The Visual Object Tracking VOT-TIR2016 Challenge Results

Michael Felsberg, Matej Kristan, Jiři Matas, Aleš Leonardis, Roman Pflugfelder, Gustav Häger, Amanda Berg, Abdelrahman Eldesokey, Jörgen Ahlberg, Luka Čehovin, Tomáš Vojiř, Alan Lukežič, Gustavo Fernández, et al.

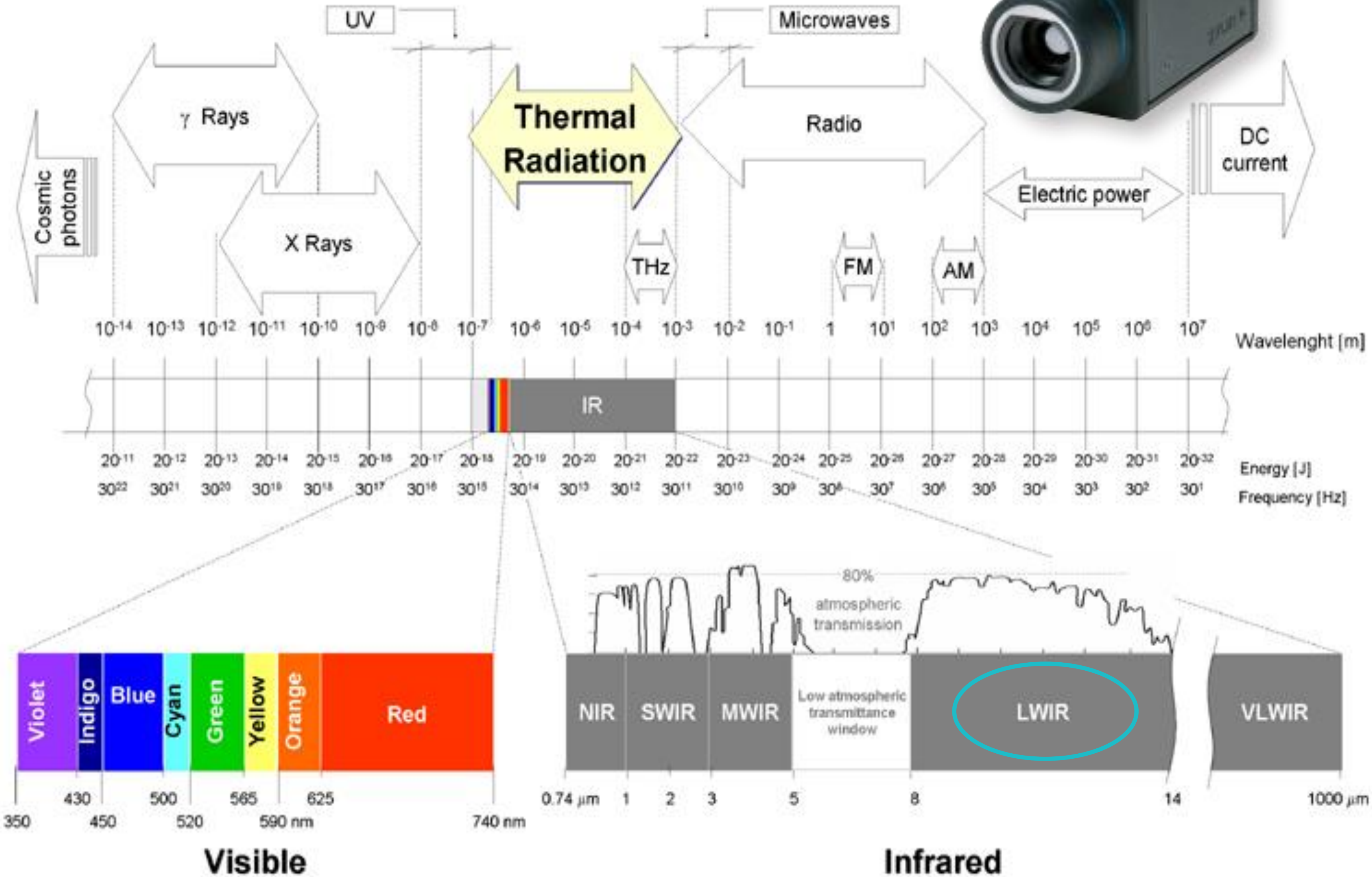
Outline

1. Scope of the VOT-TIR challenge
   - Thermal infrared imaging
2. VOT-TIR2016 challenge overview
   - Evaluation system
   - Dataset
   - Performance evaluation measures
3. VOT-TIR2016 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 bounding box 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



http://www.theses.ulaval.ca/2005/23016/apb.html

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

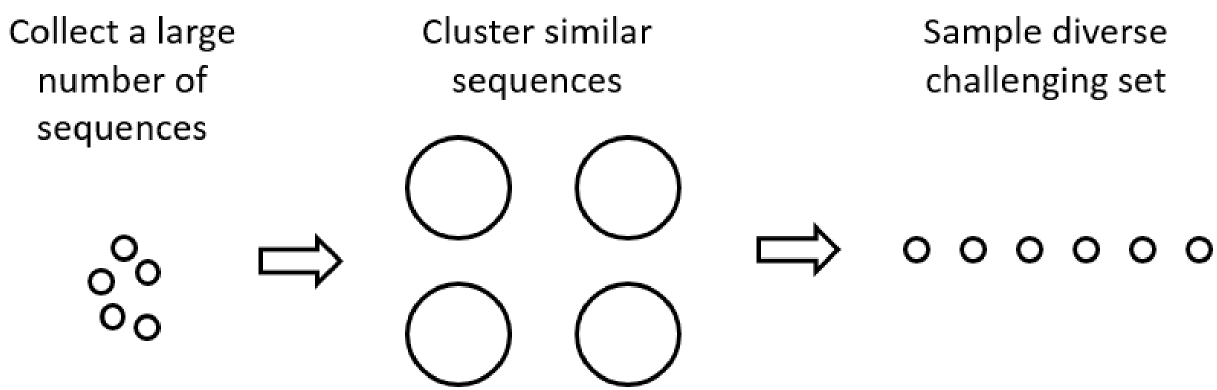
- 16-bit
- Constant values if radiometric
- Less structure/edges/texture
- No shadows
- Noise: blooming, resolution, dead pixels

