Trans2k: Unlocking the Power of Deep Models for Transparent Object Tracking

Alan Lukežič[1], Žiga Trojer[1], Jiri Matas[2], Matej Kristan[1]

[1] Faculty of Computer and Information Science, University of Ljubljana, Slovenia, [2] Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic

Tracking Transparent Objects

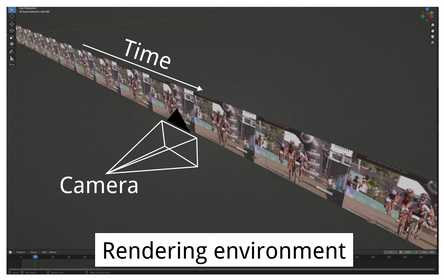
- Most of the visual object tracking research focused on opaque objects (typically well distinguishable from the background)
- Tracking transparent objects: **background visible through the object**
- Relevant for many applications

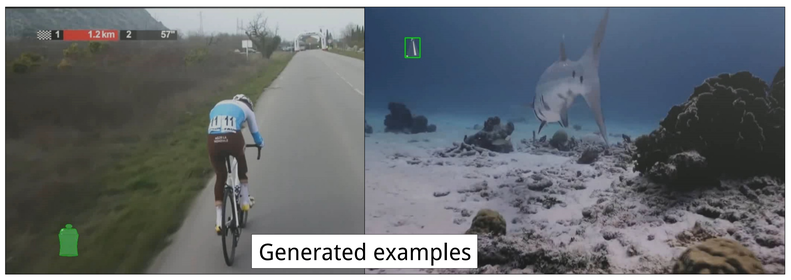
Current research:

- Transparent object tracking benchmark (TOTB) [1]
  Evaluation dataset: 225 fully annotated videos, 15 object categories
- SOTA: deep trackers
- **Drawback**: training dataset does not exist

Proposed Solution

- Construct **the first training dataset** for transparent object tracking
- Challenges: (i) Time-consuming data **acquisition**, (ii) Expensive **annotation**
- Observations:
  (i) Transparent objects can be **rendered** realistically
  (using modern renderers, e.g. BlenderProc [7])
  (ii) Human-level **realism not required** in training videos for deep learning [2,3,4,5]
- Solution: **Render the training dataset** for transparent object tracking

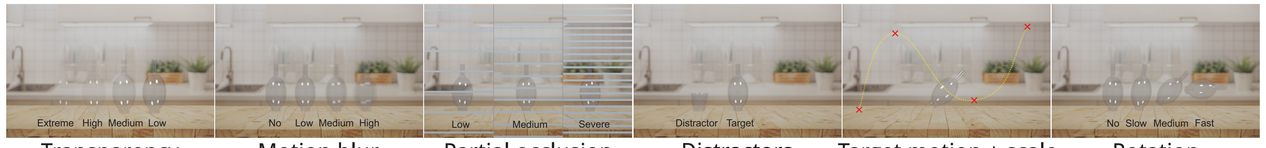




[1] H. Fan, et al., Transparent Object Tracking Benchmark, ICCV 2021
[2] P. Krahenbuhl, Free supervision from video games, CVPR 2018
[3] S. Richter et al., Playing for benchmarks, ICCV 2017
[4] G. Ros et al., The synthia dataset, CVPR 2016
[5] Hodan T. et al., BOP: Benchmark for 6D object pose estimation, ECCV 2018
[6] M. Denninger, et al., Reducing the reality gap with photorealistic rendering, ICRSS, 2020
[7] L. Huang, et al., GOT-10k: A large high-diversity benchmark for generic object tracking in the wild, IEEE TPAMI, 2019
[8] H. Fan, et al., Lasot: A high-quality benchmark for large-scale single object tracking, CVPR, 2019
[9] M. Muller, et al., TrackingNet: A largescale dataset and benchmark for object tracking in the wild, ECCV, 2018

Dataset Contruction

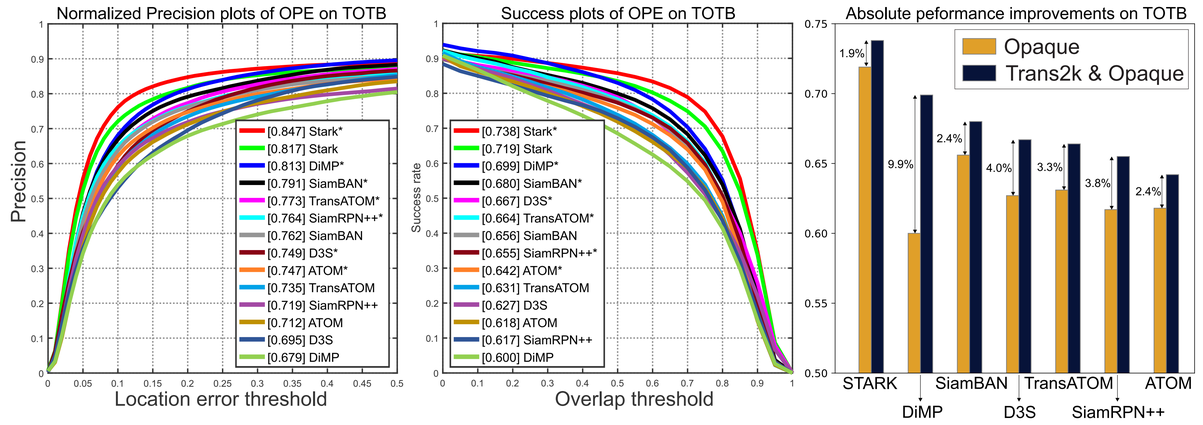
- Background simulated by a random video from GoT-10k tracking dataset [7]
- Render a transparent 3D object using BlenderProc [6]
- 25 object types, 148 3D models
- 2000 training sequences, >104k frames
- Target position annotation: **bounding box** + **segmentation**
- Dataset parameterization:



Transparency | Motion blur | Partial occlusion | Distractors | Target motion + scale | Rotation

Experimental Setup

- Re-trained state-of-the-art trackers
- Training data: Trans2k & Opaque (Opaque = GoT-10k [7], LaSoT [8], TrackingNet [9])
- Training batch sampled from Trans2k and Opaque with 5:3 ratio
- Evaluation dataset: TOTB (realistic video sequences)



Findings

- Up to 16% performance boost after re-training
- SOTA: transformer-based trackers (weakness: lack of discriminative power)
- Opaque objects important in training (otherwise “glass detector”)
- Deeper backbones outperform shallow ones (ResNet18 vs. ResNet50)
- Minimal performance loss on opaque objects

alan.lukezic@fri.uni-lj.si, https://github.com/trojerz/Trans2k