VOT
visual object tracking

The Visual Object Tracking Challenge Results
VOT-RGBT 2019

Michael Felsberg, Matej Kristan, Aleš Leonardis, Jiri Matas, Roman Pflugfelder, Joni-Kristian Kämäräinen, Amanda Berg, Abdelrahman Eldesokey, Luka Čehovin, Gustavo Fernandez, Alan Lukežič, Ondrej Drbohlav et al.

University of Ljubljana
Faculty of Computer and Information Science

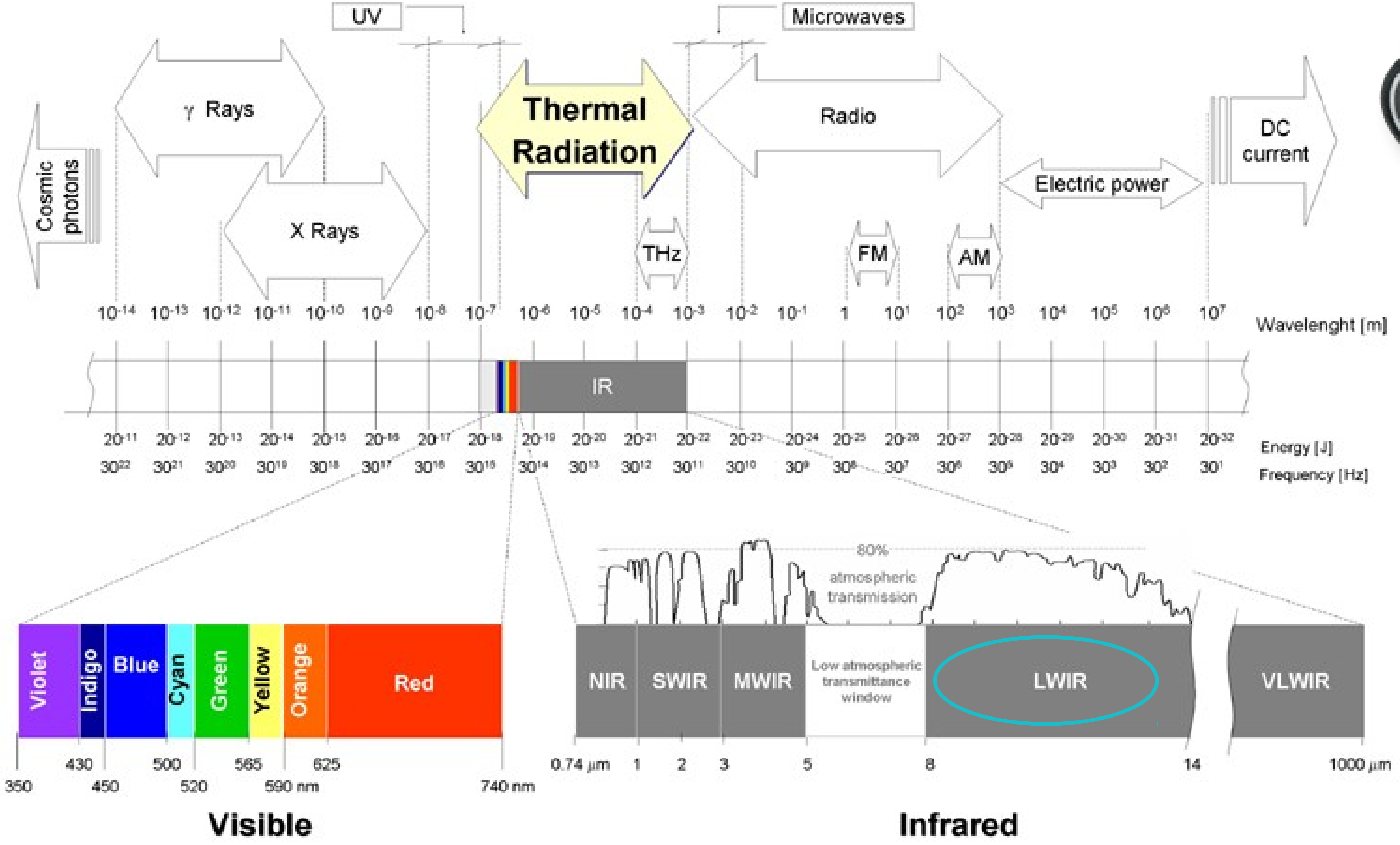
UNIVERSITY OF BIRMINGHAM

LINKÖPING UNIVERSITY

AUSTRIAN INSTITUTE OF TECHNOLOGY

Tampere University

Why adding Thermal Image Modality?



