VOT
visual object tracking

The 9th Visual Object Tracking Challenge Results VOT-RGBD2021

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

University of Ljubljana
Faculty of Computer and Information Science

UNIVERSITY OF BIRMINGHAM

LINKÖPING UNIVERSITY

AUSTRIAN INSTITUTE OF TECHNOLOGY

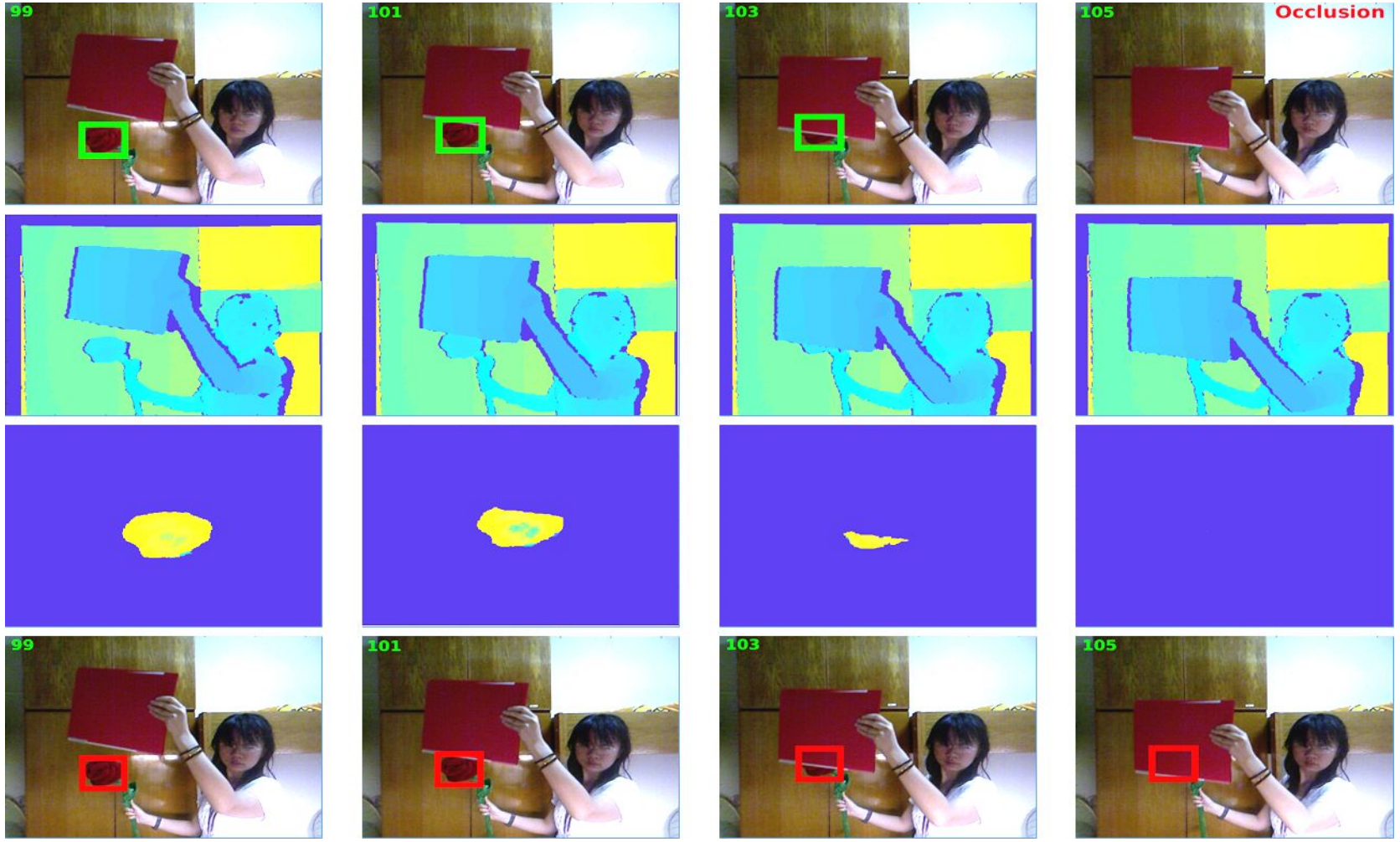
Tampere University

ETH zürich

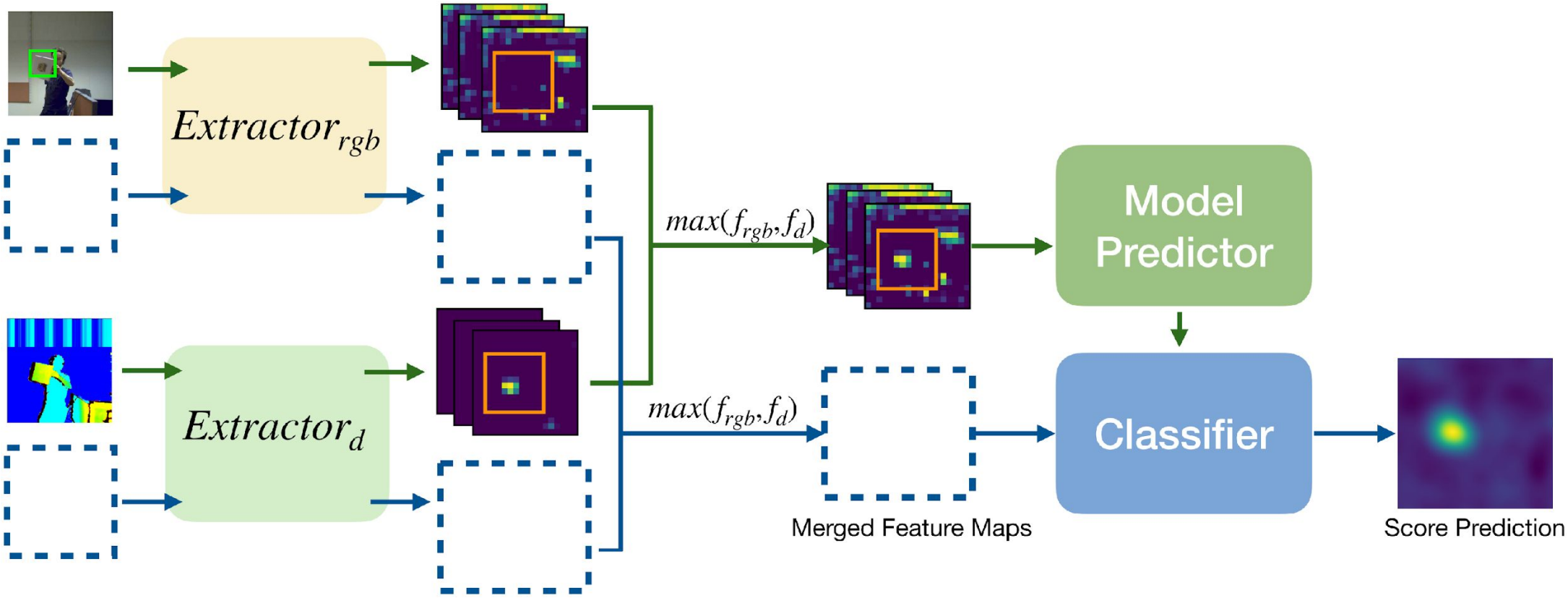
The VOT 2021 workshop

VOT2021 RGBD CHALLENGE: LET'S TALK ABOUT DEPTH

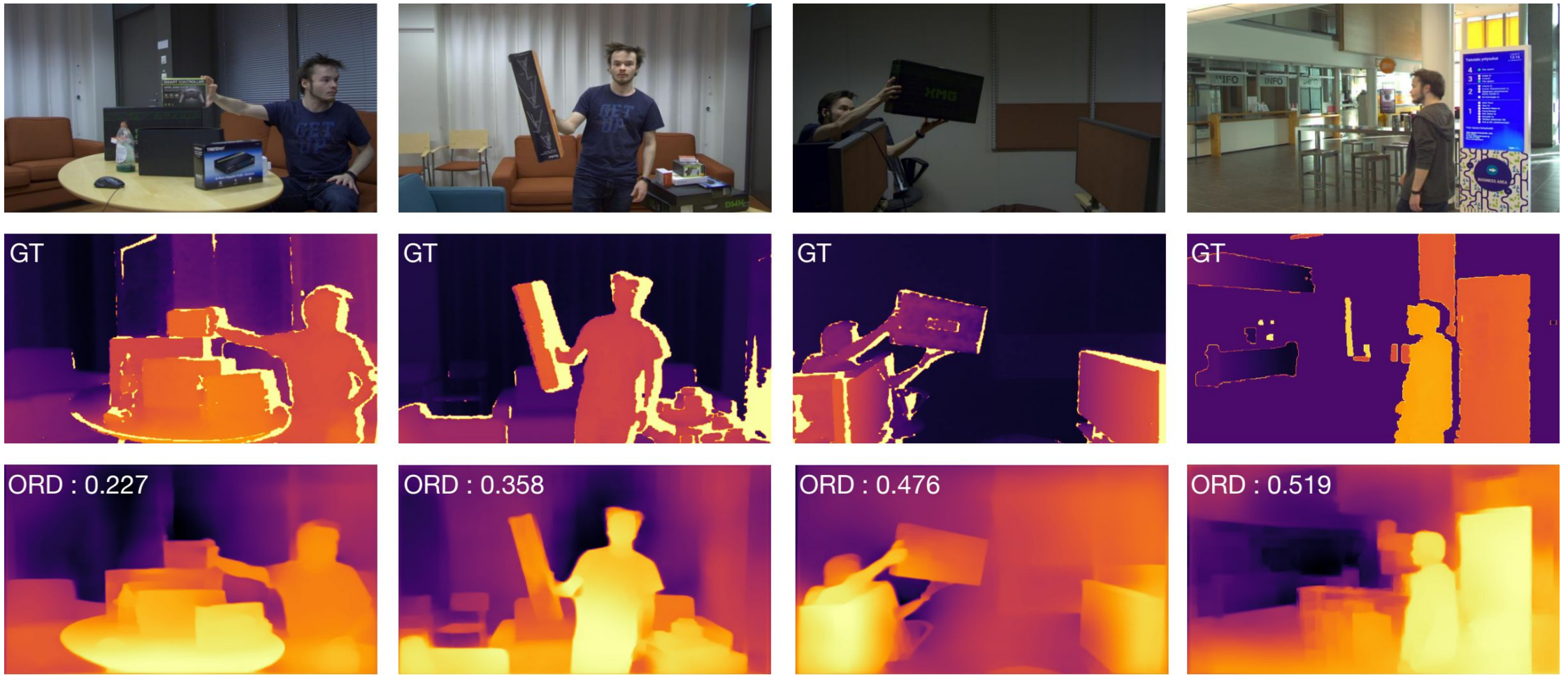
When do we need depth for tracking?



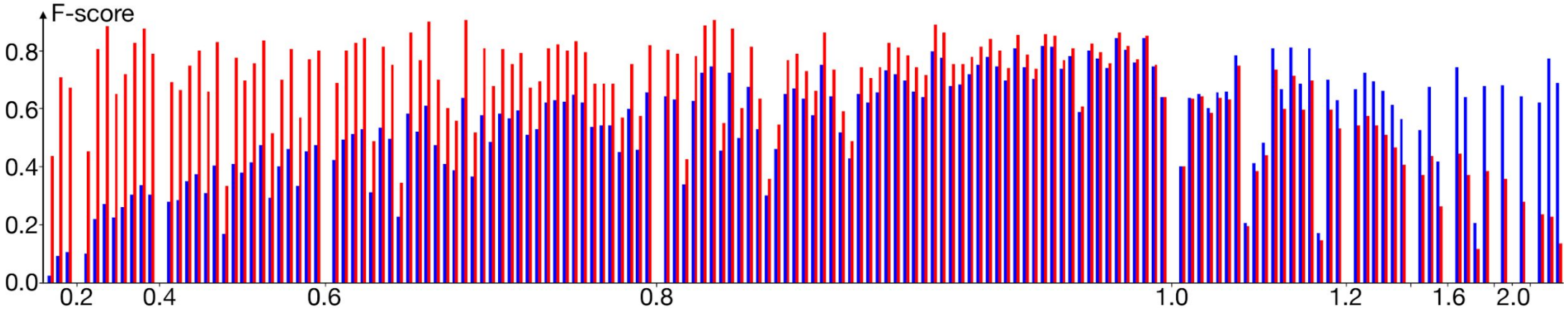
RGB-DiMP vs. D-DiMP vs. RGBD-DiMP



Generated depth for D and RGBD



Let data speak (short-term setting)



We trained two DiMPs using either **RGB** or **Depth** input and sorted all CDTB sequences based on their RGB-DiMP/Depth-DiMP ratios (CDTB was cut to short-term sequences)

