VOT
visual object tracking

The Visual Object Tracking Challenge Results VOT-RGBT 2020

Michael Felsberg, Matej Kristan, Aleš Leonardis, Jiri Matas, Roman Pflugfelder, Joni-Kristian Kämäräinen, Martin Danelljan, Linbo He, Yushan Zhang, Luka Čehovin Zajc, Alan Lukežič, Ondrej Drbohlav, Song Yan, Jinyu Yang, Gustavo Fernandez et al.

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

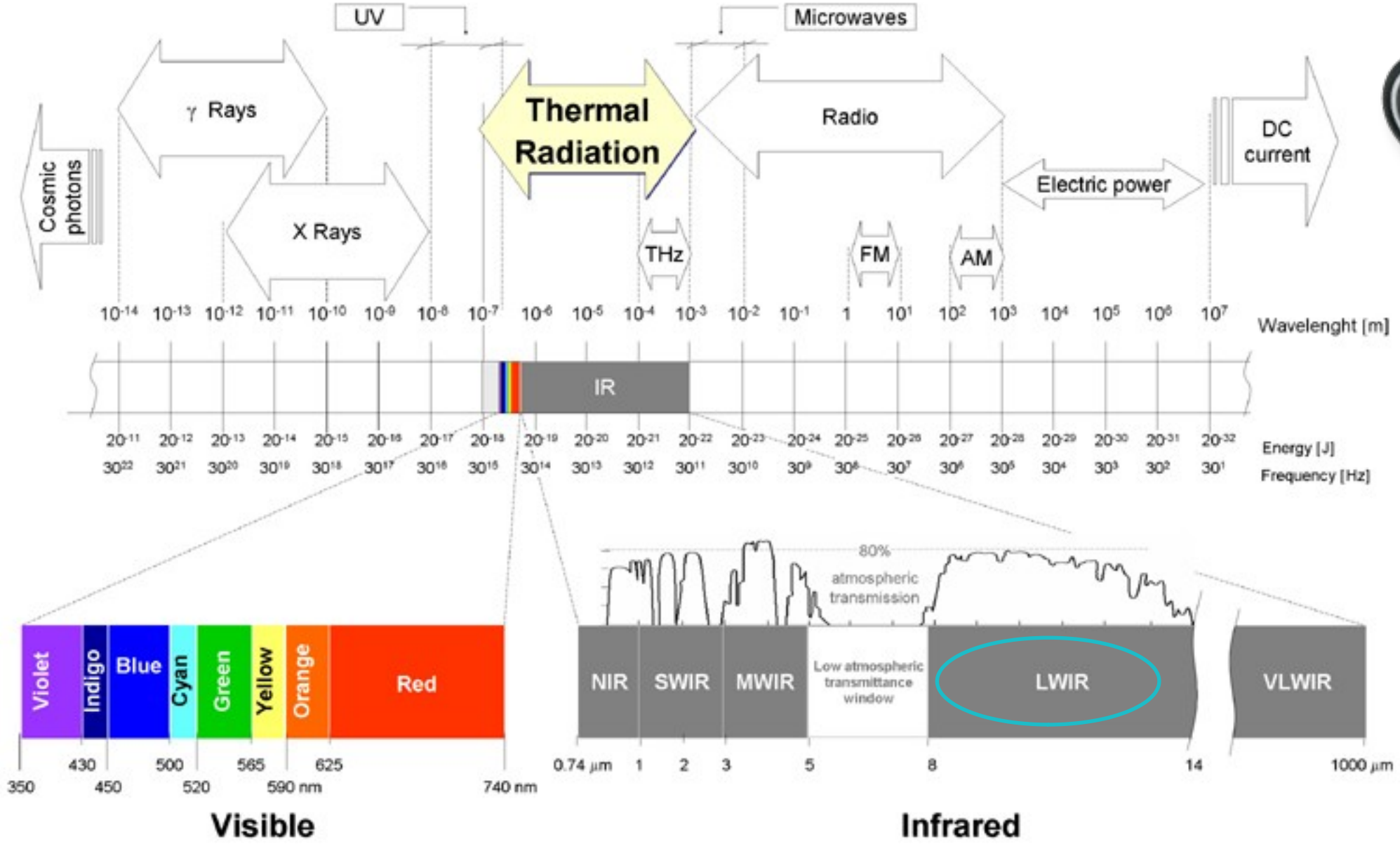
UNIVERSITY OF BIRMINGHAM

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

Tracking Challenges

- RGB and TIR - Calibration and registration
- Understanding the similarities and complementaries (VOT-TIR)
- Fusion / cross modality (VOT-RGBT)

