The Visual Object Tracking Challenge Results
VOT-RGBD 2020

Matej Kristan, Aleš Leonardis, Jiri Matas, Michael Felsberg, Roman Pflugfelder, Joni-Kristian Kämäräinen, Martin Danellljjan, Luka Čehovin Zajc, Alan Lukežič, Ondrej Drbohlav, Linbo He, Yushan Zhang, Song Yan, Jinyu Yang, et al.
Why to add depth?

- In robotics people ask »Why to add RGB?«

- **Depth simplifies:**
  - 3D sensing
  - Occlusion detection
  - Object-to-background separation

- **Challenges**
  - How to combine RGB and D?
  - OR when to switch between RGB and D?
  - Much less RGBD than RGB tracking datasets
RGBD Datasets

• 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:
  • Small number of sequences and/or they are short
  • Depth is overly stable (lack of out-of-plane rotations)
  • Single sensor used (overfitting)

The VOT-RGBD 2020 Dataset (2019)

- 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

The VOT-RGBD 2020 Dataset

- Diverse sensors:
  - Kinect-v2: [Image]
  - ToF-RGB pair: [Image]
  - Stereo-camera pair: [Image]

- Indoor and outdoor sequences, spatially aligned and temporally synchronized (ask for raw data)
VOT-RGBD setup

• **Long-term** tracking
  - Full/partial occlusions (occlusion handling needed)
  - Target may disappear (re-detection needed)
  - Evaluation uses the VOT-LT protocol

• No sequestered data (thanks to COVID-19)

• Required outputs: bounding box and presence confidence
VOT-RGBD 2020 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**: $F = (2 \times \text{Pr} \times \text{Re}) / (\text{Pr} + \text{Re})$

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 (ATCAIS, CLGS_D, DdiMP, and Siam_LTD)

We added baselines and the best trackers from all VOT categories
### VOT-RGBD 2020 Results

<table>
<thead>
<tr>
<th>Tracker</th>
<th>Pr</th>
<th>Re</th>
<th>F-Score</th>
<th>ST/LT</th>
<th>RGB/RGBD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ATCAIS</strong></td>
<td>0.709</td>
<td>0.696</td>
<td>0.702</td>
<td>LT</td>
<td>RGBD</td>
</tr>
<tr>
<td><strong>DDiMP</strong></td>
<td>0.703</td>
<td>0.689</td>
<td>0.696</td>
<td>ST</td>
<td>RGBD</td>
</tr>
<tr>
<td><strong>CLGS.D</strong></td>
<td>0.725</td>
<td>0.664</td>
<td>0.693</td>
<td>LT</td>
<td>RGBD</td>
</tr>
<tr>
<td><strong>SiamDW.D</strong></td>
<td>0.677</td>
<td>0.685</td>
<td>0.681</td>
<td>LT</td>
<td>RGBD</td>
</tr>
<tr>
<td><strong>LTDSyEd</strong></td>
<td>0.674</td>
<td>0.643</td>
<td>0.658</td>
<td>LT</td>
<td>RGBD</td>
</tr>
<tr>
<td><strong>RLT.DiMP</strong></td>
<td>0.625</td>
<td>0.632</td>
<td>0.629</td>
<td>LT</td>
<td>RGB</td>
</tr>
<tr>
<td><strong>LTMU.B</strong></td>
<td>0.680</td>
<td>0.581</td>
<td>0.626</td>
<td>LT</td>
<td>RGB</td>
</tr>
<tr>
<td><strong>Megtrack</strong></td>
<td>0.694</td>
<td>0.551</td>
<td>0.614</td>
<td>LT</td>
<td>RGB</td>
</tr>
<tr>
<td><strong>RPT</strong></td>
<td>0.601</td>
<td>0.546</td>
<td>0.572</td>
<td>ST</td>
<td>RGB</td>
</tr>
<tr>
<td><strong>Siam.LTD</strong></td>
<td>0.626</td>
<td>0.489</td>
<td>0.549</td>
<td>LT</td>
<td>RGBD</td>
</tr>
<tr>
<td><strong>OceanPlus</strong></td>
<td>0.577</td>
<td>0.502</td>
<td>0.537</td>
<td>ST</td>
<td>RGB</td>
</tr>
<tr>
<td><strong>AlphaRef</strong></td>
<td>0.491</td>
<td>0.547</td>
<td>0.518</td>
<td>ST</td>
<td>RGB</td>
</tr>
</tbody>
</table>