Evaluation system from VOT 2016

- Matlab-based kit to automatically perform a battery of standard experiments
- Download from our homepage
  - https://github.com/vicoslab/vot-toolkit
  - select the `vottir2016` 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
- OTB-like evaluation omitted

VOT-TIR2016 Dataset: LTIR2016

- Follows VOT 2013 selection and annotation approach:
  - Keep it sufficiently small, diverse and well annotated
  - Follow the VOT dataset construction methodology



- Modification of Linköping Thermal InfraRed (LTIR) dataset

  A. Berg, J. Ahlberg, M. Felsberg, *A Thermal Object Tracking Benchmark*. AVSS 2015.

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

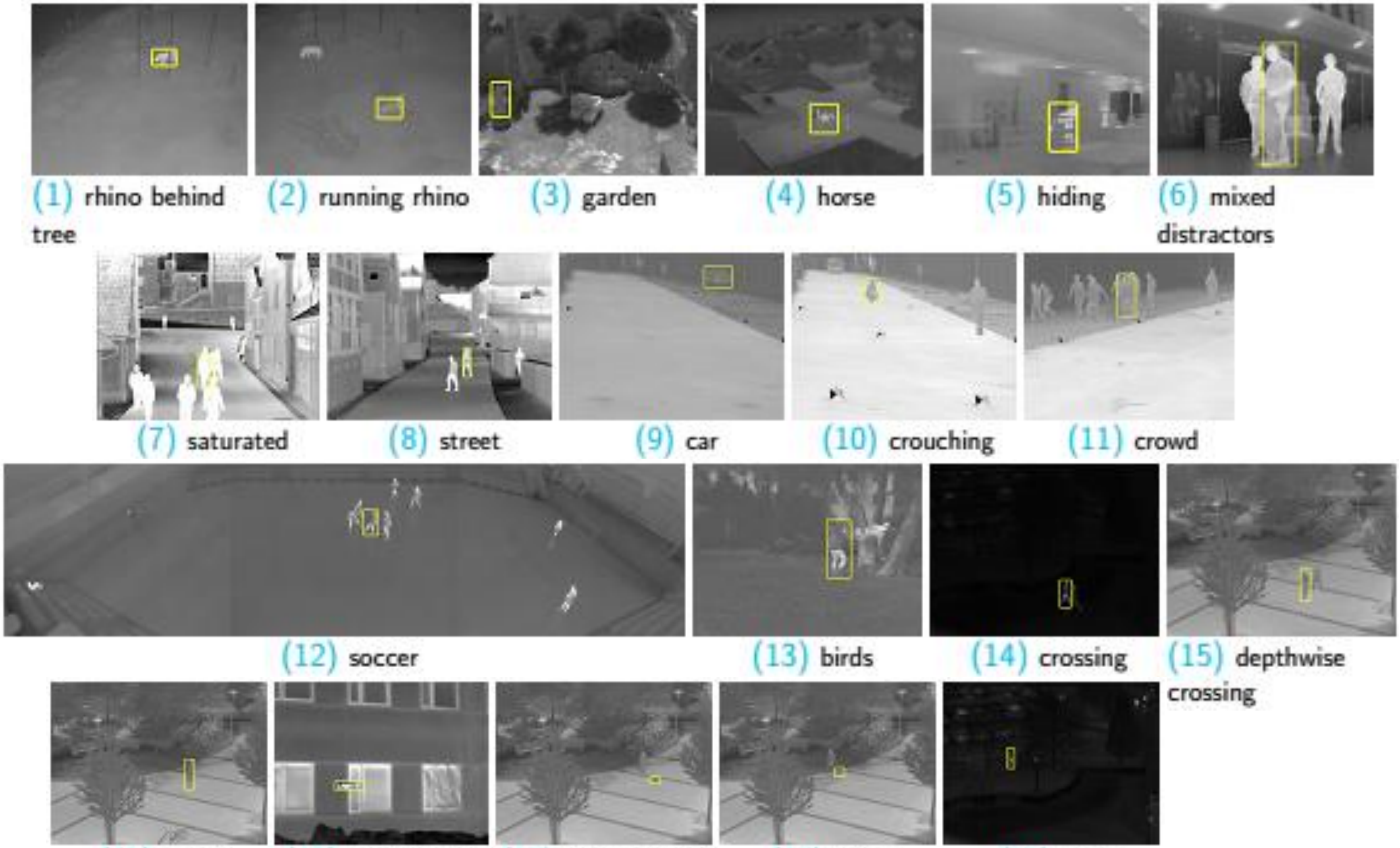
- Different sources
- Different applications
- Different sensors
- Moving + stationary sensors
- Radiometric + non-radiometric
- 8/16 bits