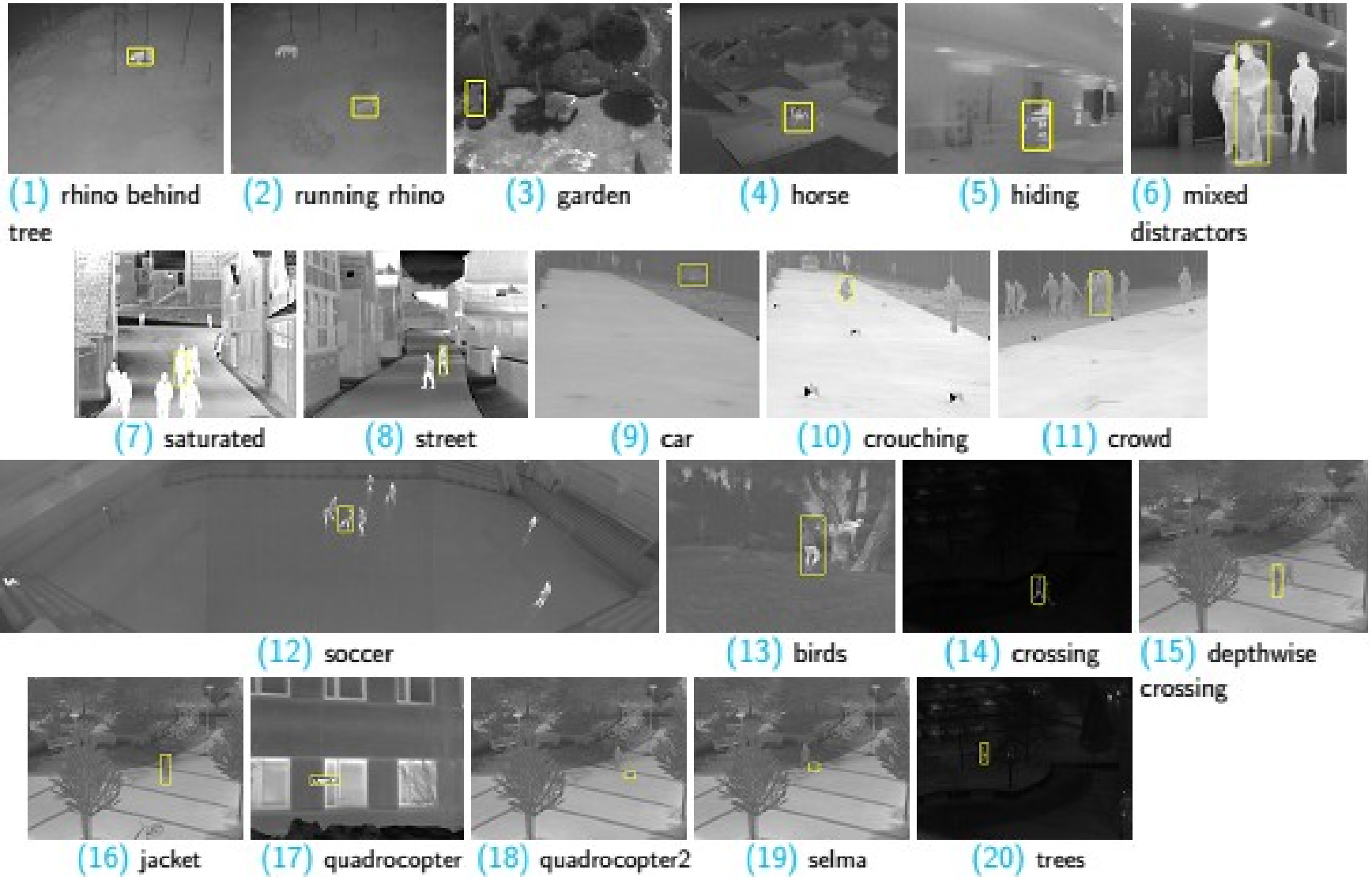
Why a separate challenge?