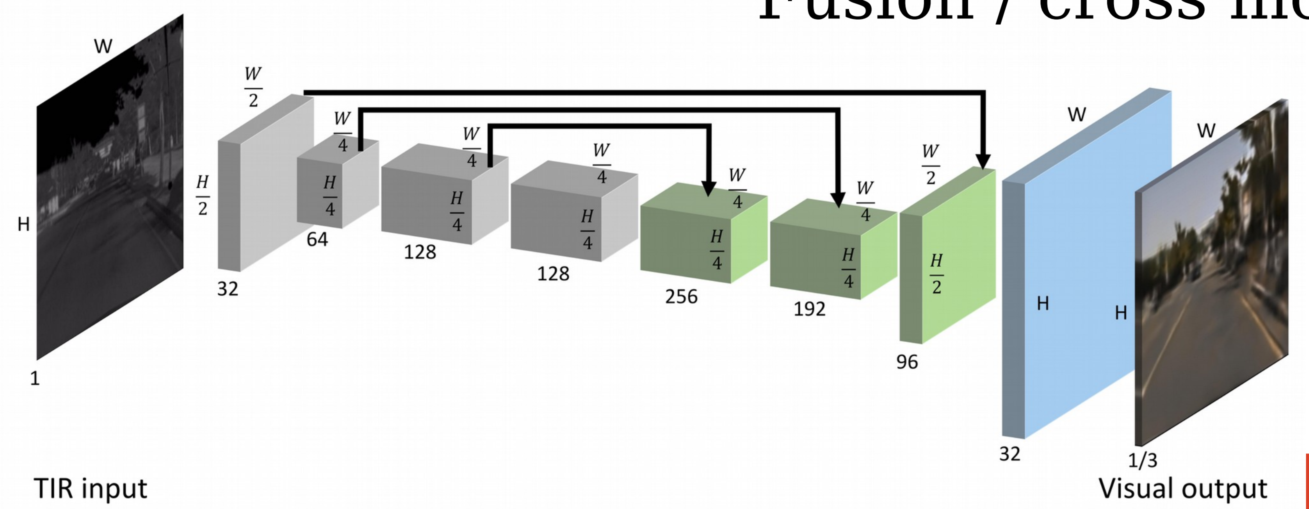
http://www.theses.ulaval.ca/2005/23016/apb.html

Applications of TIR

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

Challenges

- Interpretation of TIR images
  - TIR2RGB
- Tracking: RGB and TIR
  - Calibration and registration
  - Understanding the similarities and complementaries (VOT-TIR)
  - Fusion / cross modality (VOT-RGBT)



Pre-VOT datasets for tracking in TIR

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

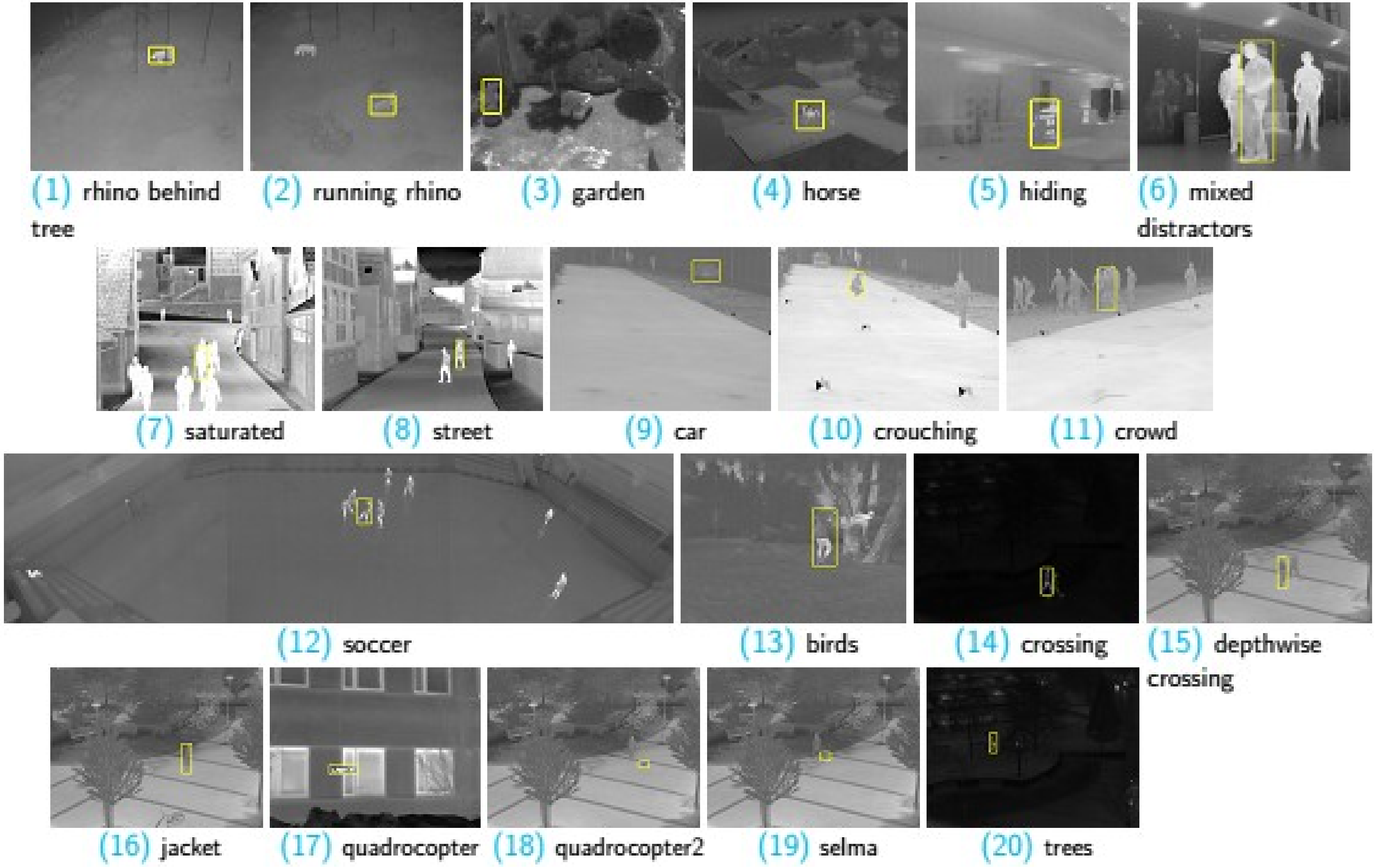
Why a separate challenge?