- 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

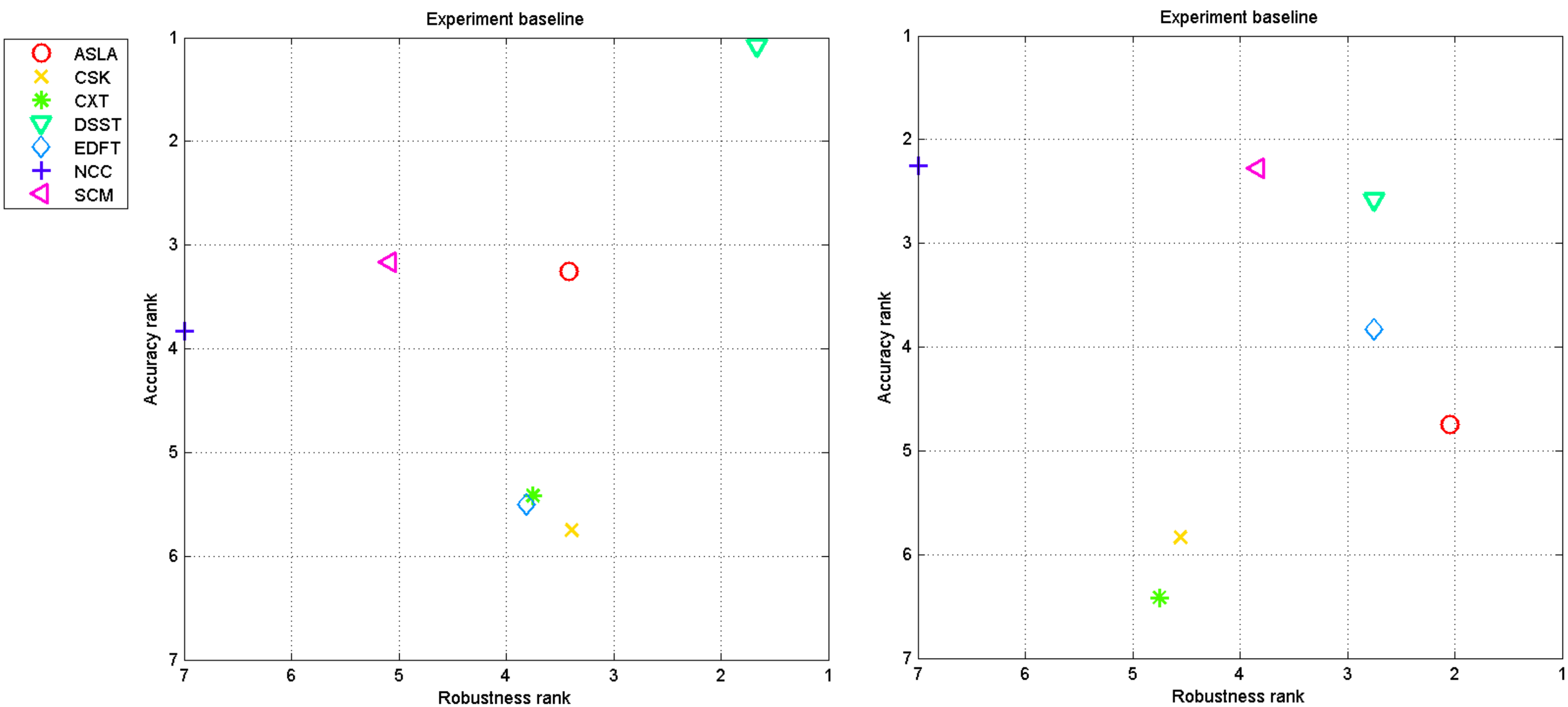
Sequence details

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) quadrocopter (18) quadrocopter2 (19) selma (20) trees

Will it be different? Test against VOT2014



VOT2014 LTIR

Problem 2015: 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<=A_f<=0.2
14<=M_f<=22

intermediate:

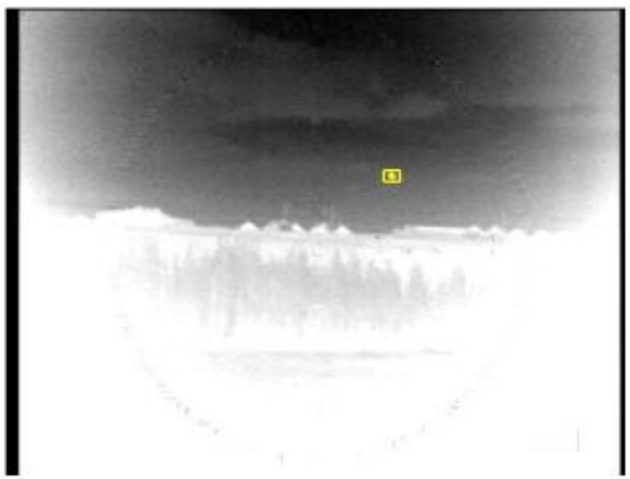
0.04<=A_f<=0.1
6<=M_f<=11

easiest:

0<=A_f<=0.04
0<=M_f<=7

Modifications of LTIR

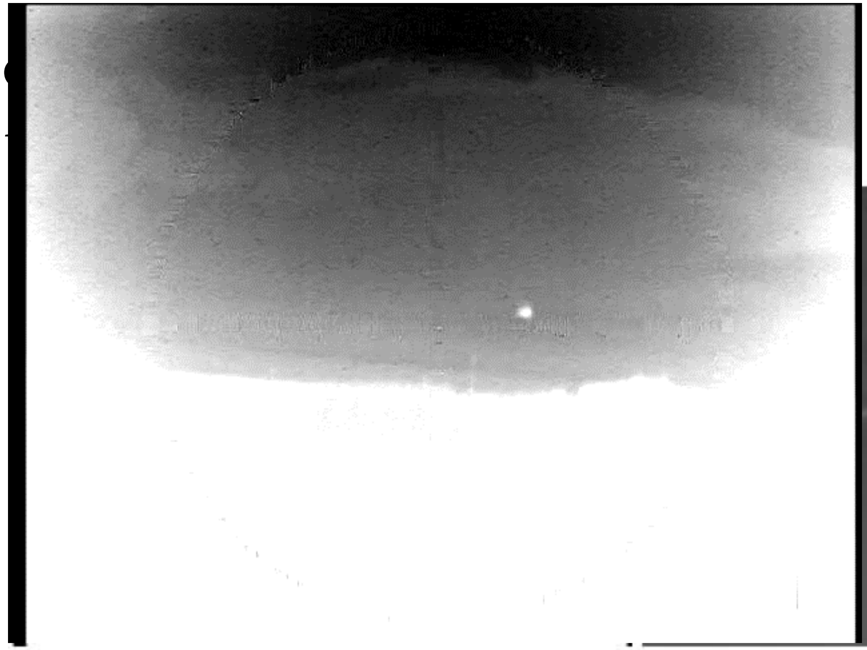
- VOT-TIR2015 was already saturated
- Call for sequences – limited success (3 new sources, too easy)
- Easiest sequences have been removed: *Crossing*, *Horse*, and *Rhino behind tree*
- New, more difficult sequences have been added: *Bird*, *Boat1*, *Boat2*, *Car2*, *Dog*, *Excavator*, *Ragged*, and *Trees2*



