VOT
visual object tracking

The Visual Object Tracking Challenge Results
VOT-RGBD 2019

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

University of Ljubljana
Faculty of Computer and Information Science

UNIVERSITY OF BIRMINGHAM

LINKÖPING UNIVERSITY

AUSTRIAN INSTITUTE OF TECHNOLOGY

Tampere University

Why Consider RGB-D?

- RGB tracking well developed
- Recently depth sensors widely used (robotics, multimedia applications)
- Two modalities: RGB + Depth
- Potentially simplifies:
  - Occlusion detection
  - Object-to-background separation
  - Object 3D structure
- New research oportunities
  - How to combine RGB and D?
  - Depth changes or not reliable?

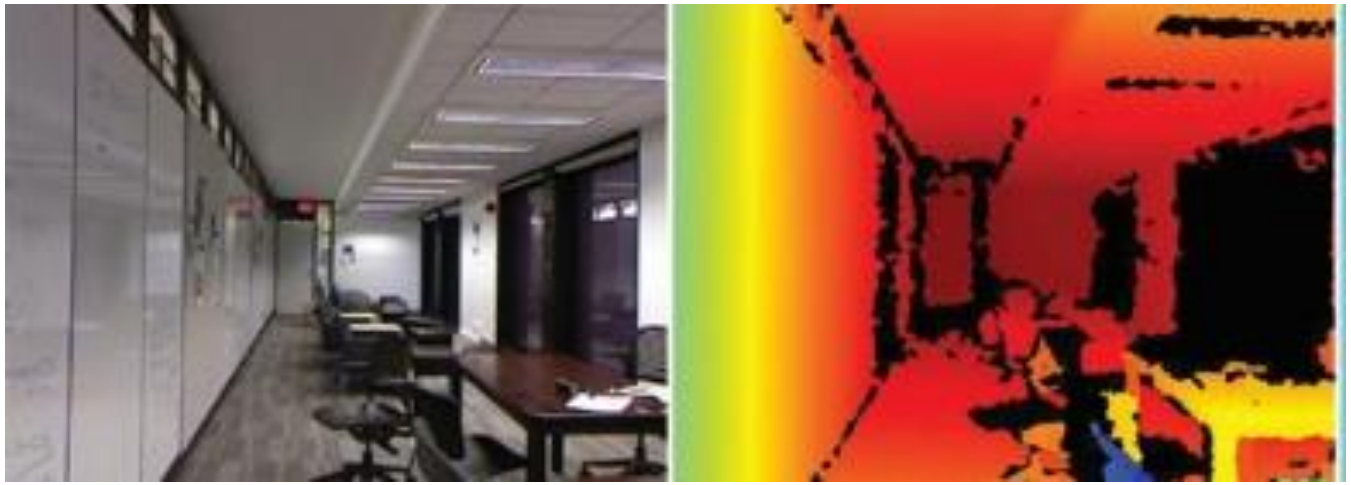
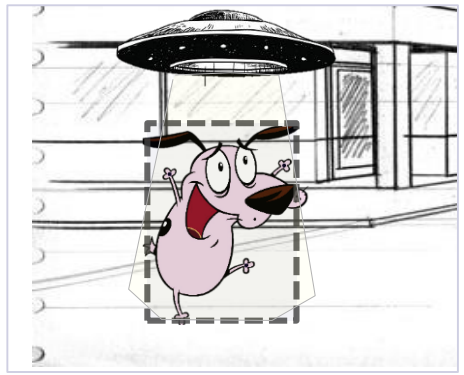
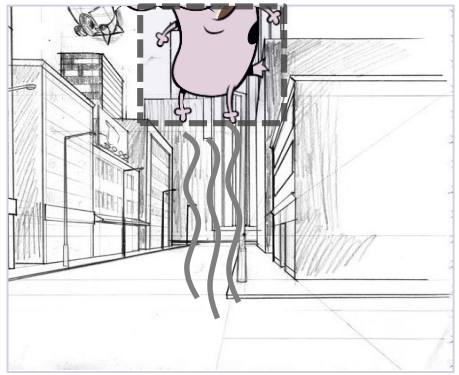
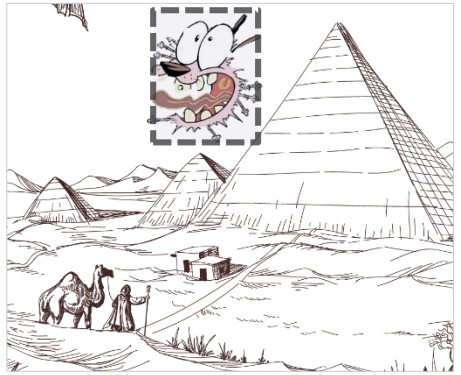