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

Many similarities but also interesting differences

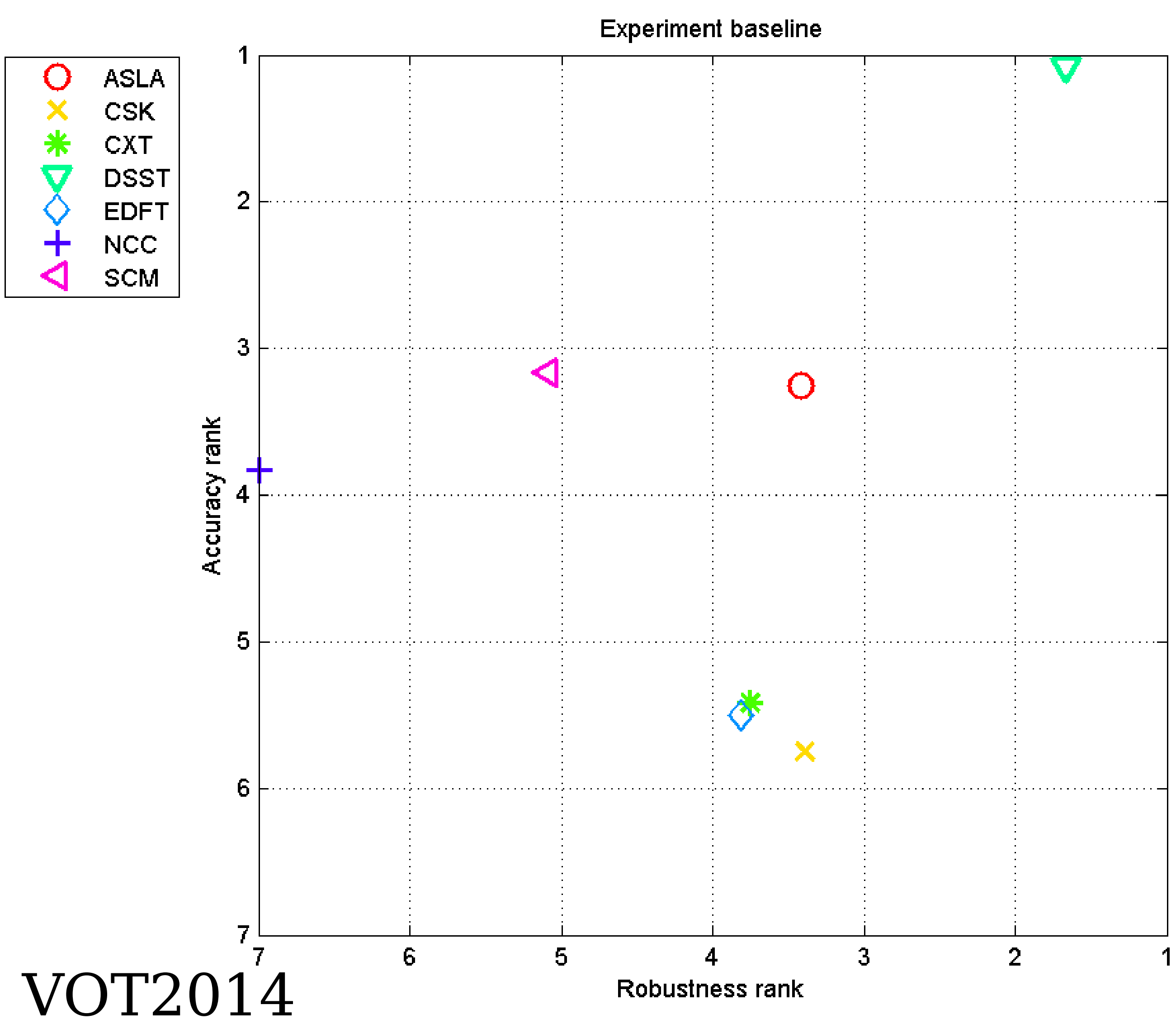
- 16-bit
- Constant values if radiometric
- Less structure/edges/texture
- No shadows
- Noise: blooming, resolution, dead pixels