RGB is better (ratio < 1.0) in 80% of the sequences, but for a substantial part of them D is almost equally good (and our Depth-DiMP is not even trained with real depth tracking data)

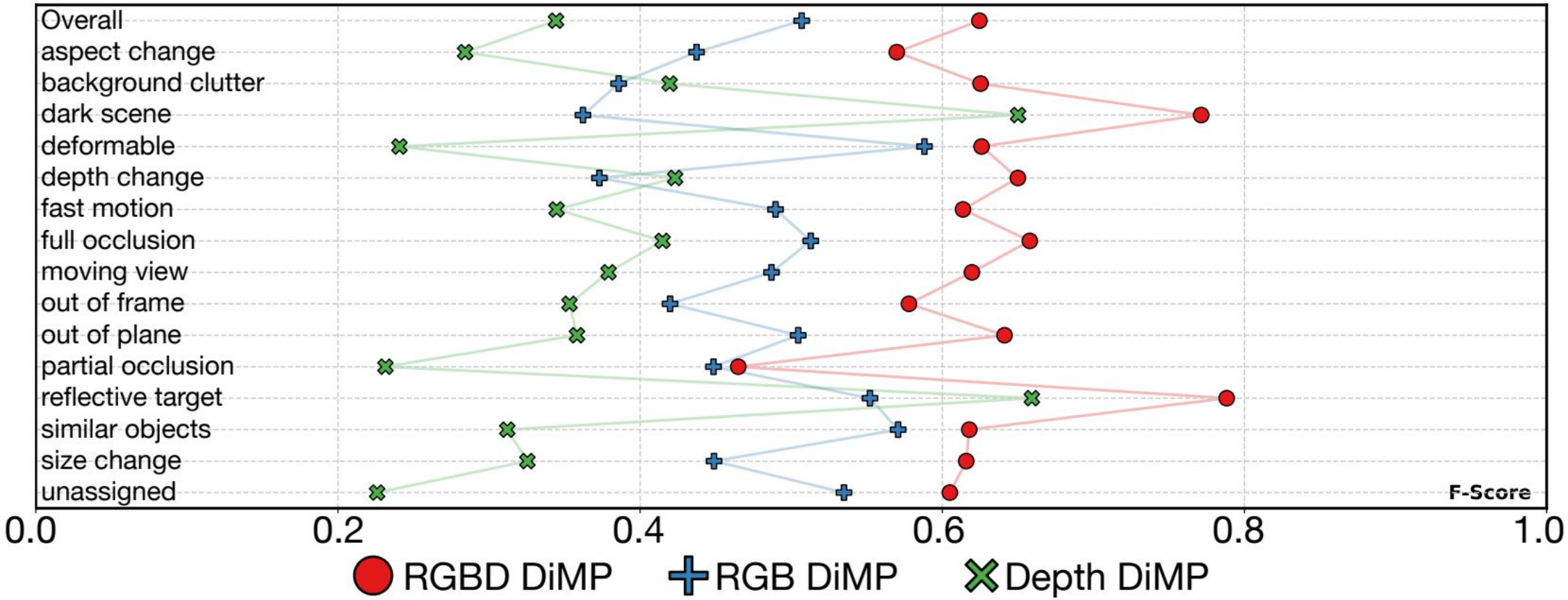
Depth is better (ratio > 1.0) in 20% of the sequences

It is still unclear (besides dark scenes)