Figure from Silberman et al. ECCV2012

VOT RGB-D Challenge Scope

- Long-term tracking
- Required tracker properties
  - Determine when the target has been lost (or disappeared)
  - Re-detection after loosing the target

- Tracker prediction
  - A single training example in the first frame, target localization in the remaining frames
  - Output required: bounding box and presence confidence

RDB-D Datasets: Related Work

- Existing datasets:
  - PTB [1]: 100 sequences, average sequence length: 214 frames, short disappearances, synchronization problems, indoor only
  - STC [2]: 36 sequences, average sequence length: 255 frames, no target disappearances, limited outdoor scenarios
- Problems:
  - Not proper long-term evaluation (short sequences, short-term occlusions)
  - Lack of out-of-plane rotations
  - Single sensor used to capture video sequences (generalization?)

[1] S. Song et al., Tracking Revisited Using RGBD Camera: Unified Benchmark and Baselines, ICCV 2013.
[2] J. Xiao et al., Robust Fusion of Color and Depth Data for RGBD Target Tracking Using Adaptive Range-Invariant Depth Models and Spatio-Temporal Consistency Constraints. IEEE TCyB 2018..

The VOT-RGBD 2019 Dataset

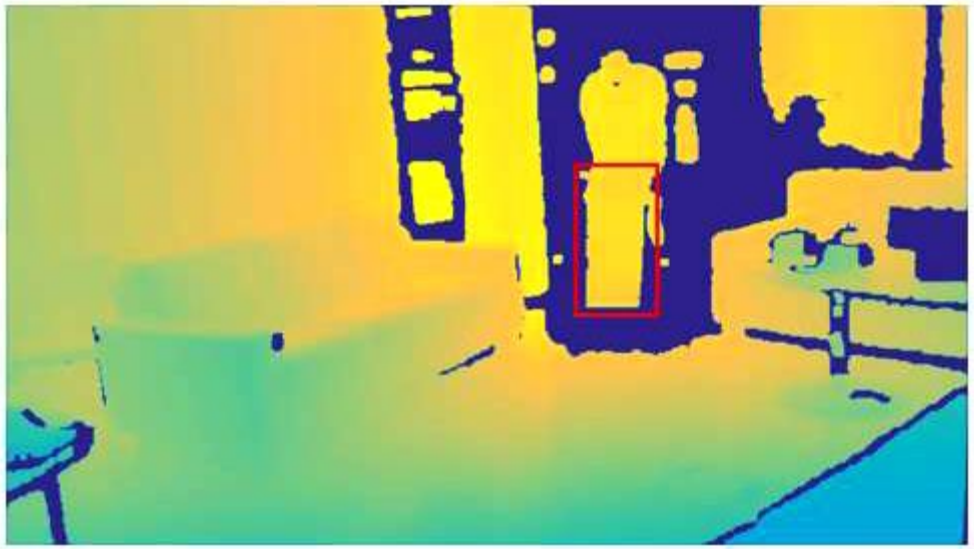
- Diverse sensors:

Kinect-v2:

ToF-RGB pair:

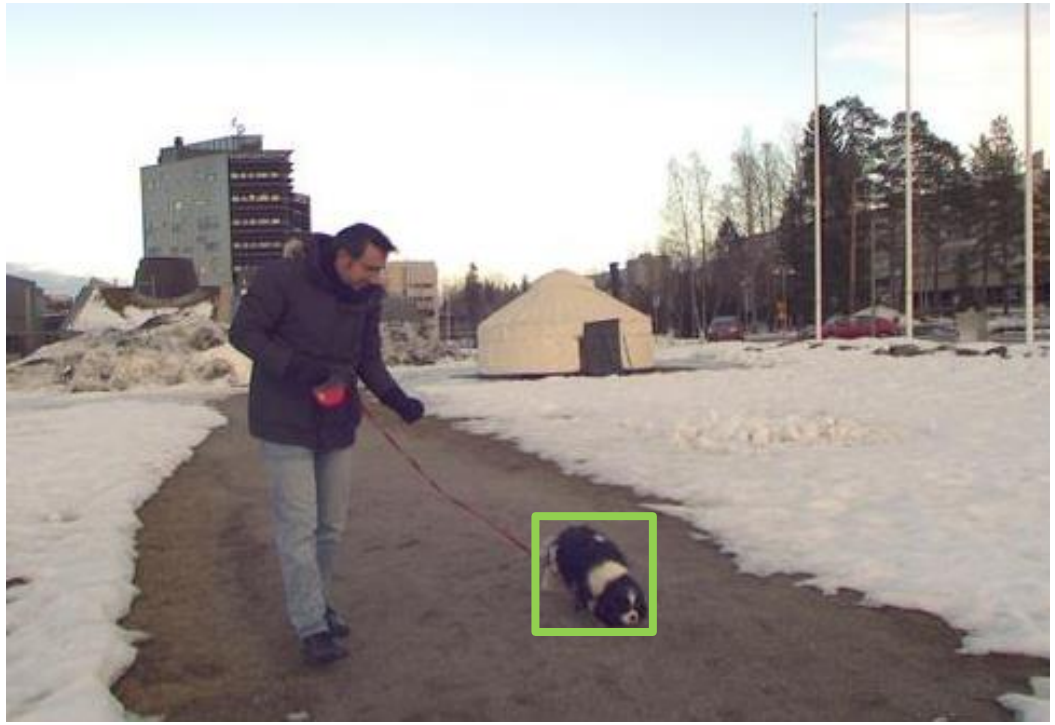
Stereo-camera pair:

- Indoor and outdoor sequences, spatially and temporally synchronized



The VOT-RGBD 2019 Dataset

- 80 sequences, average length 1274 frames
- Frequent and long-lasting target disappearances
  - Average target absence period: 56 frames
- Significant target pose change
- Axis-aligned bounding box



- Per-frame visual attributes:

  Full Occlusion, Target out-of-frame,
  Partial occlusion, Aspect change, Size change, Fast motion, Similar objects, Out-of-plane rotation,
  Reflective target, Depth change, Deformable target, Dark scene, Unassigned

VOT-RGBD 2019 Performance Evaluation

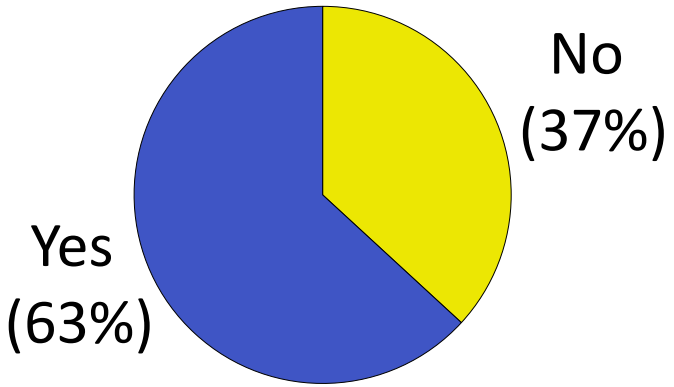
- No-reset experiment
  - Tracker initialized in the first frame, tracks to the end of the sequence
- Long-term tracking performance measures [1,2]
  - Tracking Precision (Pr): accuracy of predicted bboxes (when predictions given)
  - Tracking Recall (Re): accuracy of predicted bboxes (when target visible)
  - Tracking F-measure: F = (2 * Pr * Re) / (Pr + Re)

[1] A. Lukežič et al., Performance Evaluation Methodology for Long-TermVisual Object Tracking, arxiv: abs/1906.08675.
[2] M. Kristan et al., The sixth Visual Object Tracking VOT2018 challenge results, ECCVW 2018.