Beihang University

Modifications of LTIR

- VOT-TIR2015 was already saturated
- Call for sequences – limited success (3 new sources, too easy)
- Easiest sequences have been removed: *Crossing*, *Horse*, and *Rhino behind tree*
- New, more *Bird*, *Boat1*, *Boat2*, *Car2*, *Dog*,



Beihang University

Properties

- 25 Sequences
- Average sequence length 740
- Annotations in accordance with VOT
  - Bounding-box
  - 11 global attributes (per-sequence)

    Blur, dynamics change, temperature change, object motion, size change, camera motion, background clutter, aspect ratio change, object deformation, scene complexity, neutral

  - 6 local attributes (per-frame)

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

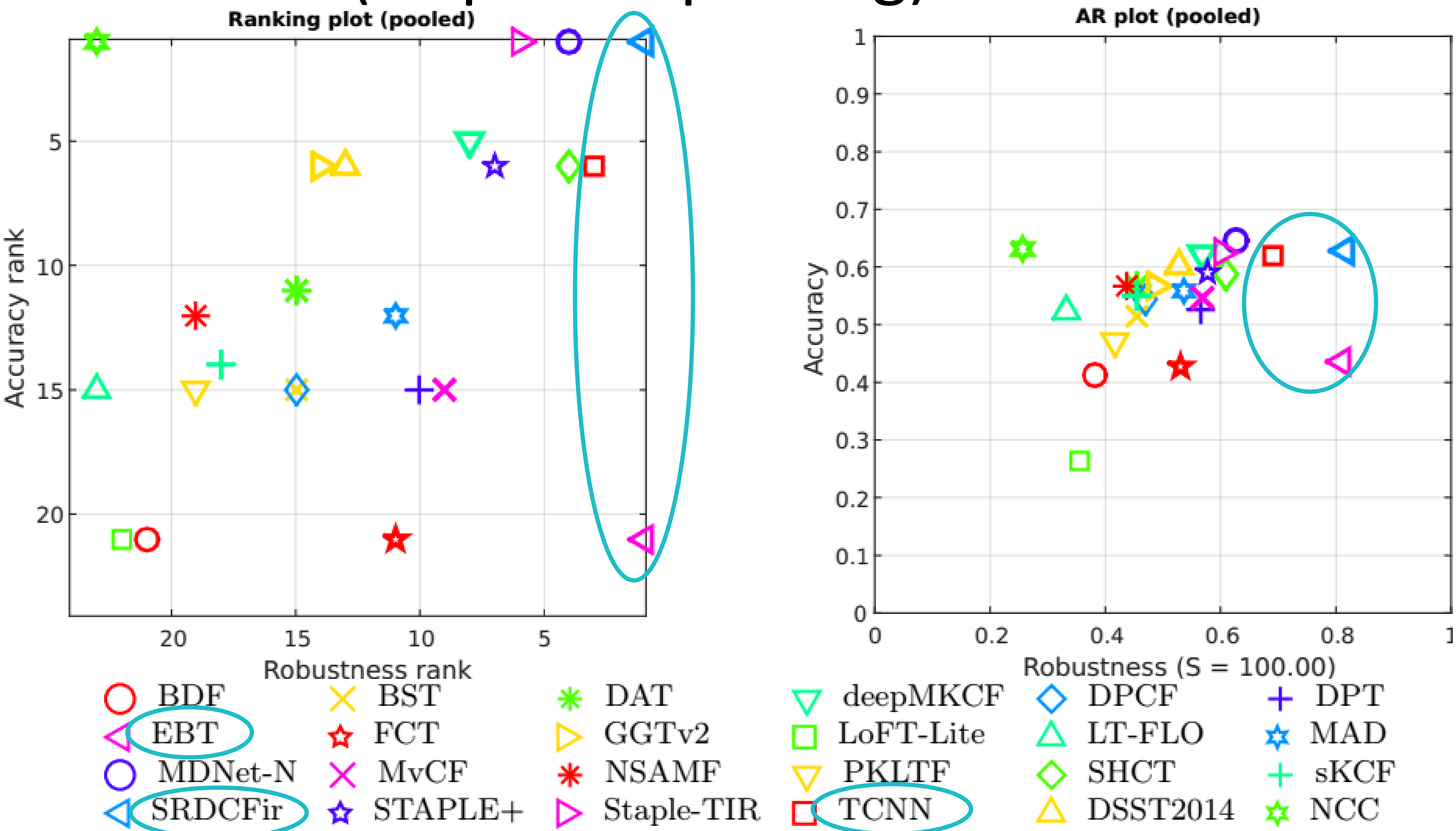
Performance evaluation measures

- Basically the same as VOT2016 (based on 8-bit)
  - accuracy
  - robustness
- Evaluated pooled and normalized per-attribute
  - raw value
  - rank
- Overall: expected average overlap
- Speed in EFO units