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

Many similarities but also interesting differences

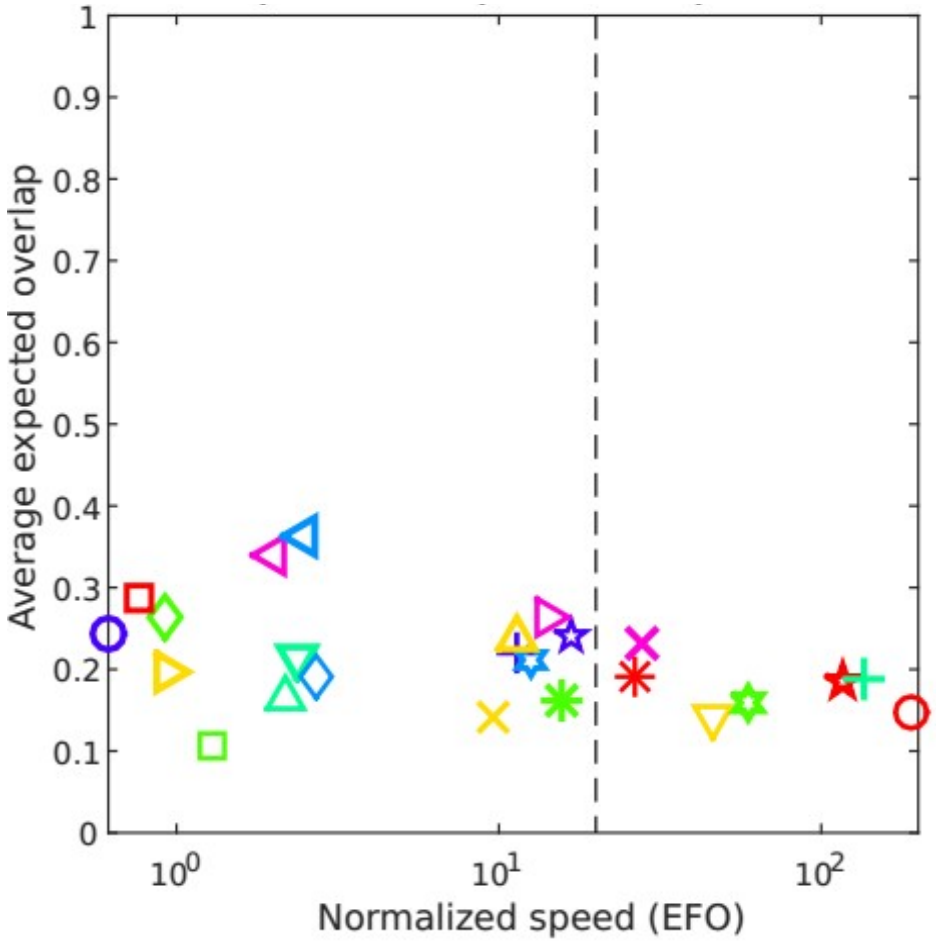
- 16-bit
- Constant values if radiometric
- Less structure/edges/texture
- No shadows
- Noise: blooming, resolution, dead pixels