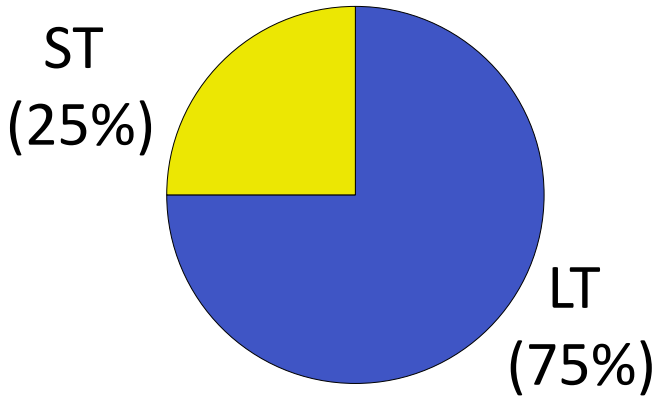
VOT-RBD 2019: Trackers Summary

- 12 trackers tested
- Diverse approaches

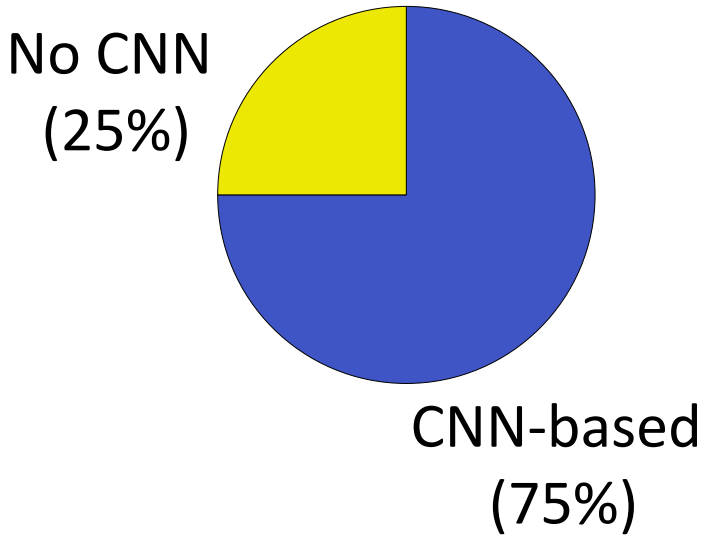
Is depth used?



ST/LT category:



Architecture

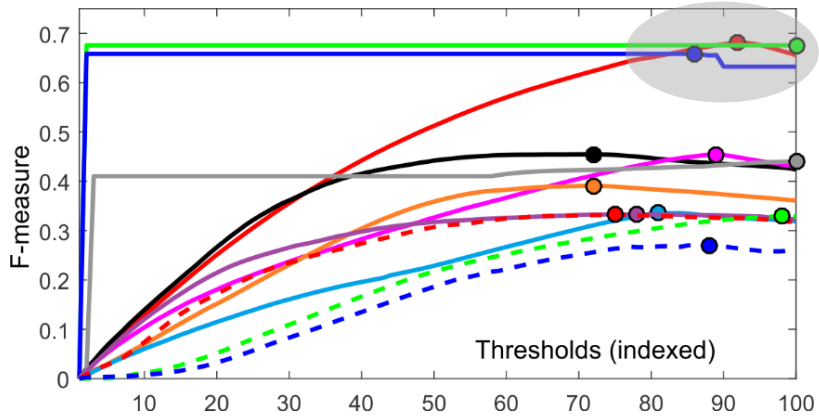


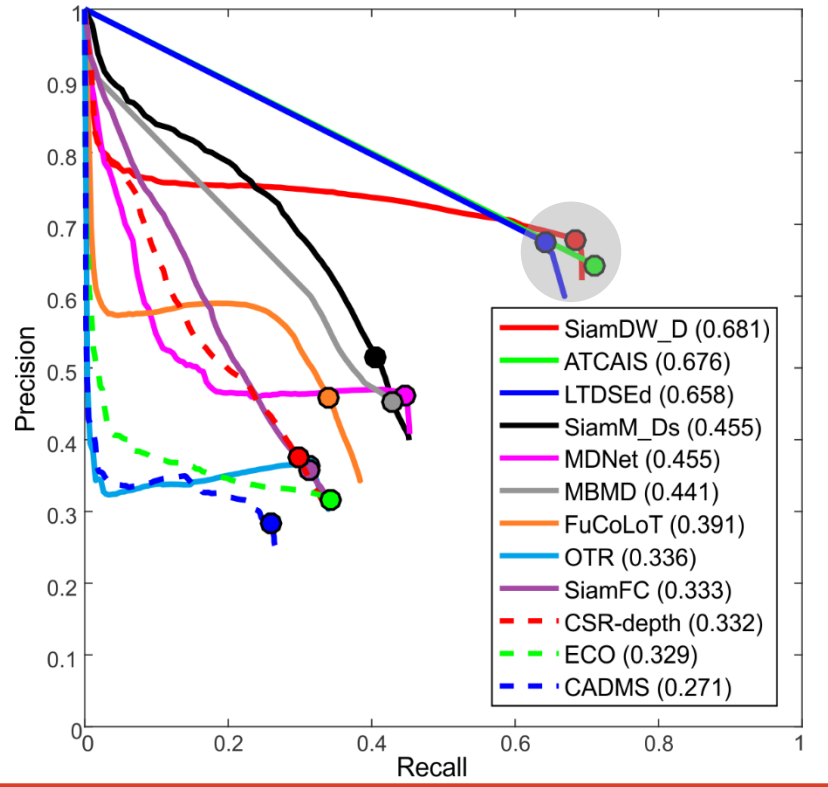
VOT-RGBD 2019 Results

TOP PERFORMERS

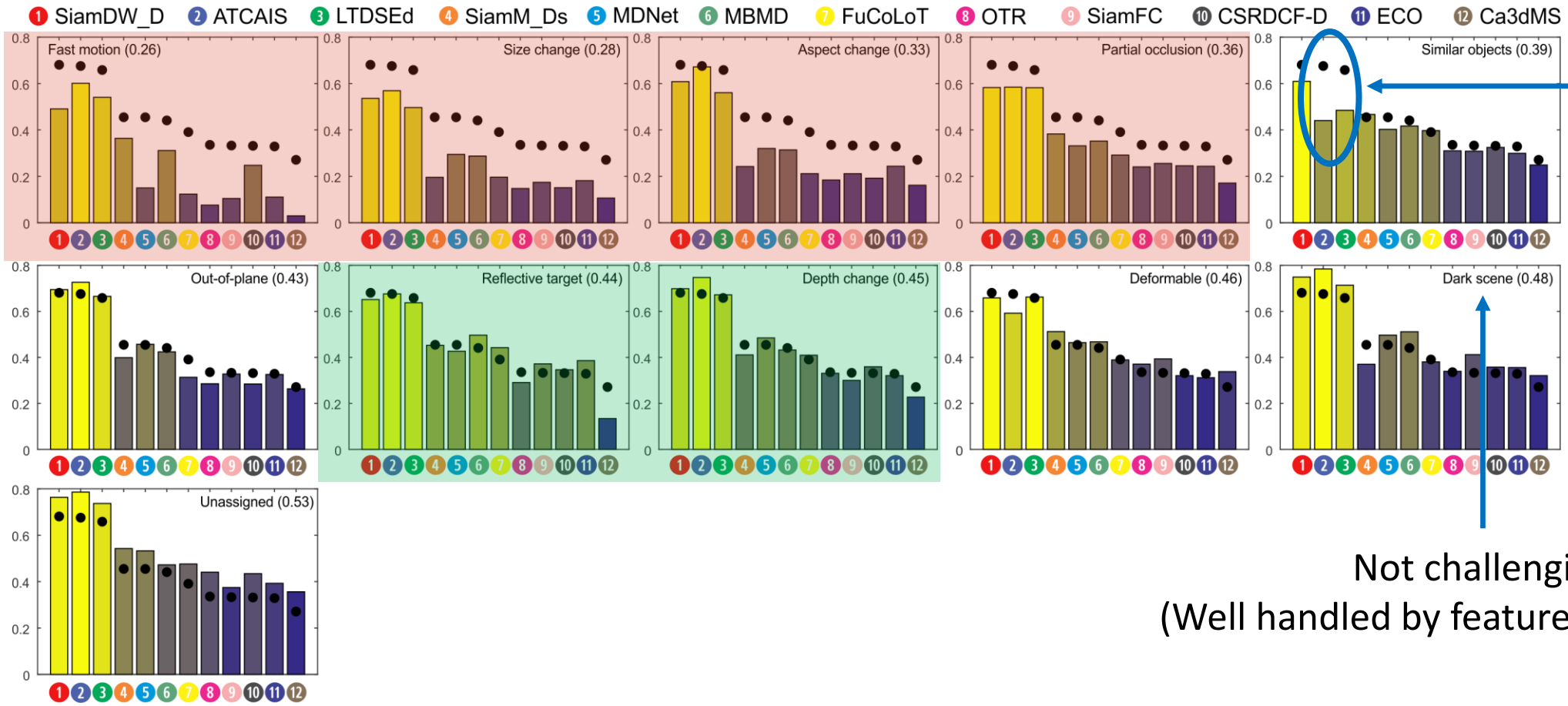
BASELINES (run by committee)

Tracker	Short-Term	Re-detection	Depth	Class
SiamDW_D	SiamRPN++, ATOM (ResNet50 + SENet154)	RPN (MMDet), IoU	Disappearance detection	LT_1
ATCAIS	ATOM, HTC instance segmentation	RPN (MMDet)	Disappearance detection, re-detection	LT_1
LTDSEd	ATOM, SiamMask, RT-MDNet verifier	RPN, RT-MDNet verifier	Disappearance detection	LT_1
SiamM_Ds	SiamMask	SiamMask, enlarged search window	Disappearance detection	LT_1
MDNet	Deep CNN classifier, random sampling	-	-	ST_1
MBMD	Siam bbox regression, MDNet verifier	Sampling, MDNet-based verifier (image-wide)	-	LT_1
FuCoLoT	DCF, color segmentation	DCF (image-wide)	-	LT_1
OTR	DCF, color segmentation	DCF (image-wide)	3D model	LT_1
SiamFC	Deep template, correlation	-	-	ST_1
CSRDCF-D	DCF, color segmentation	DCF (image-wide)	Segmentation	LT_1
ECO	DCF (deep features)	-	-	ST_1
Ca3dMS	3D mean-shift	Occluder tracking, verified by depth	3D tracking, disappearance detection	LT_0





VOT-RGBD 2019 Per-Attribute Analysis (F-measure)