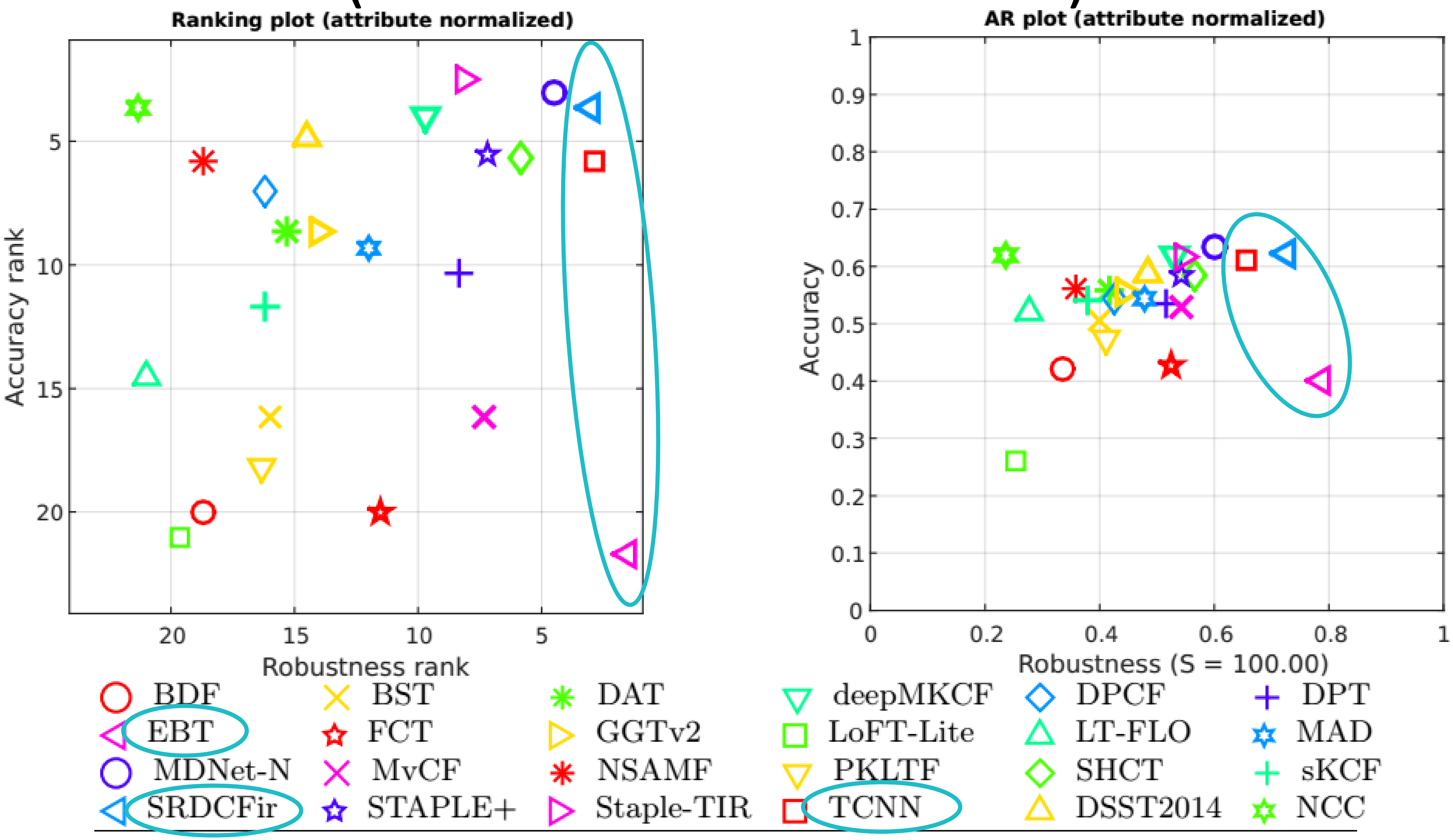
Results

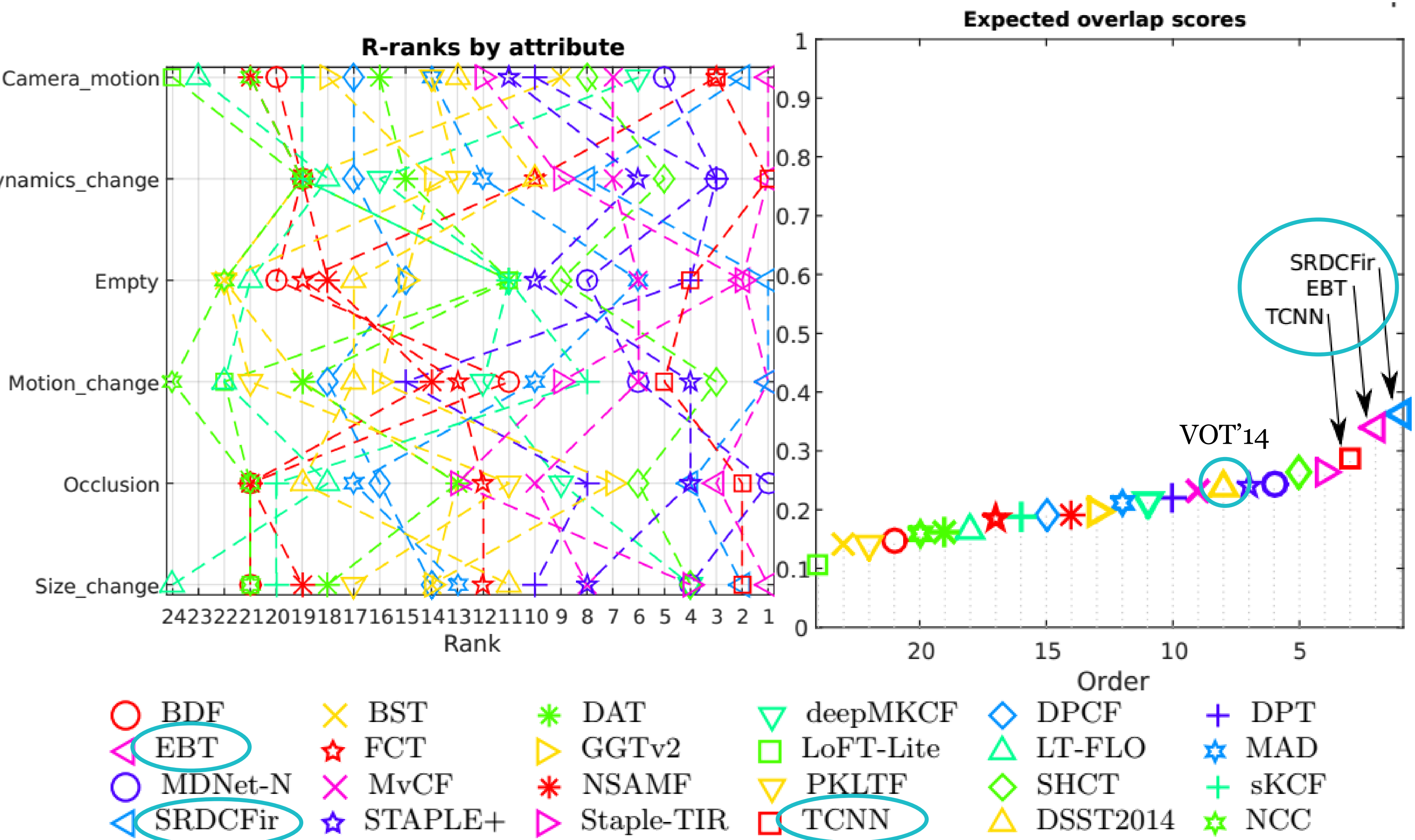
- 21 submitted trackers
- 3 added by VOT committee (NCC, DSST2014, SRDCFir)
- All 24 trackers in both challenges
- Various classes of trackers
  - 8 part-based trackers: BDF, BST, DPCF, DPT, FCT, GGTv2, LT-FLO, and SHCT
  - 7 trackers based on DCFs: DSST2014, MvCF, NSAMF, sKCF, SRDCFir, Staple-TIR, and STAPLE+
  - 3 trackers based on deep features/learning: deepMKCF, TCNN, and MDNet-N
  - 2 fusion based trackers: MAD and LOFT-Lite
  - 4 other: EBT, PKLTF, DAT, and NCC

