

Supplementary Materials : Depth-only Object Tracking

Song Yan¹

song.yan@tuni.fi

Jinyu Yang²³

jinyu.yang96@outlook.com

Aleš Leonardis³

a.leonardis@cs.bham.ac.uk

Joni-Kristian Kämäräinen¹

joni.kamarainen@tuni.fi

¹ Tampere University

Finland

² Southern University of Science and

Technology

China

³ University of Birmingham

United Kingdom

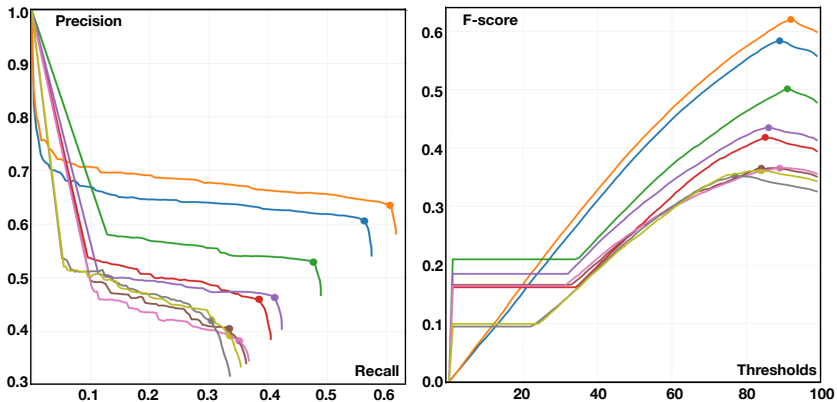
1 Performance analysis

Pr-Re plots. The overall performances, including Precision-Recall and F-score plots, are illustrated in Figure 1. Our proposed RGBD-DiMP trackers, using RGB+3×D inputs and using RGB+ColMap inputs, achieve the promising performances on both the DepthTrack-ST test set and CDTB-ST compared to the corresponding RGB DiMP trackers. The results verify that pseudo RGBD training datasets bring significant benefits for tracking performances.

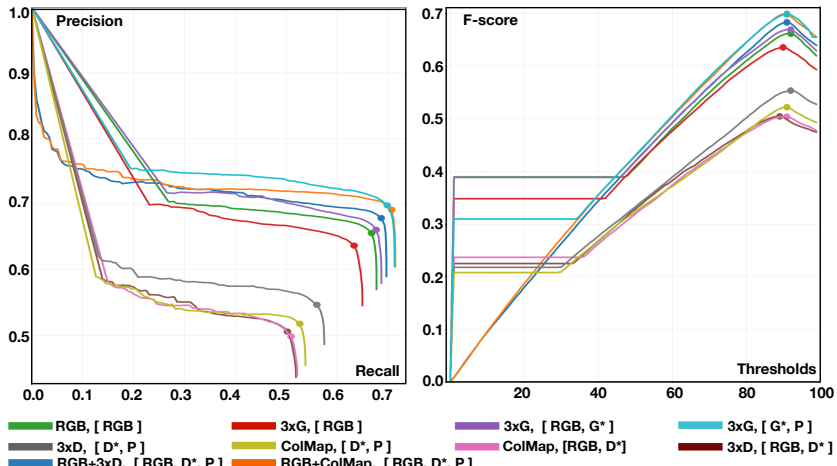
Visual attributes. The results are illustrated in Figure 2. Compared the corresponding RGB DiMP and Depth-DiMP trackers, our proposed RGBD-DiMP tracker (RGB+ColMap) achieves the best F-scores in 8 out of the 13 attribute categories in CDTB-ST and 14 out of the 15 attributes on the DepthTrack-ST test set. It is noteworthy that on the CDTB-ST our RGBD-DiMP tracker (RGB+ColMap) performs particularly poor on *out-of-frame* and *similar-objects* scenarios because the original DiMP is the short-term tracker.

2 Sequence level analysis

To investigate the complementary properties of RGB and D we computed the graph in Figure 5 where all DepthTrack-ST test sequences are sorted from left to right so that for the leftmost RGB is more dominant and for the rightmost D is more dominant. In the middle are sequences for which color and depth DiMP trackers perform equally well. For the major part of the sequences RGB is better than D (ratio < 1.0), but on the other hand, in 30% (37 out of 124) the depth is better than RGB (≥ 1.0) and in half of them depth is clearly better.