Towards VOT-TIR: Linköping Thermal InfraRed (LTIR) dataset

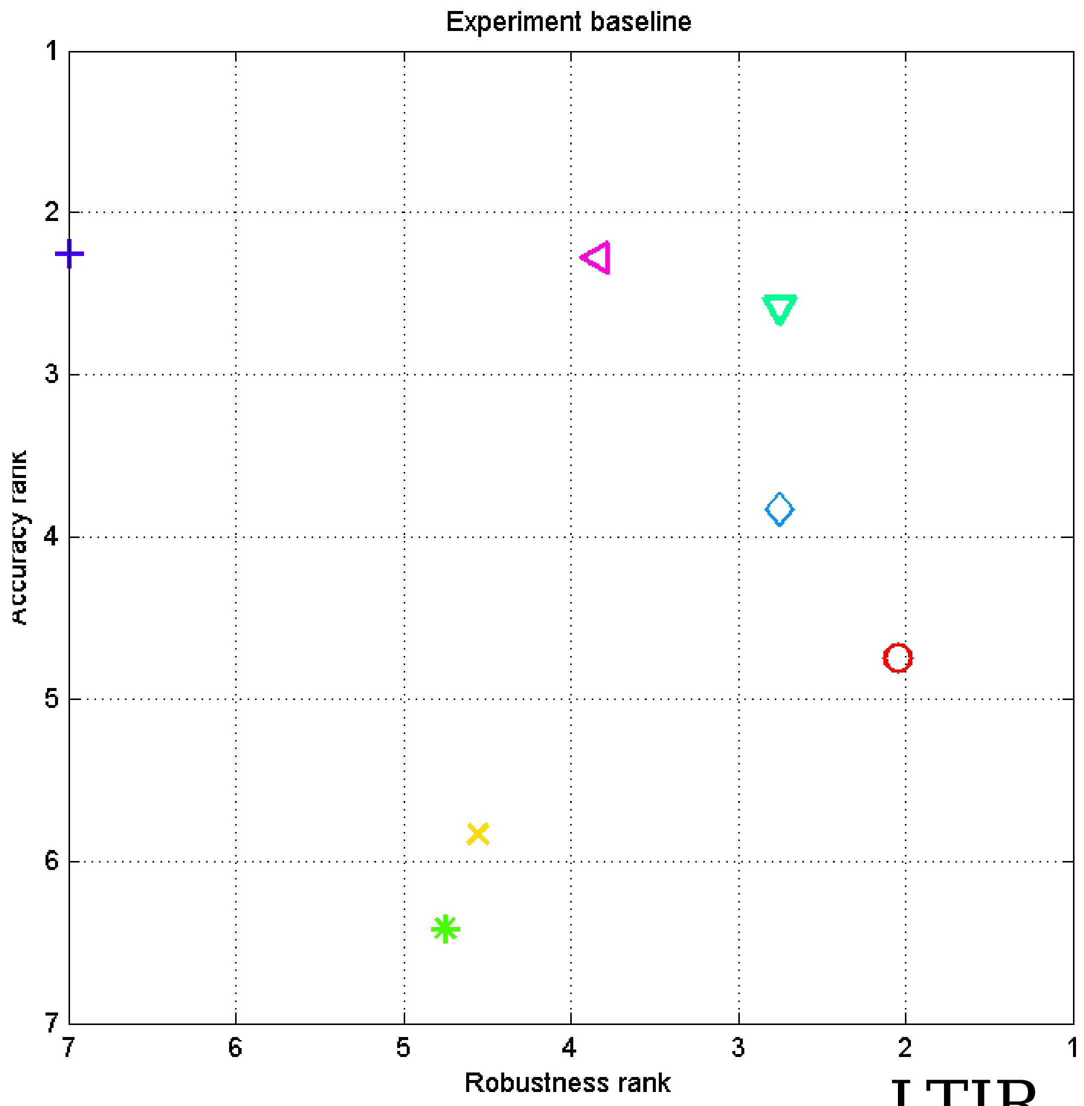


A. Berg,
J. Ahlberg,
M. Felsberg,
A Thermal Object Tracking Benchmark.
AVSS 2015.

Will it be different? Test against VOT2014

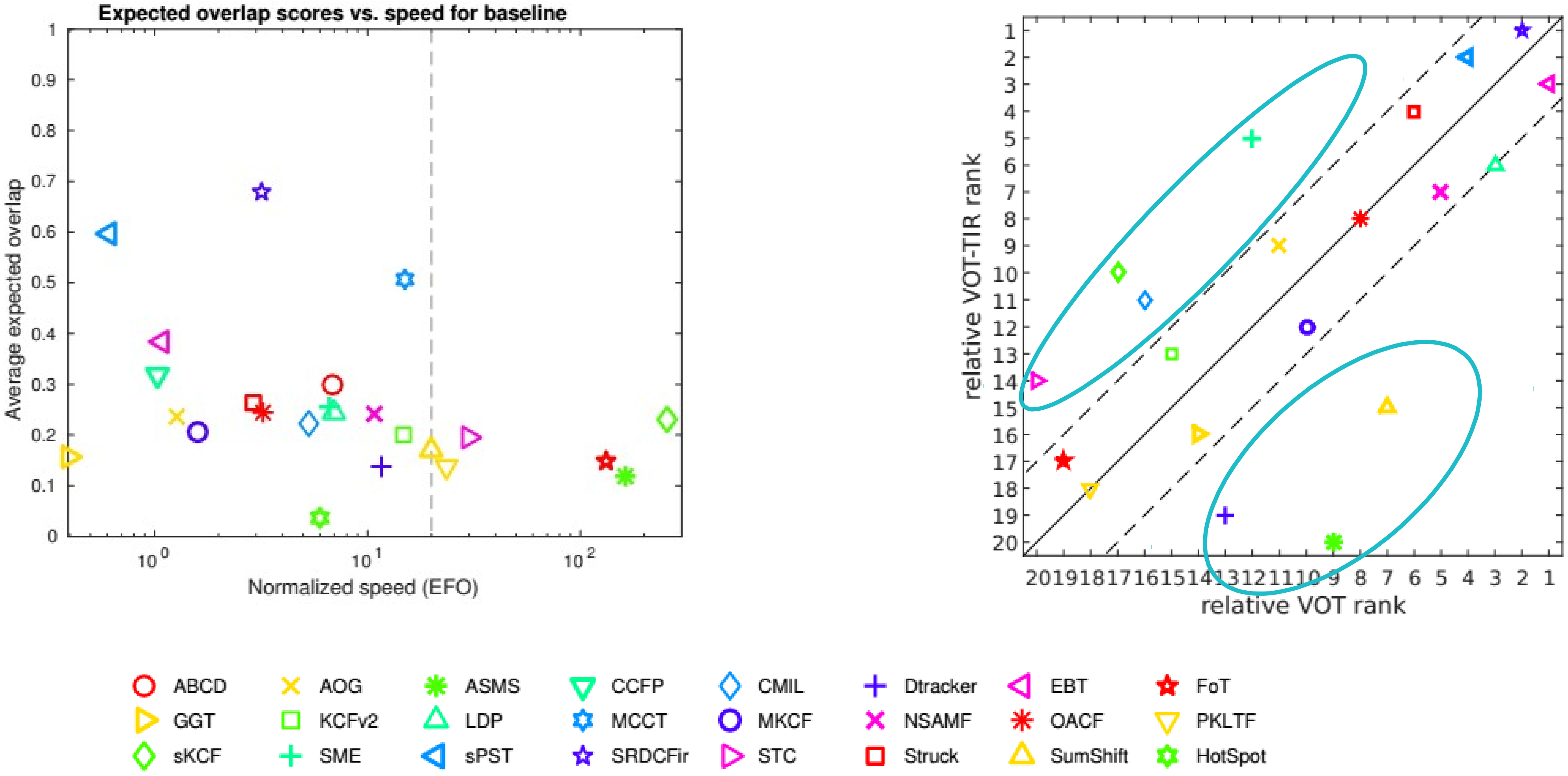


VOT2014



LTIR