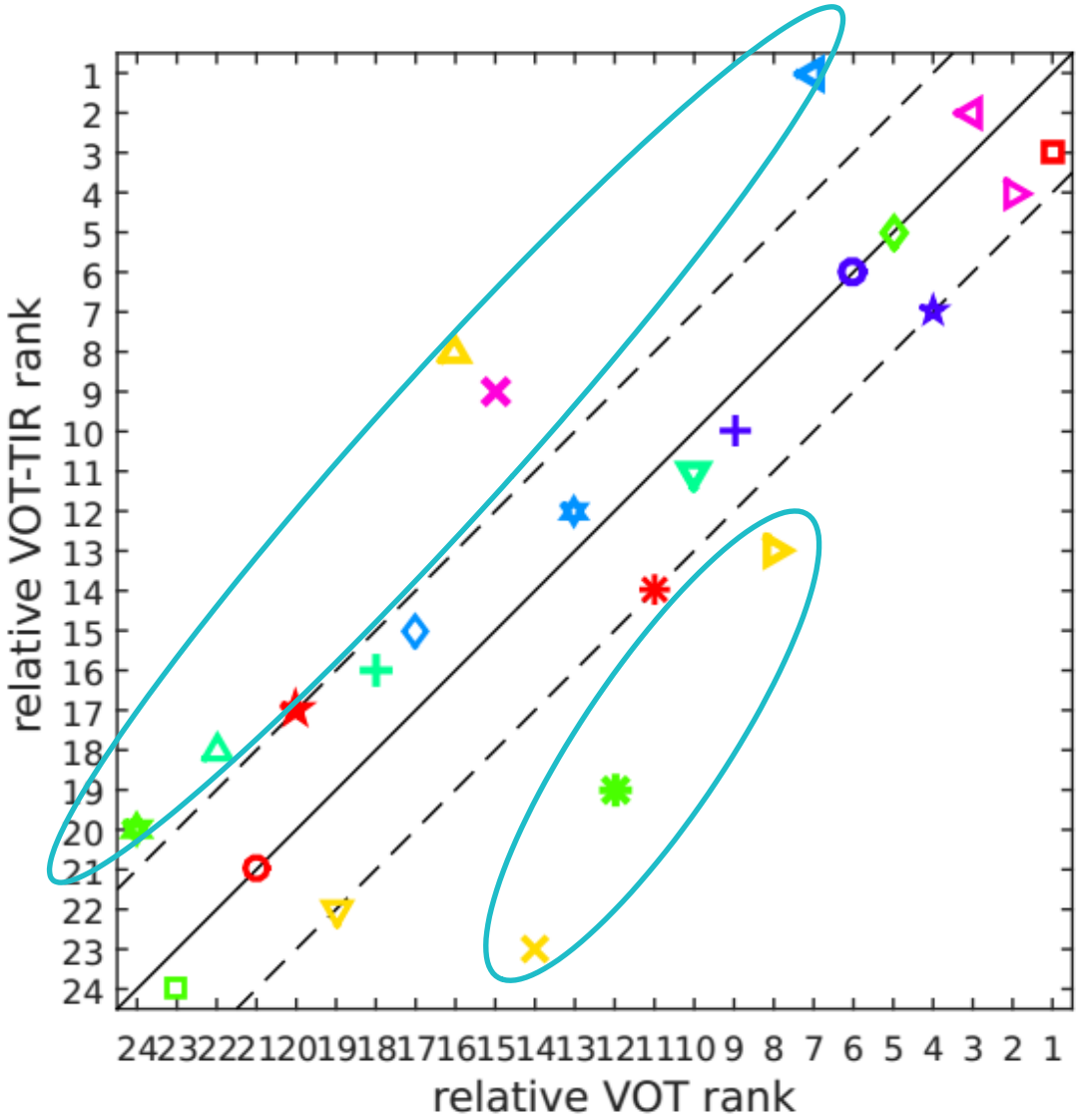
VOT-TIR: Linköping Thermal InfraRed (LTIR) dataset