		D DiMP				RGB DiMP		
id	*Seq. name*	Pr	Re	F	Ratio	Pr	Re	F
01	box_room_noocc_3_1	0.689	0.689	0.689	← 5.1	0.135	0.135	0.135
02	XMG_outside_2	0.771	0.771	0.771	← 3.4	0.229	0.229	0.229
03	boxes_office_occ_1_2	0.621	0.621	0.621	← 2.6	0.235	0.235	0.235
04	box_darkroom_noocc_9_1	0.643	0.643	0.643	← 2.3	0.278	0.278	0.278
05	boxes_room_occ_1_1	0.669	0.693	0.681	← 1.9	0.354	0.360	0.357
06	box_darkroom_noocc_5_1	0.677	0.679	0.678	← 1.8	0.385	0.387	0.386
07	box_room_occ_1_1	0.595	0.688	0.638	← 1.7	0.346	0.403	0.372
08	bag_outside_3	0.741	0.741	0.741	← 1.7	0.444	0.444	0.444
09	box_room_noocc_2_1	0.417	0.417	0.417	← 1.6	0.262	0.262	0.262
10	toy_office_occ_1_2	0.642	0.708	0.673	← 1.5	0.541	0.367	0.437
11	trophy_room_occ_1_2	0.083	0.103	0.092	7.7→	0.653	0.770	0.706
12	two_mugs_5	0.108	0.106	0.107	6.3→	0.707	0.640	0.672
13	thermos_office_occ_1_2	0.091	0.112	0.100	4.5→	0.416	0.494	0.452
14	box_room_noocc_8_1	0.220	0.220	0.220	3.7→	0.804	0.804	0.804
15	humans_corridor_occ_2_B_4	0.246	0.298	0.270	3.3→	0.882	0.882	0.882
16	humans_corridor_occ_2_B_1	0.160	0.372	0.224	2.9→	0.538	0.825	0.651
17	jug_3	0.259	0.259	0.259	2.8→	0.718	0.718	0.718
18	bottle_box_2	0.232	0.434	0.303	2.7→	0.846	0.808	0.827
19	trashcan_room_occ_1_6	0.333	0.342	0.337	2.6→	0.882	0.868	0.875
20	bottle_room_occ_1_2	0.558	0.209	0.304	2.6→	0.787	0.787	0.787

Table 3: 10 best CDTB-ST sequences for D and RGB DiMP trackers respectively. Ratios are converted to indicate how many times better the other modality is as denoted by the arrow.

But RGB+D indeed is good



Back to 2019 – Lessons learned

- We should have used short-term evaluation instead of long-term
- We should have included depth-only track as well

The VOT 2021 workshop

VOT2021 RGBD CHALLENGE: DATASETS

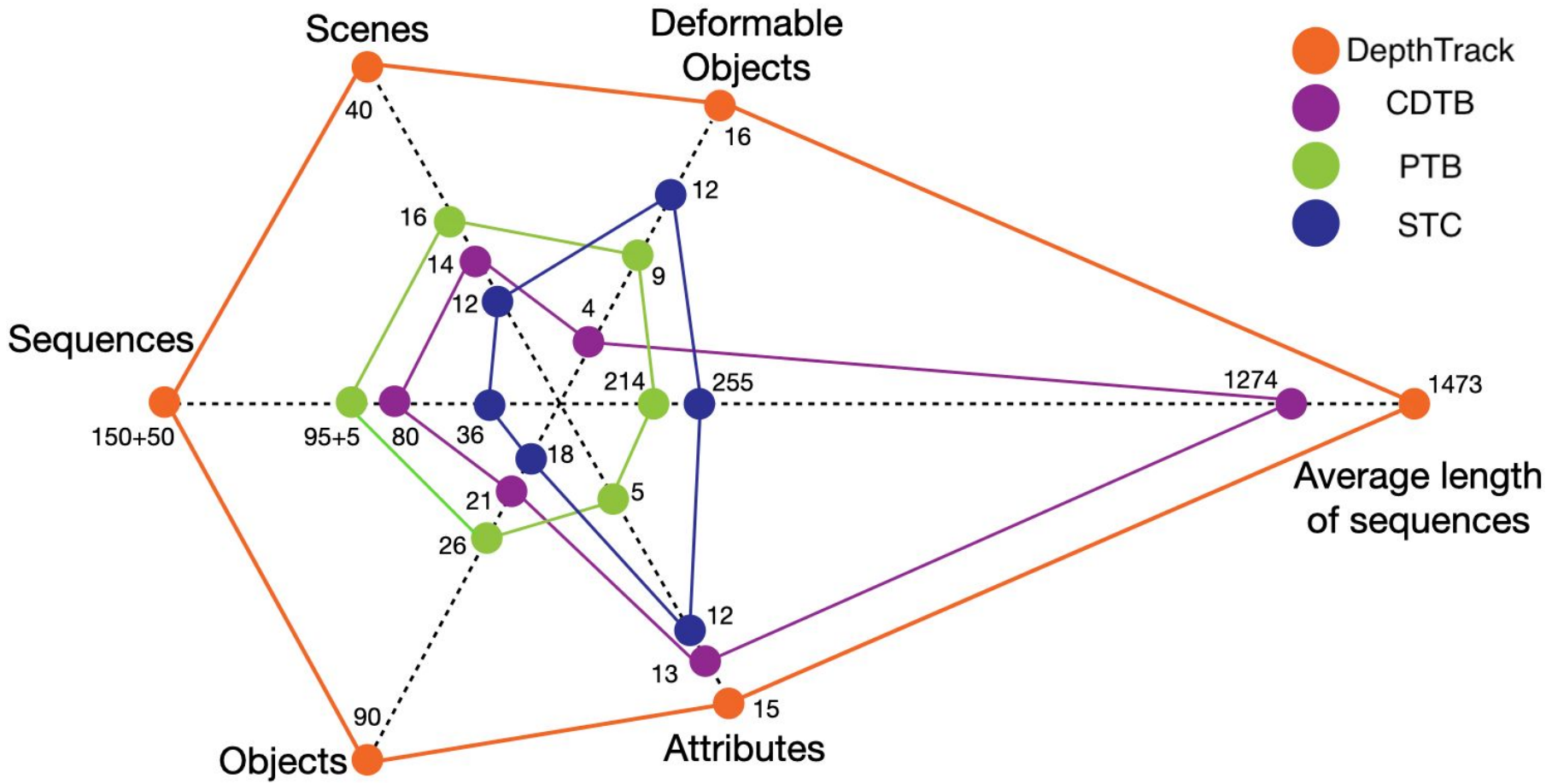
The VOT-RGBD 2021 Dataset (2019-2020)