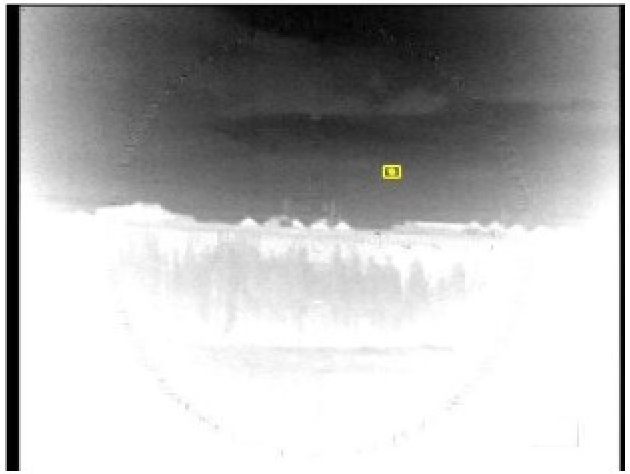
VOT2015 vs VOT-TIR2015



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



Properties

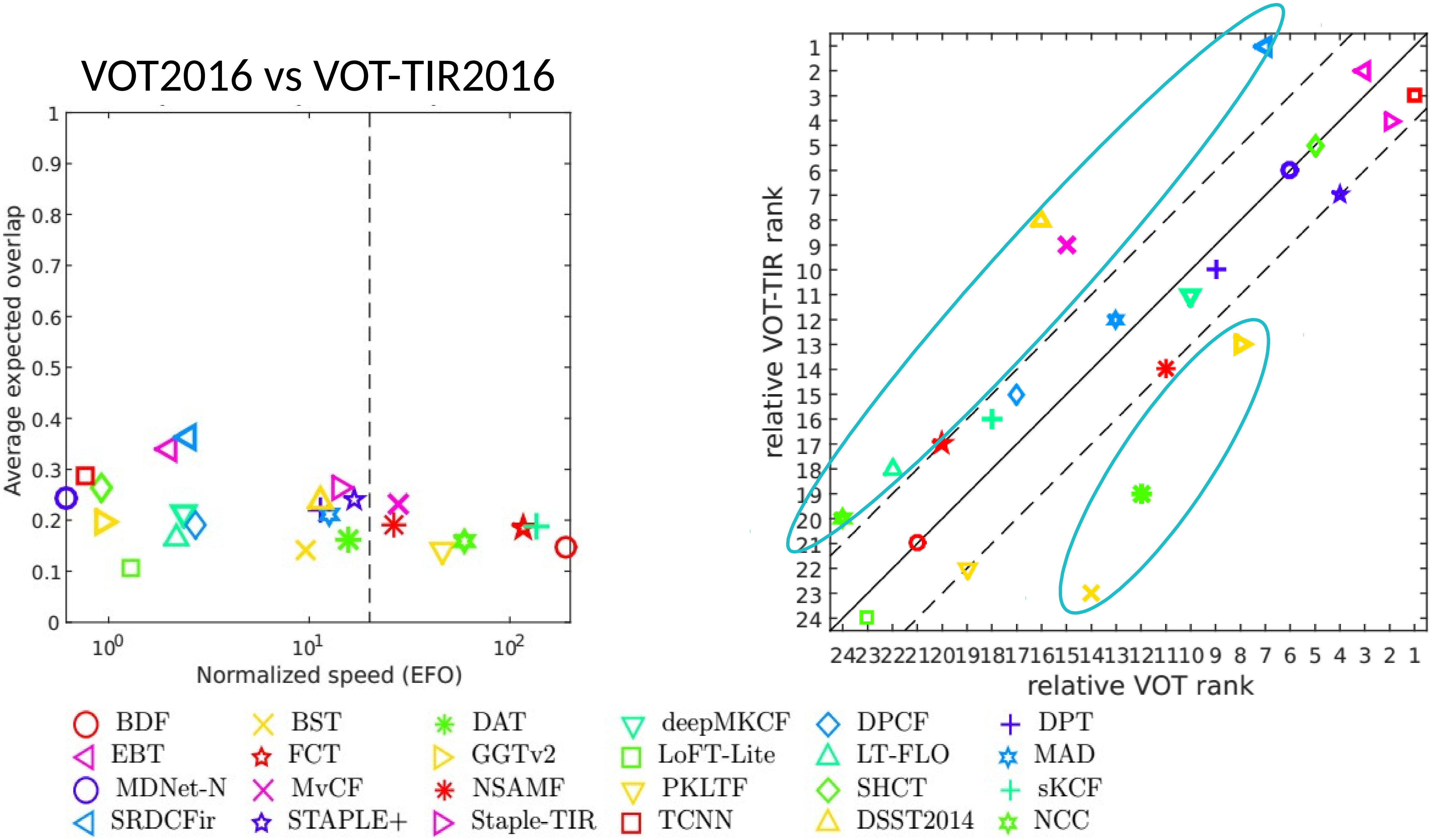
- 25 Sequences
- Average sequence length 740
- Annotations in accordance with VOT
  - Bounding-box
  - 11 global attributes (per-sequence)