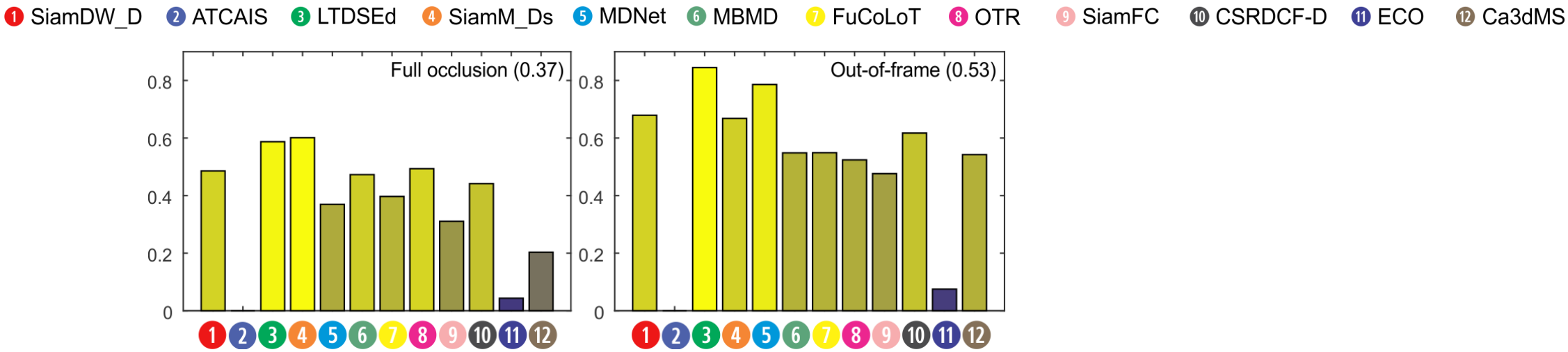


Not discriminative enough (Segmentation)

Not challenging (Well handled by features: deep or HC)

- The most challenging: fast motion, size change, aspect change, partial occlusion
- Well handled: depth change, reflective target

VOT-RGBD 2019 Per-Attribute Analysis (TNR)



- F-measure where target is not visible not defined (Recall = 0)
- True Negative Rate (TNR)
  Percentage of correctly predicted *target not visible*
- Out-of-frame easier to predict than full occlusion
  - Close to the image edge, similar objects at full occlusion

VOT-RGBD 2019 Challenge Summary

Top performers apply:

- Deep features to cope with appearance change (e.g., ResNet [1])
- Robust discriminative localizer (eg., ATOM [2])
- Localization refinement: IoUNet [2], Segmentation [3], RPN [4]
- Powerful detector (for target re-detection)

Research opportunities:

- Depth is mostly used only for target loss detection
- Depth for training: localization, detection or segmentation
- Depth for 3D modeling and foreground-background separation

[1] K. He et al., Deep Residual Learning for Image Recognition, CVPR 2016.
[2] M. Danelljan et al., ATOM: Accurate tracking by overlap maximization, CVPR 2019.
[3] Q. Wang, et al., Fast online object tracking and segmentation: A unifying approach, CVPR 2019.
[4] B. Li et al., High Performance Visual Tracking With Siamese Region Proposal Network, CVPR 2018.

VOT-RGBD 2019 Online Resources

- Paper, presentation, results, trackers code available online on: http://www.votchallenge.net/vot2019/
- Evaluation toolkit on github: https://github.com/votchallenge/vot-toolkit

- Dataset downloaded automatically by the toolkit

VOT-RGBD 2019 Awards

Winners of the VOT-RGBD 2019 challenge:

SiamDW-D by: H. Yu, H. Peng, Z. Wu, Y. Huang, J. Fu, L. Wang

“Online Deeper and Wider Siamese Networks for RGBD Visual Tracking”

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

Thanks

- The VOT2019 committee

M. Kristan J. Matas A. Leonardis M. Felsberg R. Pflugfelder J.K. Kamarainen G. Fernandez L. Čehovin A. Lukežič A. Eldesokey

- Everyone who participated or contributed

Matej Kristan1, Ji˘r´ı Matas2, Ale˘s Leonardis3, Michael Felsberg4, Roman Pflugfelder5,6, Joni-Kristian Kamarainen7, Luka C˘ ehovin Zajc1, Ondrej Drbohlav2, Alan Lukez˘ic˘1, 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´a˘s 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*