- 80 sequences, average length 1274 frames
- Frequent and long-lasting target disappearances
  - Average target absence period: 56 frames
- Significant target 3D pose changes
- 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

[3] CDTB: A Color and Depth Visual Object Tracking Dataset and Benchmark (A. Lukezic, U. Kart, J. Käpylä, A. Durmush, J.-K. Kämäräinen, J. Matas and M. Kristan), In Int. Conf. on Computer Vision (ICCV), 2019.

The DepthTrack Dataset (2021 “sequestered”)



DepthTrack: Unveiling the Power of RGBD Tracking (S. Yan, J. Yang, J. Käpylä, F. Zheng, A. Leonardis, J.-K. Kämäräinen), In Int. Conf. on Computer Vision (ICCV), 2021.

The VOT 2021 workshop

VOT2021 RGBD CHALLENGE: RESULTS

VOT-RGBD 2021 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-Term Visual Object Tracking, arxiv: abs/1906.08675.
[2] M. Kristan et al., The sixth Visual Object Tracking VOT2018 challenge results, ECCVW 2018.

VOT-RGBD from 2019 to 2020

2019 – 4 valid entries (1. SiamDW-D, 2. ATCAIS, 3. LTDSEd, and 4. SiamM_Ds)

2020 – 4 valid entries (1. ATCAIS, 3. CLGS_D, 2. DDiMP, and 4. Siam_LTD)

2021 – 5 valid entries (sttc_rgbd, STARK_RGBD, SLMD, TALGD, DRefine)

VOT-RGBD 2021 Results

Tracker	Pr	Re	F-Score	Year
STARK_RGBD	0.742①	0.769①	0.756①	2021
TALGD	0.728②	0.717②	0.722②	2021
DRefine	0.707	0.715③	0.711③	2021
ATCAIS	0.709	0.696	0.702	2020
DDiMP	0.703	0.689	0.696	2020
CLGS_D	0.725③	0.664	0.693	2020
SLMD	0.701	0.685	0.693	2021
sttc_rgbd	0.692	0.685	0.689	2021