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

VOT2016 vs VOT-TIR2016





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

Properties TIR

- 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

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

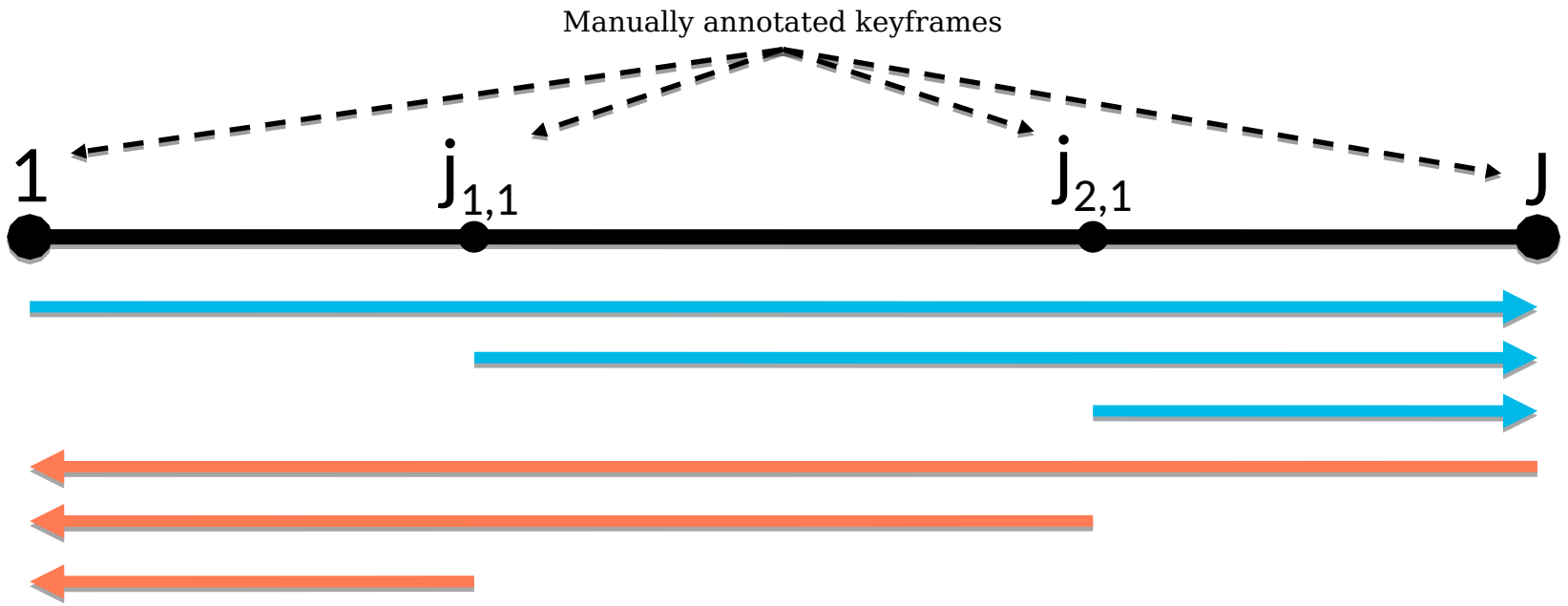
Issues

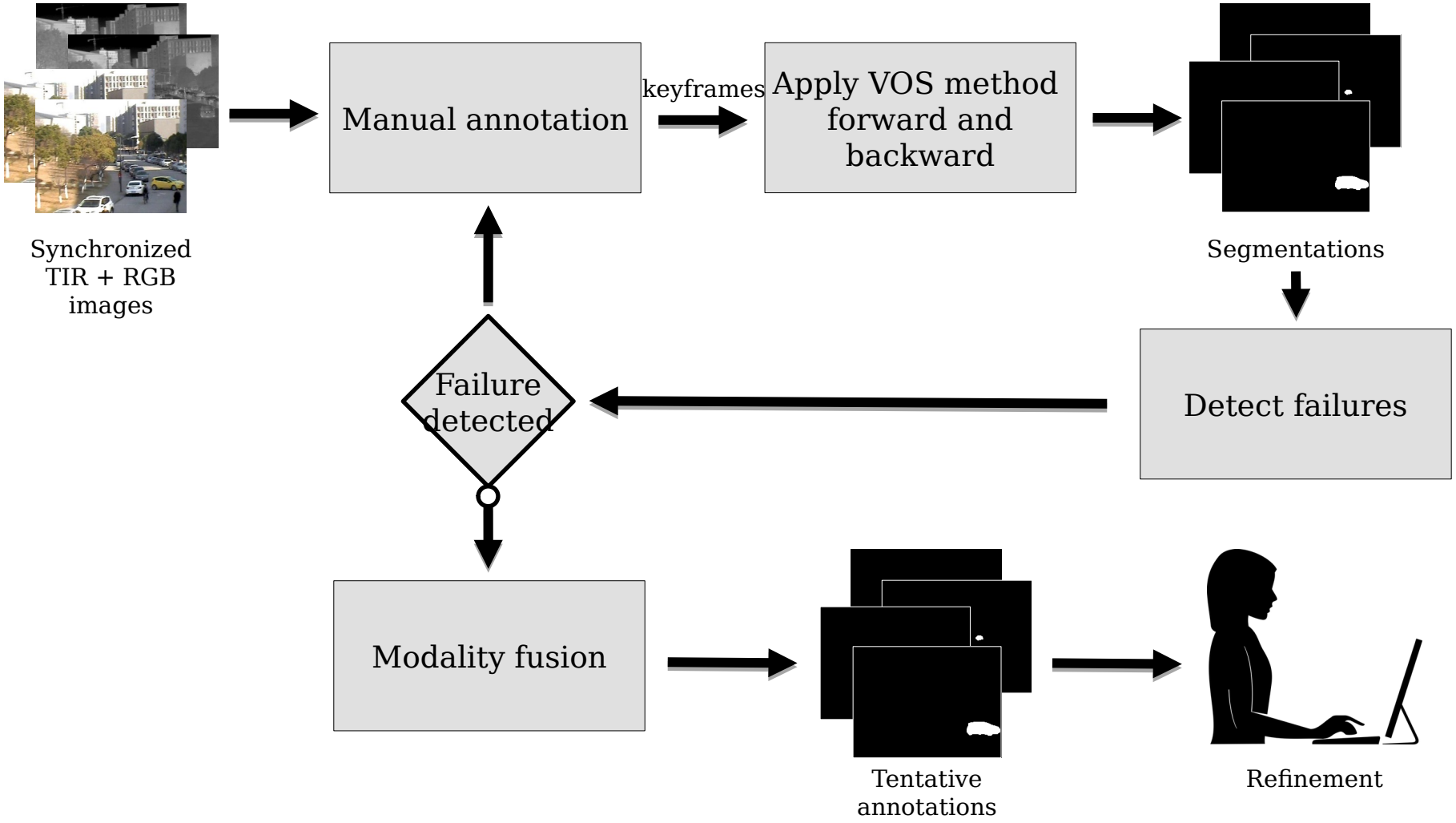
- Spatial accuracy (addressed by re-annotation)
- Synchronization (considered part of challenge)