**Table 6.** List of trackers that participated in the VOT-RGBD2020 challenge along with their performance scores (Pr, Re, F-score) and categorizations (ST/LT, RGB/RGBD). 2020 submissions are ATCAIS, DDiMP, CLGS.D and Siam.LTD. SiamDW.D and LTDSyEd are 2019 submissions (SiamDW.D was the winner). RGB trackers are the three top performers of VOT-ST2020 and VOT-LT2020.
Figure 16. VOT-RGBD2019 challenge: tracking performance w.r.t. visual attributes. The first eleven attributes correspond to scenarios with a visible target (showing F-measure). The overall tracking performance is shown in each graph with black dots. The attributes full occlusion and out of view represent periods when the target is not visible and true negative rate is used to measure the performance.
Fig. 14. VOT-RGBD2020 challenge: tracking performance w.r.t. visual attributes.
VOT-RGBD 2020 General Findings

Number of sequences per scene type

- Lab: 11
- Corridor: 12
- Lobby: 2
- Office desk: 5
- Dark office: 10
- Outside: 27
- Robotics lab: 2
- Desk: 1

Number of failures per scene type

- Corridor: 295
- Office: 1138
- Office desk: 5
- Lobby: 473
- Dark office: 261
- Outside: 406
- Robotics lab: 2
- Desk: 6

Number of sequences per object type

- Mug: 3
- Toy: 2
- Dog: 2
- Human: 13
- Backpack: 3
- Bicycle: 2
- Box: 1
- Trash can: 3

Number of failures per object type

- Mug: 523
- Box: 1
- Bicycle: 1
- Cart: 1
- Bottle: 1
- Trash can: 1
- Mug: 1
- Box: 1
VOT-RGBD 2020 Challenge Summary

Ingredients of the best RGBD trackers:

• Take the best available long-term RGB tracker with all tricks and hacks
• Use depth-based hacks and tweaks for occlusion detection

Open questions:

• Can depth provide discriminative features for tracking?
• Can we learn optimal depth usage from data?
• Depth for 3D tracking?

Winner of the VOT-RGBD 2020 challenge:

ATCAIS by Y. Wang, L. Wang, D. Wang, H. Lu and X. Yang

<table>
<thead>
<tr>
<th>Tracker</th>
<th>Pr</th>
<th>Re</th>
<th>F-Score</th>
<th>ST/LT</th>
<th>RGB/RGBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATCAIS</td>
<td>0.709</td>
<td>0.696</td>
<td>0.702</td>
<td>LT</td>
<td>RGBD</td>
</tr>
<tr>
<td>DDiMP</td>
<td>0.703</td>
<td>0.689</td>
<td>0.696</td>
<td>ST</td>
<td>RGBD</td>
</tr>
<tr>
<td>CLGS_D</td>
<td>0.725</td>
<td>0.664</td>
<td>0.693</td>
<td>LT</td>
<td>RGBD</td>
</tr>
<tr>
<td>SiamDW_D</td>
<td>0.677</td>
<td>0.685</td>
<td>0.681</td>
<td>LT</td>
<td>RGBD</td>
</tr>
<tr>
<td>LTDSEd</td>
<td>0.674</td>
<td>0.643</td>
<td>0.658</td>
<td>LT</td>
<td>RGBD</td>
</tr>
<tr>
<td>RLT_DiMP</td>
<td>0.625</td>
<td>0.632</td>
<td>0.629</td>
<td>LT</td>
<td>RGB</td>
</tr>
<tr>
<td>LTMU_B</td>
<td>0.680</td>
<td>0.581</td>
<td>0.626</td>
<td>LT</td>
<td>RGB</td>
</tr>
<tr>
<td>Megtrack</td>
<td>0.694</td>
<td>0.551</td>
<td>0.614</td>
<td>LT</td>
<td>RGB</td>
</tr>
<tr>
<td>RPT</td>
<td>0.601</td>
<td>0.546</td>
<td>0.572</td>
<td>ST</td>
<td>RGB</td>
</tr>
<tr>
<td>Siam.LTD</td>
<td>0.626</td>
<td>0.489</td>
<td>0.549</td>
<td>LT</td>
<td>RGBD</td>
</tr>
<tr>
<td>OceanPlus</td>
<td>0.577</td>
<td>0.502</td>
<td>0.537</td>
<td>ST</td>
<td>RGB</td>
</tr>
<tr>
<td>AlphaRef</td>
<td>0.491</td>
<td>0.547</td>
<td>0.518</td>
<td>ST</td>
<td>RGB</td>
</tr>
</tbody>
</table>

Table 6. List of trackers that participated in the VOT-RGBD2020 challenge along with their performance scores (Pr, Re, F-score) and categorizations (ST/LT, RGB/RGBD). 2020 submissions are ATCAIS, DDiMP, CLGS_D and Siam.LTD. SiamDW_D and LTDSEd are 2019 submissions (SiamDW_D was the winner). RGB trackers are the three top performers of VOT-ST2020 and VOT-LT2020.