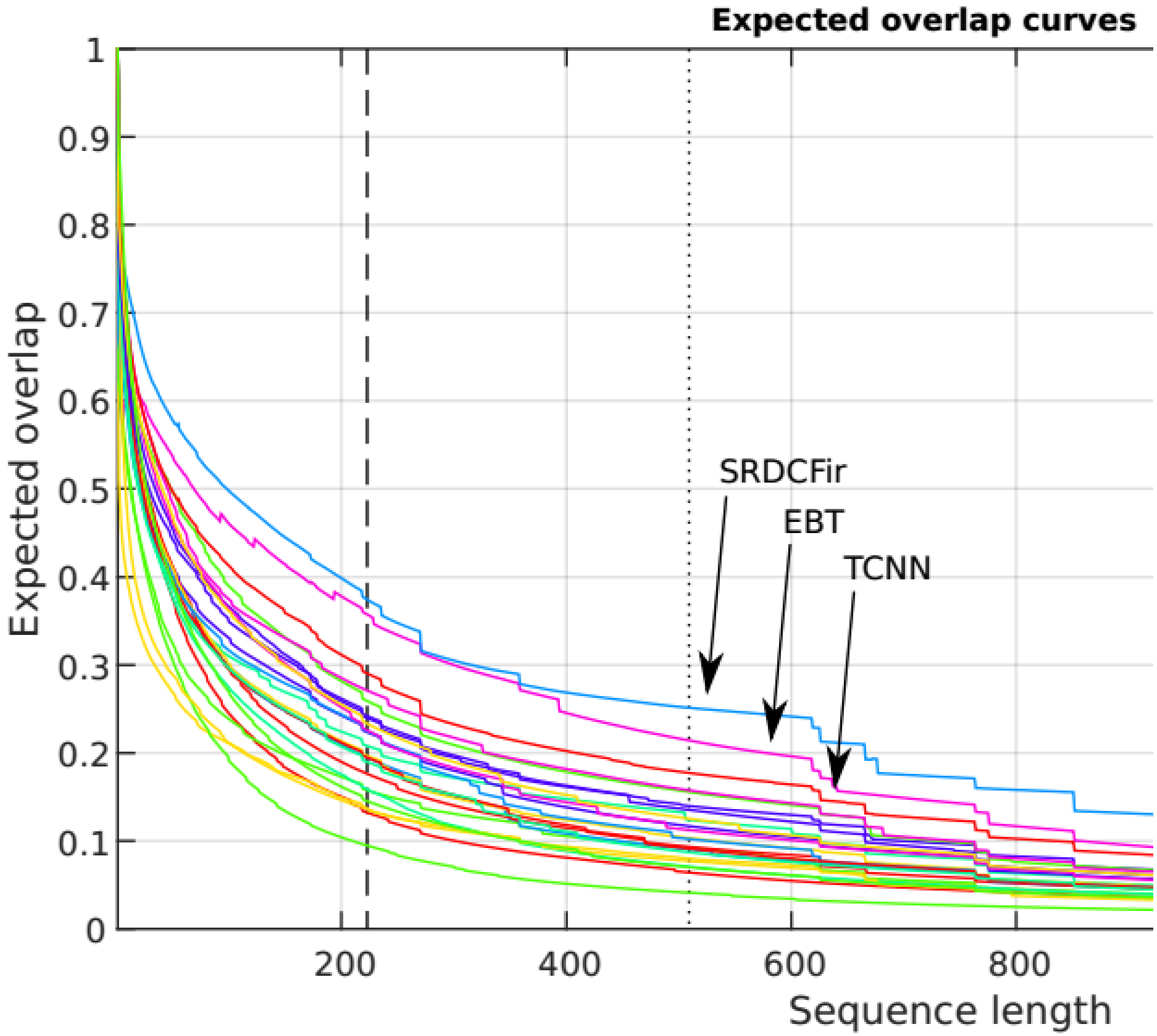
Results (sequence pooling)