Semi-automatic (re-)annotation

- Procedure described in: A. Berg, J. Johnander, F. D. de Gevigney, J. Ahlberg, and M. Felsberg. Semi-automatic annotation of objects in visual-thermal video. In VOT2019.
  - Step 1: semi-automatic video segmentation based on: J. Johnander, M. Danelljan, E. Brissman, F. S. Khan, and M. Felsberg. 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, Czech Technical University, Prague, January 2017.

Detect failures based on forward-backward consistency




Synchronized TIR + RGB images
Manual annotation
keyframes
Apply VOS method forward and backward
Segmentations
Detect failures
Failure detected
Modality fusion
Tentative annotations
Refinement

Results

- Enabling technique for realizing VOT-RGBT 2019.
- We estimated a 78% reduction in workload compared to full manual annotation of the VOT-RGBT 2019 dataset.
- Synchronization issue: TIR is used as reference
- Spatial accuracy: EAO RGB-TIR 0.75

Challenge Winner Protocol

- Evaluation is performed similar to VOT-ST 2020
  - Initialization points (anchors) are used
  - rotated bounding boxes had to be used (due to issues with sync and calibration)
- 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

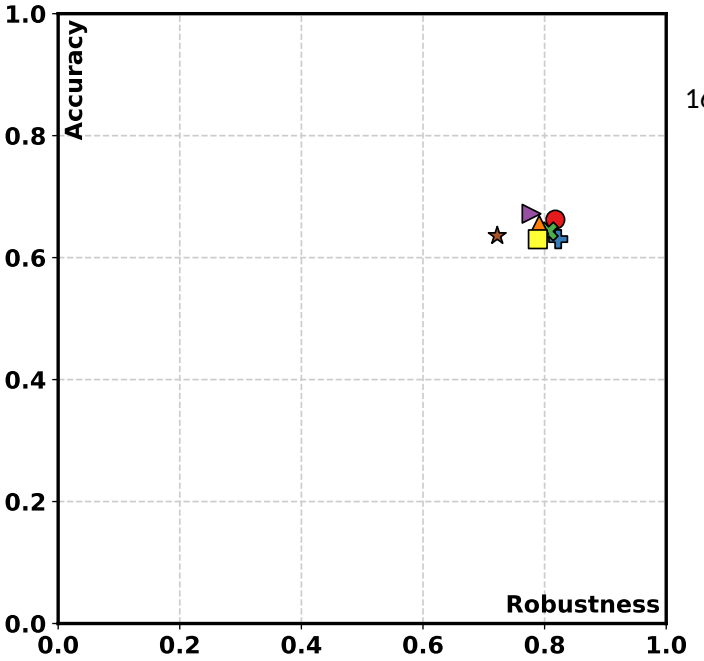
- 7 trackers in total, 5 submissions with code, 2 by committee (mfDiMP and SiamDW-T)
  - 2 ST_1, 3 ST_0
  - All 5 uniform dynamic model
  - All 5 based on discriminative model and holistic representation