    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

VOT2016 vs VOT-TIR2016



RGBT-dataset

- RGBT234-dataset from: C. Li, X. Liang, Y. Lu, N. Zhao, and J. Tang. RGB-T object tracking: Benchmark and baseline. Pattern Recognition (96), 2019
- 234 sequences with an average length of 335 frames
- Same clustering in 11-dim attribute space, but now 60 sequences
- Local attribute illumination/dynamics change not used
- Original axis-aligend annotation has been replaced with new rotated bboxes

Issues

- Spatial accuracy (addressed by re-annotation)
- Synchronization (considered part of challenge)

Semi-automatic (re-)annotation

- Procedure described in paper #2:
  A. Berg et al. *Semi-automatic annotation of objects in visual-thermal video*.
- Step 1: semi-automatic video segmentation based on: J. Johnander et al. *A generative appearance model for end-to-end video object segmentation*. In CVPR, 2019.
- Step 2: bounding box determination: T. Vojir and J. Matas. *Pixel-wise object segmentations for the VOT 2016 dataset*. Research Report CTU–CMP–2017–01.

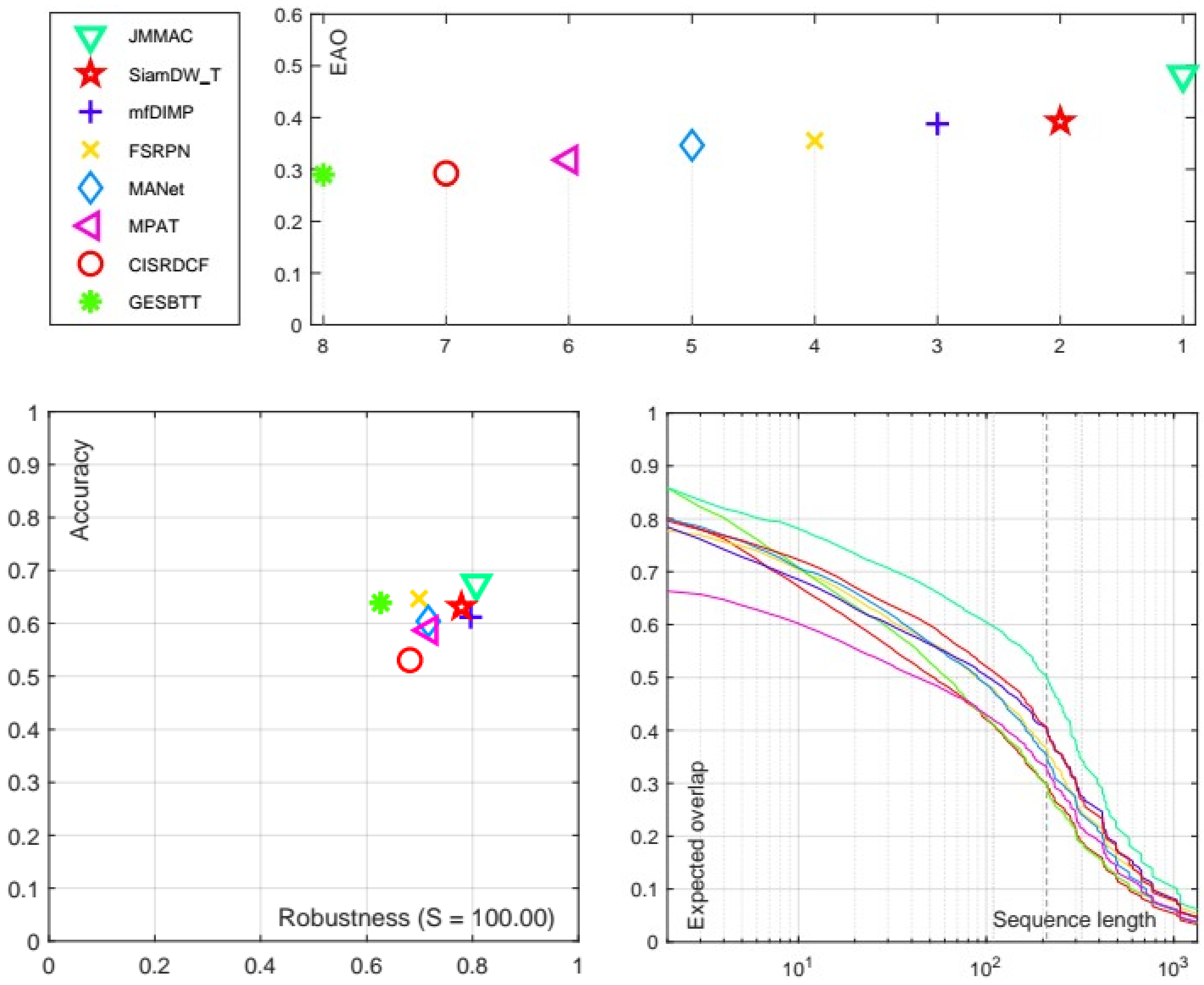
- Synchronization issue: TIR is used as reference
- Spatial accuracy: EAO RGB-TIR 0.75
- Evaluation is performed in the same way as for VOT-ST 2019
- Top-ranked trackers on the public dataset run by the committee on the sequestered dataset
- Top-ranked tracker on the sequestered dataset is the winner