Results (attribute normalization)



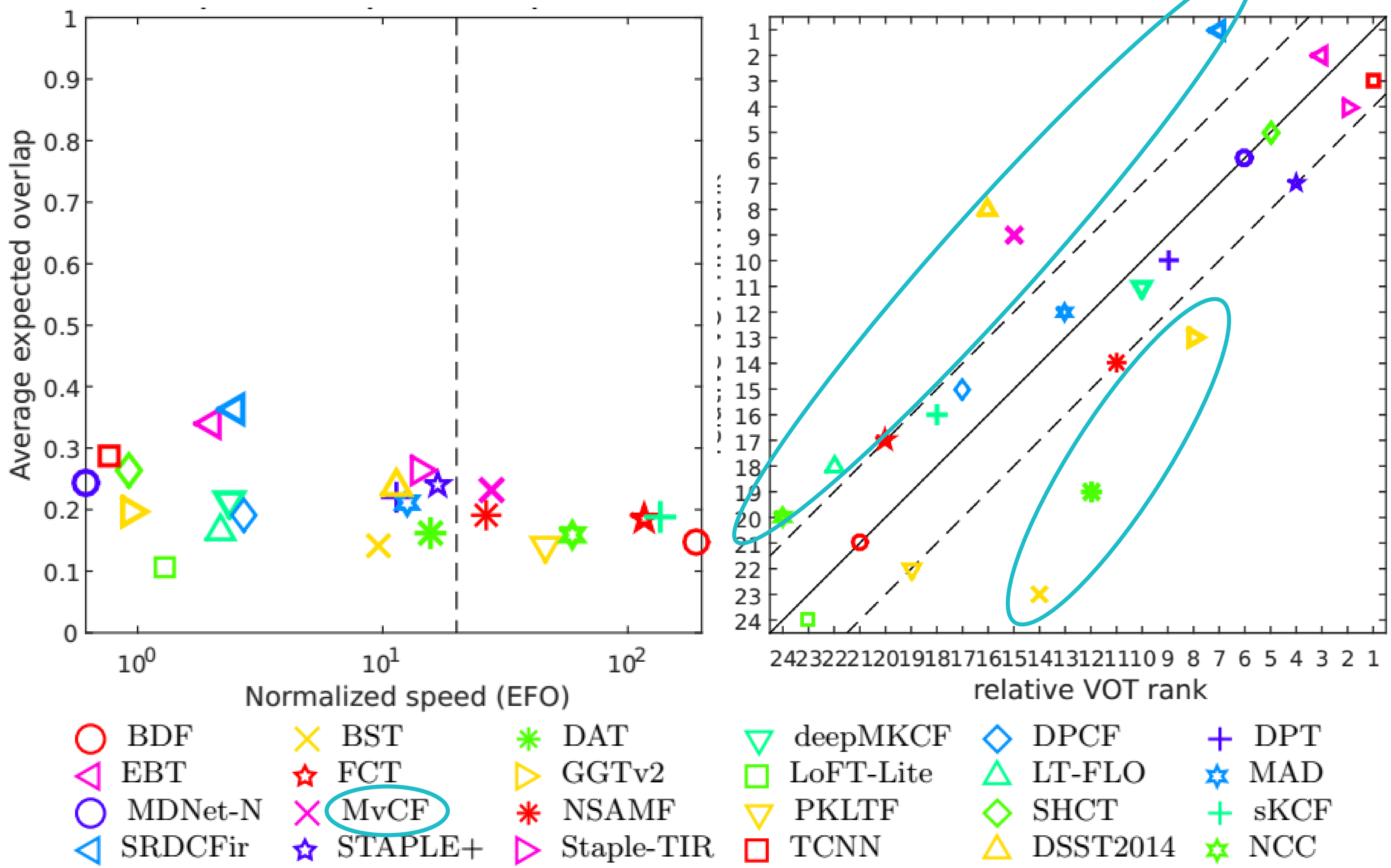

R-ranks by attribute
Camera_motion
Dynamics_change
Empty
Motion_change
Occlusion
Size_change
24 23 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1
Rank
Expected overlap scores
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
20
15
10
5
Order
VOT'14
SRDCFir
EBT
TCNN
BDF
BST
DAT
deepMKCF
DPCF
DPT
EBT
FCT
GGTv2
LoFT-Lite
LT-FLO
MAD
MDNet-N
MvCF
NSAMF
PKLTF
SHCT
sKCF
SRDCFir
STAPLE+
Staple-TIR
TCNN
DSST2014
NCC



Tracker	EAO	A	R
SRDCFir*	0.364	0.63	0.82
EBT*	0.340	0.43	0.81
TCNN*	0.287	0.62	0.69
Staple-TIR*	0.264	0.63	0.60
SHCT*	0.263	0.59	0.61
MDNet-N*	0.243	0.65	0.63
STAPLE+*	0.241	0.59	0.58
DSST2014*	0.236	0.60	0.53
MvCF*	0.231	0.55	0.57
DPT*	0.219	0.53	0.57
deepMKCF	0.213	0.62	0.57
MAD*	0.211	0.56	0.54
GGTv2*	0.197	0.57	0.49
NSAMF*	0.192	0.57	0.44
DPCF*	0.191	0.54	0.47
sKCF*	0.188	0.55	0.46
FCT*	0.186	0.43	0.53
LT-FLO	0.163	0.52	0.33
DAT*	0.162	0.57	0.46
NCC*	0.160	*0.63*	0.26
BDF*	0.147	0.41	0.38
PKLTF*	0.141	0.47	0.42
BST*	0.140	0.51	0.46
LoFT-Lite*	0.107	0.26	0.36

LINKÖPING UNIVERSITY

Felsberg et al., VOT-TIR2016 results



Sequence ranking

- A_f: average number of trackers that failed per frame
- M_f: maximum number of trackers that failed at a single frame