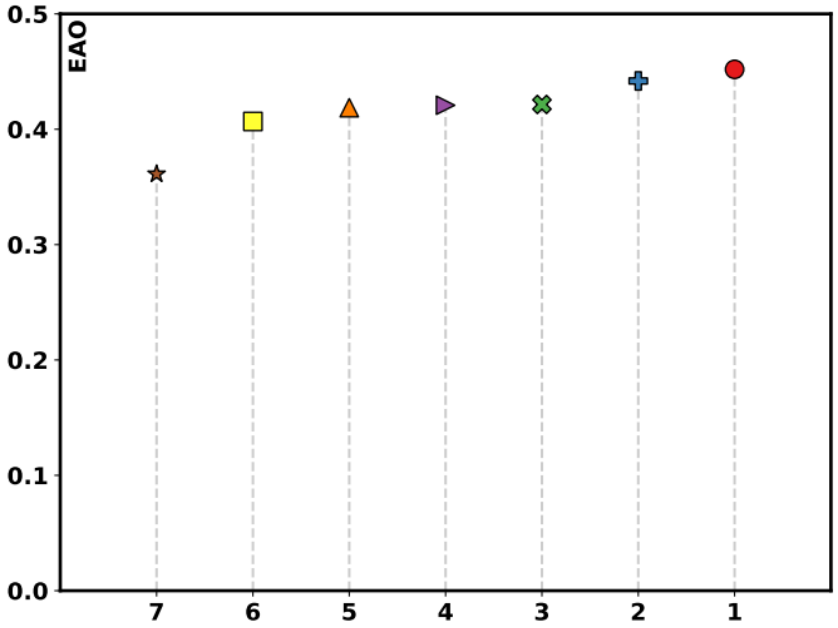
- 4 single-stage based on DCFs: M2C2Frgbt, JMMAC, AMF, SNDCFT
- 1 multi-stage based on Siamese network: DFAT
- 1 makes use of subspace methods and hand-crafted features: M2C2Frgbt
- 4 make use of deep features, 2 of them train the backbone: AMF and DFAT
- 1 makes use of ransac: JMMAC

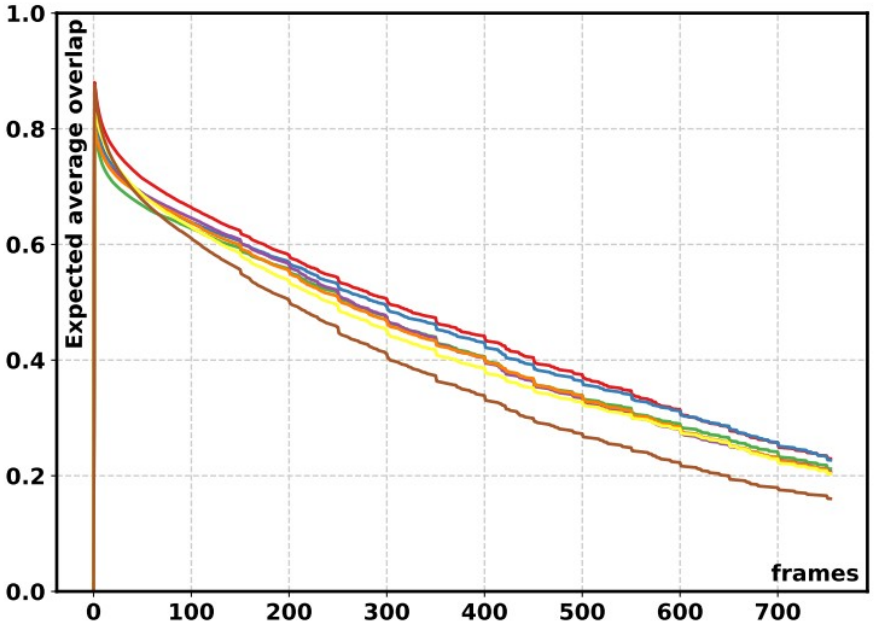
Results on public dataset

- All top-5 trackers use CNN features
- 3 out of these 5 trackers use DCFs
- 2 use Siamese networks
- #2 and #3 do backbone training

Tracker	EAO		A		R	
JMMAC	0.420	①	0.662	②	0.818	②
AMF	0.412	②	0.630		0.822	①
DFAT	0.390	③	0.672	①	0.779	
SiamDW-T	0.389		0.654	③	0.791	
mfDiMP	0.380		0.638		0.793	③
SNDCFT	0.378		0.630		0.789	
M2C2Frgbt	0.332		0.636		0.722	



