_distractors	3
saturated	3,5
selma	3,5
street	3,5
birds	4
crouching	4

Sequence	Score
jacket	4
hiding	4,5
car	5
crossing	5
depthwise_crossing	5
horse	5
rhino_behind_tree	5
running_rhino	5
soccer	5
trees	5

challenging:

$$0.06 \leq A_f \leq 0.2$$

$$14 \leq M_f \leq 22$$

intermediate:

$$0.04 \leq A_f \leq 0.1$$

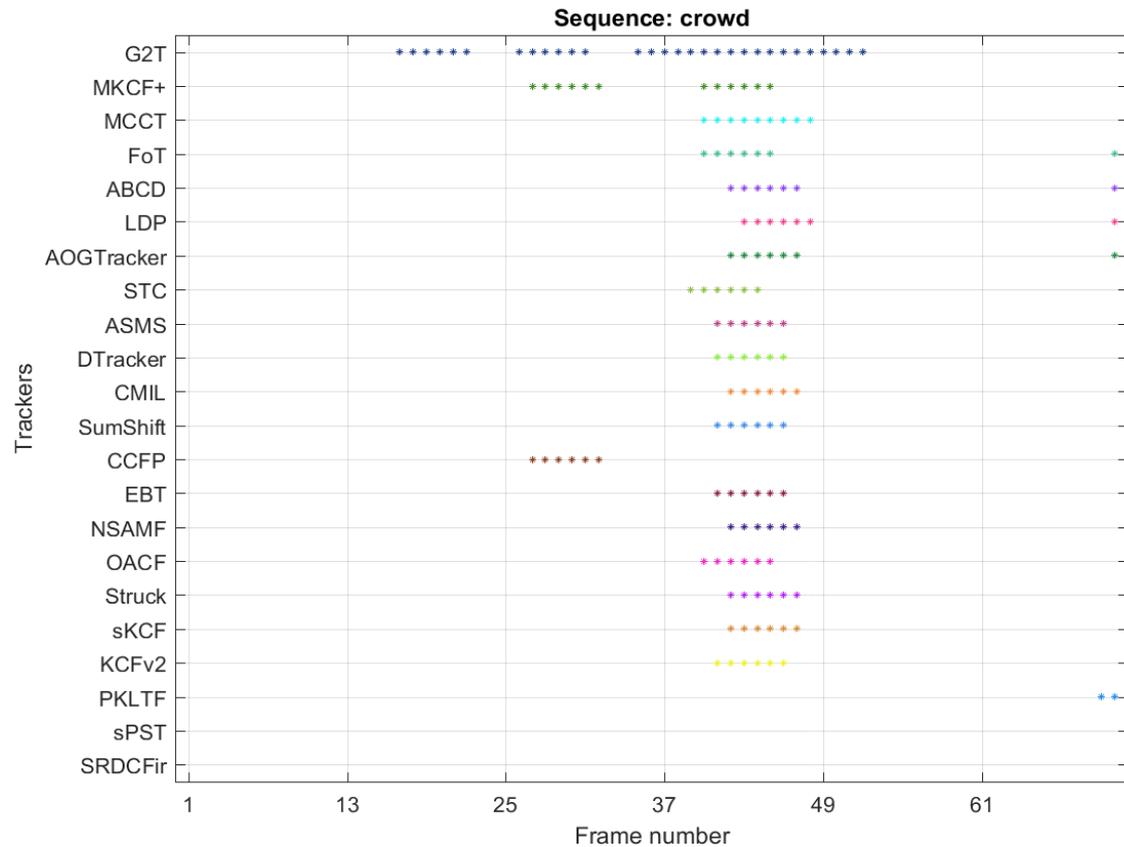
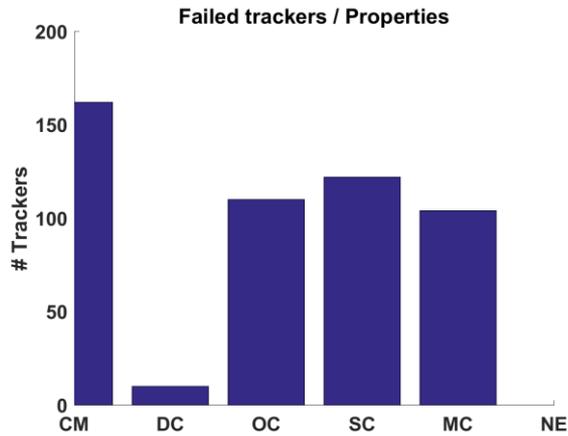
$$6 \leq M_f \leq 11$$