Submitted tracker

- 10 trackers in total, 8 unique submissions with code
  - 5 ST_1, 3 ST_0
  - 7 uniform dynamic model, 1 random walk
  - 4 trackers based on discriminative correlation filters: CISRDCF, GESBTT, JMMAC, and mfDiMP
  - 4 trackers based on multiple CNNs: MANet, mfDiMP, MPAT, and SiamDW_T
  - 4 trackers make use of Siamese CNNs: FSRPN, mfDiMP, MPAT, and SiamDW_T
  - 2 trackers apply a Kalman filter: GESBTT and JMMAC
  - 1 tracker makes use of optical flow: GESBTT
  - 1 tracker makes use of ransac: JMMAC
  - 5 trackers use combinations of several features
  - 6 trackers use CNN features
  - 3 trackers use hand-crafted features
  - 2 trackers use keypoints
  - 2 trackers use grayscale features

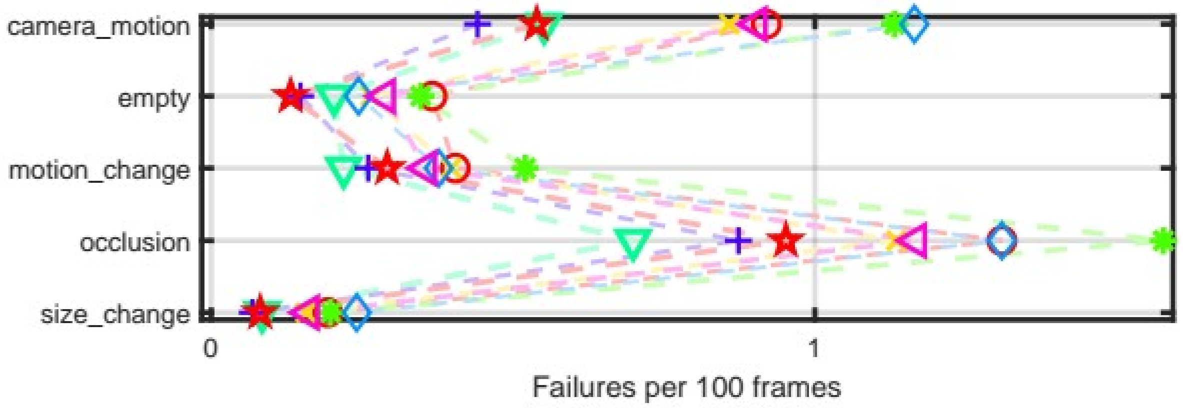
Results on public dataset

- All top-5 trackers use CNN features
- Respectively 3 out of these 5 trackers use
  - DCFs
  - Multiple CNNs
  - Siamese CNNs
  - JMMAC is working significantly better than the other two DCF-based trackers – RANSAC reason?



Further results

- EAO is stronger correlation to robustness than accuracy
- Robustness is most challenging for occlusion and camera motion
- Changed order for sequestered dataset



	Tracker	EAO	A	R
1.	mfDiMP	0.2347 ①	0.6133	0.3160 ①
2.	SiamDW_T	0.2143 ②	0.6515 ②	0.2714 ②
3.	MANet	0.2041 ③	0.5784	0.2592 ③
4.	JMMAC	0.2037	0.6337 ③	0.2441
5.	FSRPN	0.1873	0.6561 ①	0.1755

VOT-ST2019 Winners

Winners of the VOT-RGBT 2019 challenge:

mfDiMP by: L. Zhang, A. Gonzalez-Garcia, J. van de Weijer

"Multi-modal fusion for end-to-end RGB-T tracking"

(The talk up next!)