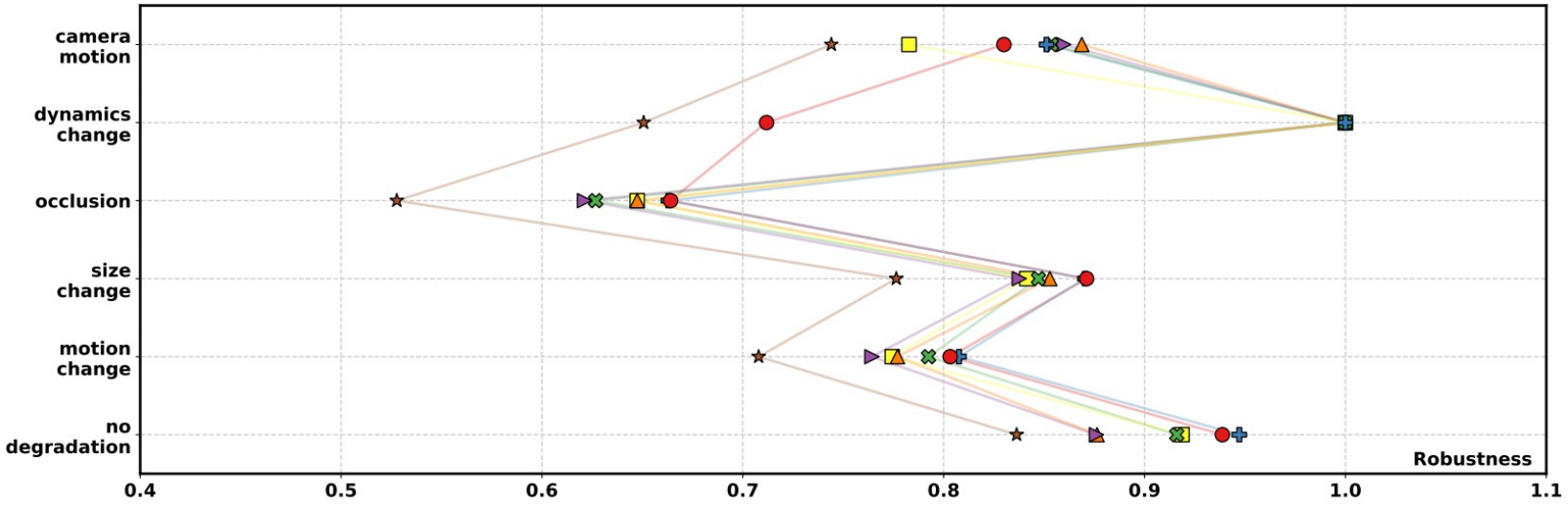




Example: AMF on man9

Further results

- EAO is stronger correlated to robustness than accuracy
- Robustness is most challenging for occlusion
- Changed order for sequestered dataset



Tracker	EAO	A	R
1. SiamDW-T	0.403 ①	0.664 ①	0.702 ③
2. mfDiMP	0.402 ②	0.623 ③	0.734 ①
3. DFAT	0.385 ③	0.654 ②	0.674
4. AMF	0.373	0.590	0.705 ②
5. JMMAC	0.158	0.576	0.287

VOT-RGBT2020 Winners

Winners of the VOT-RGBT 2020 challenge:

… by: H. Li, Z. Tang, T. Xu, X. Zhu, X. Wu, J. Kittler

“ Decision Fusion Adaptive Tracker (DFAT)”

(The talk in the next live session!)

Summary

- CNN-features dominating
- The ranking changes on sequestered dataset
- Overall performance decreases slightly on sequestered dataset
- Robustness most important
- Occlusion largest challenge
- For the future:
  - Attract more participants
  - Mittigate the effect of spatial missalignment and synchronization errors so that we can switch to segmentation-based evaluation

Thanks

- The VOT2020 committee


M. Kristan


J. Matas


A. Leonardis


M. Felsberg


R. Pflugfelder


J. K. Kamarainen

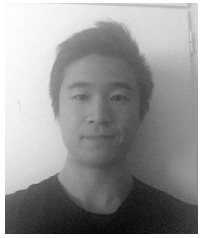

M. Danelljan


G. Fernandez


L. Čehovin


A. Lukežič

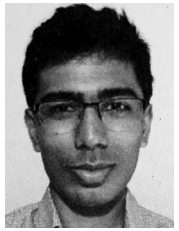

O. Drbohlav


H. Linbo


Y. Song


Y. Jinyu


Z. Yushan


G. Bhat

- Everyone who participated or contributed
- VOT2020 sponsor:

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