easiest:

$$0 \leq A_f \leq 0.04$$

$$0 \leq M_f \leq 7$$

Difficulty analysis



Summary

- New challenge with LTIR dataset
- Dataset too easy (or trackers too good)
- Large fraction of trackers show similar ranking in both challenges
 - in contrast to VOT2014
- Top-performing triple: SRDCFir, sPST, MCCT
- Best real-time method: sKCF
- Available at <http://www.votchallenge.net/vot2015>

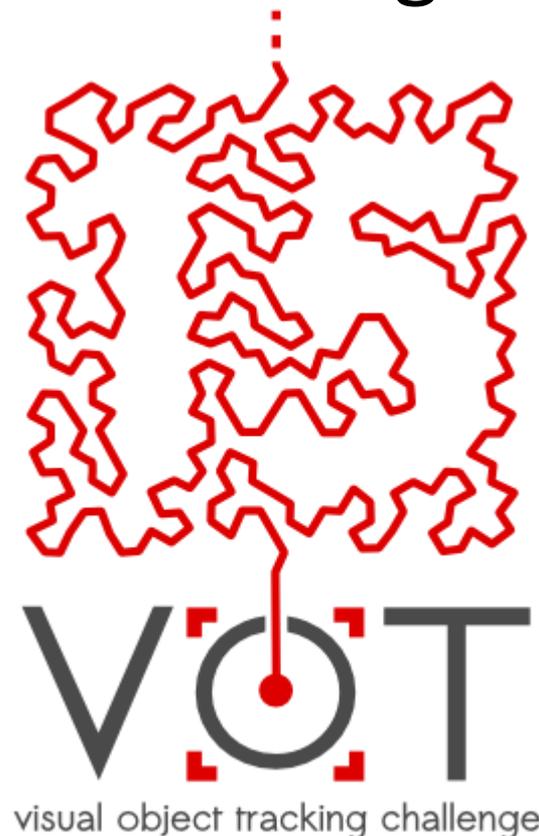
Winners of the VOT-TIR2015 Challenge:

Yang Hua, Karteek Alahari,
and Cordelia Schmid:

Simplified Proposal Selection
Tracker (sPST)

Presentation at VOT2015
today at 17:35

Award sponsored by



Thanks

- The VOT2015 Committee
- Jörgen and Amanda
- Everyone who participated:

Alan Lukezic, Alvaro Garcia-Martin, Amir Saffari, Ang Li, Andres Solis Montero, Baojun Zhao, Cordelia Schmid, Dapeng Chen, Dawei Du, Fahad Shahbaz Khan, Fatih Porikli, Gao Zhu, Guibo Zhu, Hanqing Lu, Hilke Kieritz, Hongdong Li, Honggang Qi, Jae-chan Jeong, Jae-il Cho, Jae-Yeong Lee, Jianke Zhu, Jiatong Li, Jiayi Feng, Jinqiao Wang, Ji-Wan Kim, Jochen Lang, Jose M. Martinez, Kai Xue, Karteek Alahari, Liang Ma, Lipeng Ke, Longyin Wen, Luca Bertinetto, Martin Danelljan, Michael Arens, Ming Tang, Ming-Ching Chang, Ondrej Miksik, Philip H S Torr, Rafael Martin-Nieto, Robert Laganiere, Sam Hare, Siwei Lyu, Song-Chun Zhu, Stefan Becker, Stephen L Hicks, Stuart Golodetz, Sunglok Choi, Tianfu Wu, Wolfgang Hubner, Xu Zhao, Yang Hua, Yang Li, Yang Lu, Yuezun Li, Zejian Yuan, and Zhibin Hong