Sequence	Score
Bird	1,5
Quadrocopter2	1,5
Trees2	1,5
Car2	2
Crowd	2
Garden	2
Quadrocopter	2,5
Ragged	2,5
Excavator	3
Boat2	3,5
Crouching	3,5
Dog	3,5
Mixed_distractors	3,5
Selma	3,5
Street	3,5
Trees1	3,5
Boat1	4
Jacket	4
Birds	4,5
Car1	4,5
Saturated	4,5
Soccer	4,5
Depthwise_crossing	5
Hiding	5
Running_rhino	5

challenging:

0.08<=A_f
17<=M_f

intermediate:

0.03<=A_f<=0.08
4<=M_f<=11

easiest:

0.01<=A_f<=0.05
2<=M_f<=7

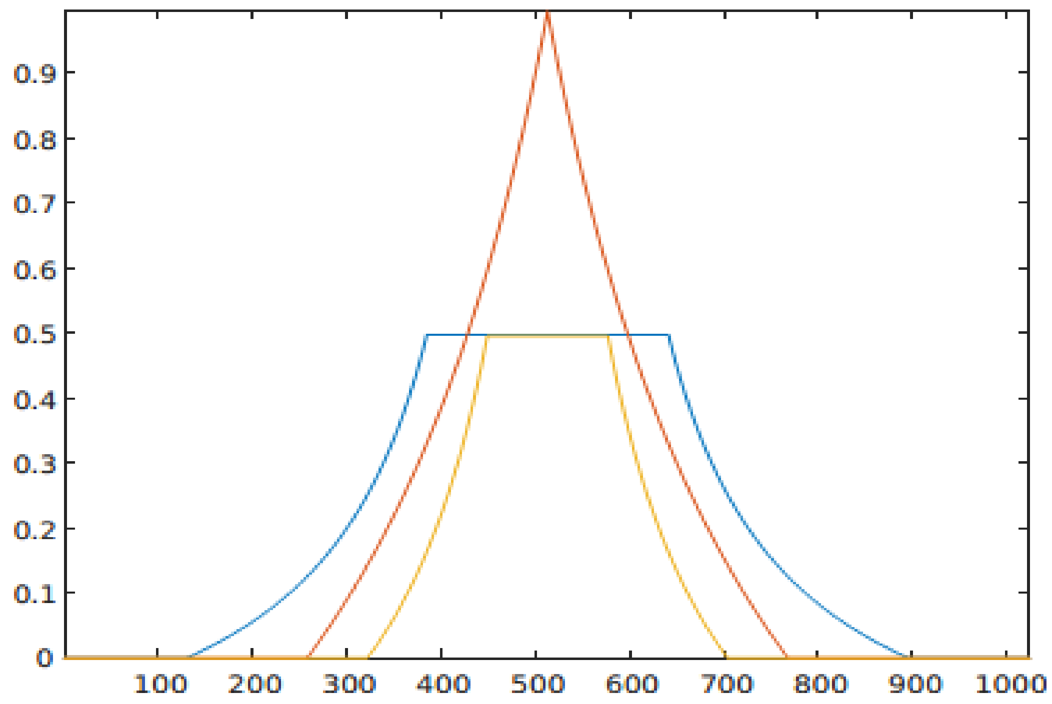
Problems with Overlap Measure

- Systematic overestimation of the bounding box
  - avoids failures
  - at cost of moderate accuracy degradation
- Paper suggests a new, quantization-based criterion

$$|A_t^G \cap A_t^T| \frac{|A_t^G|}{|A_t^T|} > \frac{1}{2}$$

Bias in Overlap Precision Measures

- assume:
  - 1D case
  - position equally distributed
- expectation
  - GT: 0.19
  - 0.5xGT: 0.11
  - 2xGT: 0.21



Summary

- More difficult challenge with LTIR2016 dataset
- 1/3 of trackers show different ranking than in VOT2016
  - in contrast to VOT-TIR2015
- Top-performing triple: SRDCFir, EBT, TCNN
- Best real-time method: MvCF
- Issues with the overlap measure
- Available at http://www.votchallenge.net/vot2016

Winners of the VOT-TIR2016 Challenge:

Gao Zhu, Fatih Porikli,
and Hongdong Li:

Edge Box Tracker (EBT)

Presentation at VOT2016
right after this talk

Award sponsored by

SICK
Sensor Intelligence.

VOT
visual object tracking challenge

Thanks

- The VOT2016 Committee
- Everyone who participated:

Alfredo Petrosino, Alvaro Garcia-Martin, Andres Solis Montero, Anton Varfolomieiev, Aykut Erdem, Bohyung Han, Chang-Ming Chang, Dawei Du, Erkut Erdem, Fahad Shahbaz Khan, Fatih Porikli, Fei Zhao, Filiz Bunyak, Francesco Battistone, Gao Zhu, Guna Seetharaman, Hongdong Li, Honggang Qi, Horst Bischof, Horst Possegger, Hyeonseob Nam, Jack Valmadre, Jianke Zhu, Jiayi Feng, Jochen Lang, Jose M. Martinez, Kannappan Palaniappan, Karel Lebeda, Ke Gao, Krystian Mikolajczyk, Longyin Wen, Luca Bertinetto, Mahdieh Poostchi, Mario Maresca, Martin Danelljan, Michael Arens, Ming Tang, Mooyeol Baek, Nana Fan, Noor Al-Shakarji, Ondrej Miksik, Osman Akin, Philip H. S. Torr, Qingming Huang, Rafael Martin-Nieto, Rengarajan Pelapur, Richard Bowden, Robert Laganiere, Sebastian B. Krah, Shengkun Li, Shizeng Yao, Simon Hadfield, Siwei Lyu, Stefan Becker, Stuart Golodetz, Tao Hu, Thomas Mauthner, Vincenzo Santopietro, Wenbo Li, Wolfgang Hubner, Xin Li, Yang Li, Zhan Xu, and Zhenyu He