Summary

- CNN-features dominating
- The ranking changes on sequestered dataset
- Overall performance decreases on sequestered dataset
- Robustness most important
- Occlusion and camera motion largest challenges
- For the future:
  - Attract more participants
  - Measure the effect of spatial missalignment and synchronization errors?
  - Potential other changes in the evaluation system

Thanks

- The VOT2019 committee


M. Kristan


J. Matas


A. Leonardis


M. Felsberg


R. Pflugfelder


G. Fernandez


L. Čehovin


A. Lukežič


A. Eldesokey

- Everyone who participated or contributed

Matej Kristan1, Jiˇr´ı Matas2, Aleˇs Leonardis3, Michael Felsberg4, Roman Pflugfelder5,6, Joni-Kristian Kamarainen7, Luka Cˇ ehovin Zajc1, Ondrej Drbohlav2, Alan Lukezˇicˇ1, Amanda Berg4,8, Abdelrahman Eldesokey4, Jani K¨apyl¨a7, Gustavo Fern´andez5, Abel Gonzalez-Garcia18, Alireza Memarmoghadam50, Andong Lu9, Anfeng He52, Anton Varfolomieiev37, Antoni Chan17, Ardhendu Shekhar Tripathi23, Arnold Smeulders45, Bala Suraj Pedasingu29, Bao Xin Chen58, Baopeng Zhang12, Baoyuan Wu43, Bi Li28, Bin He10, Bin Yan19, Bing Bai20, Bing Li16, Bo Li40, Byeong Hak Kim25,33, Chao Ma41, Chen Fang35, Chen Qian40, Cheng Chen38, Chenglong Li9, Chengquan Zhang10, Chi-Yi Tsai42, Chong Luo34, Christian Micheloni55, Chunhui Zhang16, Dacheng Tao54, Deepak Gupta45, Dejia Song28, Dong Wang19, Efstratios Gavves45, Eunu Yi25, Fahad Shahbaz Khan4,30, Fangyi Zhang16, Fei Wang40, Fei Zhao16, George De Ath49, Goutam Bhat23, Guangqi Chen40, Guangting Wang52, Guoxuan Li40, Hakan Cevikalp21, Hao Du34, Haojie Zhao19, Hasan Saribas22, Ho Min Jung33, Hongliang Bai11, Hongyuan Yu16,34, Houwen Peng34, Huchuan Lu19, Hui Li32, Jiakun Li12, Jianhua Li19, Jianlong Fu34, Jie Chen57, Jie Gao57, Jie Zhao19, Jin Tang9, Jing Li26, Jingjing Wu27, Jingtuo Liu10, Jinqiao Wang16, Jinqing Qi19, Jinyue Zhang57, John K. Tsotsos58, Jong Hyuk Lee33, Joost van de Weijer18, Josef Kittler53, Jun Ha Lee33, Junfei Zhuang13, Kangkai Zhang16, Kangkang Wang10, Kenan Dai19, Lei Chen40, Lei Liu9, Leida Guo59, Li Zhang51, Liang Wang16, Liangliang Wang28, Lichao Zhang18, Lijun Wang19, Lijun Zhou48, Linyu Zheng16, Litu Rout39, Luc Van Gool23, Luca Bertinetto24, Martin Danelljan23, Matteo Dunnhofer55, Meng Ni19, Min Young Kim33, Ming Tang16, Ming-Hsuan Yang46, Naveen Paluru29, Niki Martinel55, Pengfei Xu20, Pengfei Zhang54, Pengkun Zheng38, Pengyu Zhang19, Philip H.S. Torr51, Qi Zhang , Qiang Wang16,31, Qing Guo44, Radu Timofte23, Rama Krishna Gorthi29, Richard Everson49, Ruize Han44, Ruohan Zhang57, Shan You40, Shao-Chuan Zhao32, Shengwei Zhao16, Shihu Li10, Shikun Li16, Shiming Ge16, Shuai Bai13, Shuosen Guan59, Tengfei Xing20, Tianyang Xu32, Tianyu Yang17, Ting Zhang14, Tom´aˇs Voj´ı˜r47, Wei Feng44, Weiming Hu16, Weizhao Wang38, Wenjie Tang14, Wenjun Zeng34, Wenyu Liu28, Xi Chen60, Xi Qiu56, Xiang Bai28, Xiao-Jun Wu32, Xiao-Jun Wu32, Xiaoyun Yang15, Xier Chen57, Xin Li26, Xing Sun59, Xingyu Chen16, Xinmei Tian52, Xu Tang10, Xue-Feng Zhu32, Yan Huang16, Yanan Chen57, Yanchao Lian57, Yang Gu20, Yang Liu36, Yanjie Chen40, Yi Zhang59, Yinda Xu60, Yingming Wang19, Yingping Li57, Yu Zhou28, Yuan Dong13, Yufei Xu52, Yunhua Zhang19, Yunkun Li32, Zeyu Wang , Zhao Luo16, Zhaoliang Zhang14, Zhen-Hua Feng53, Zhenyu He26, Zhichao Song20, Zhihao Chen44, Zhipeng Zhang16, Zhirong Wu34, Zhiwei Xiong52, Zhongjian Huang57, Zhu Teng12, and Zihan Ni10

- VOT2019 sponsor:

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