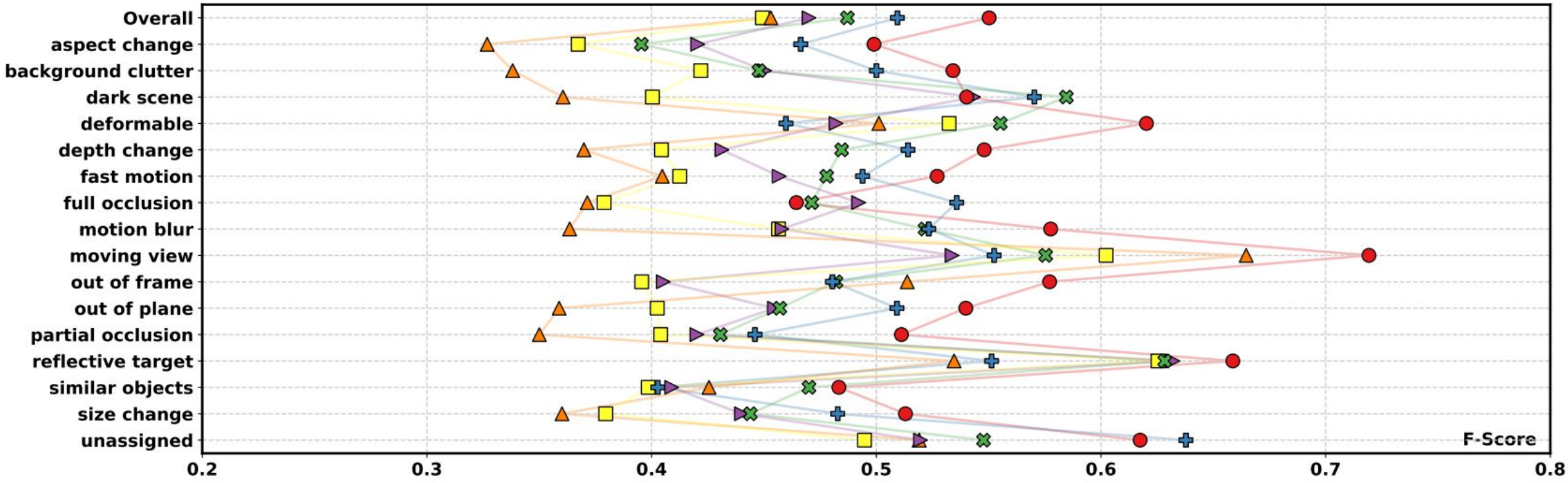
Table 4. Results of the five submitted RGBD trackers for the public VOT 2021 RGBD test data (CDTB). The numbers were computed using the user provided data. The three additional trackers are the three best from the last year (2020).

VOT-RGBD 2021 Results (sequestered)

Tracker	Pr	Re	F-Score	Year
STARK_RGBD	0.558②	0.543①	0.550①	2021
TALGD	0.540③	0.482②	0.509②	2021
DDiMP	0.505	0.470③	0.487③	2020
ATCAIS	0.491	0.451	0.470	2020
CLGS_D	0.585①	0.370	0.453	2020
DRefine	0.468	0.432	0.449	2021

Table 5. RGBD results of the best three submitted trackers for 2021 sequestered RGBD sequences. The other tree trackers are from the previous year.

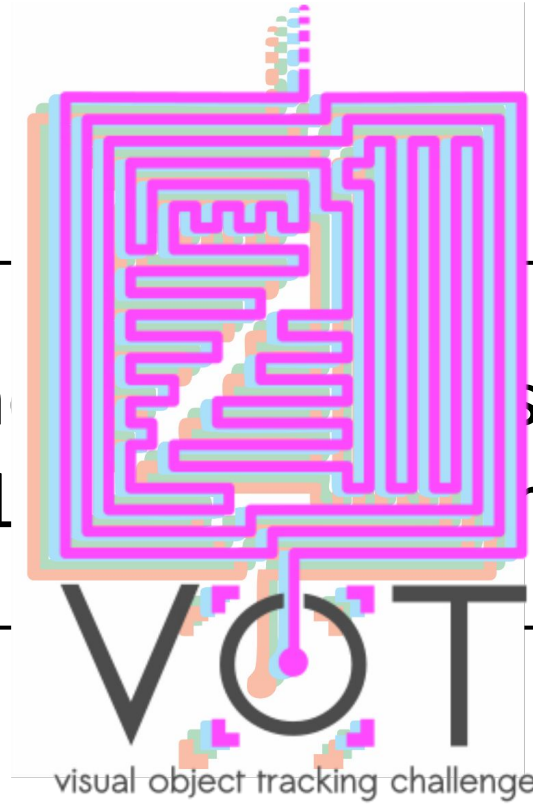
VOT-RGBD 2021 Attributes (sequestered)



The VOT 2021 workshop

VOT2021 RGBD WINNER ANNOUNCEMENTS

Winn sion III
(15:30-1 me zone)



VOT-RGBD2021 Winners:

STARK_RGBD by: X. Zhang, B. Yan, L. Wang, H. Peng, D. Wang, H. Lu and X. Yang