(a) Pr-Re and F-score plots on the DepthTrack-ST test set.



(b) Pr-Re and F-score plots on the CDTB-ST test set.

Figure 1: Precision-Recall curves (left) and F-score curves (right). For each tracker, $X, [Y]$ denotes the input data encoding (X) and the training or finetuning data ($[Y]$). RGB denotes the RGB data for training RGB DiMP tracker and ResNet50, *e.g.* LaSOT and Got10k, while $*$ denotes the data from the DepthTrack training set, and P denotes the pseudo RGBD training data, including G channels and the generated depth D^\dagger .

Example sequences. The 10 best and worst sequences in the DepthTrack-ST test set for the depth-only tracking (D^\dagger -CDTB- $3\times D$) are listed in Table 1 and Figure 3 shows three example sequences in which target shares same texture and color (the background clutter) and target is in the dark scenes.

Scenes for which depth-only fails are illustrated in Figure 4. Multiple distractors share the same depth as that of target but have different texture, *e.g.* yellow and green footballs. It is worth to note that trackers struggle to find the boundaries between hand-held targets and hands or arms and finally fails to wrong objects.

DepthTrack test set is designed to challenge both RGB and depth trackers and they may fail at the very beginning of sequences due to the fast motion of targets.

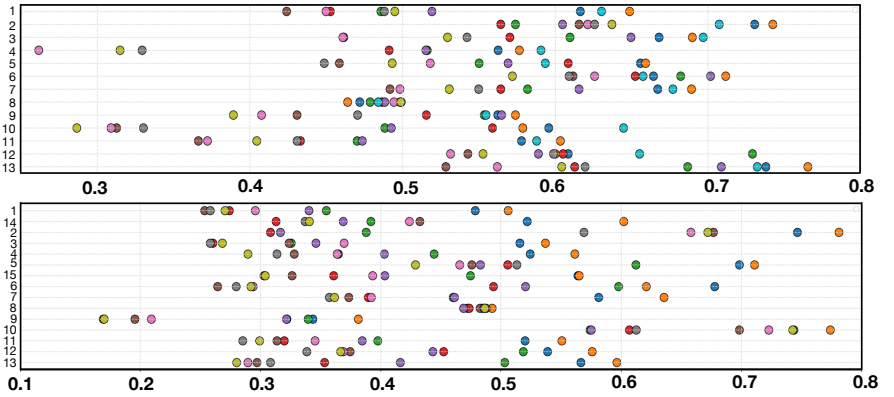


Figure 2: F-scores of trackers on each visual attributes for CDTB-ST (top) and the DepthTrack-ST test set (bottom). X-axis denotes the F-scores and Y-axis denotes the visual attributes, from top to bottom, 1) aspect change(AC), 2) dark scene (DS), 3) depth change (DC), 4) fast motion (FM), 5) full occlusion (FO), 6) deformable (ND), 7) out of plane (OP), 8) out of frame (OF), 9) partial occlusion (PO), 10) reflective target (RT), 11) size change (SC), 12) similar objects (SO), 13) unassigned (Nan), 14) background clutter (BC), 15) moving view (MV). Color markers denote the trackers (see the description in Figure 1).

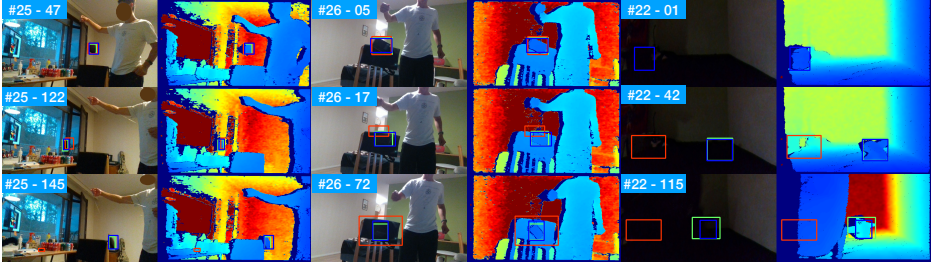


Figure 3: DepthTrack-ST test example sequences where the Depth-DiMP tracker (D^{\dagger} -CDTB-3×D) (blue boxes) is clearly superior to the RGB DiMP tracker (red boxes). Green boxes denote the groundtruth bounding boxes. # S - N denotes the N -th frame of the S -th sequence in Table 1.

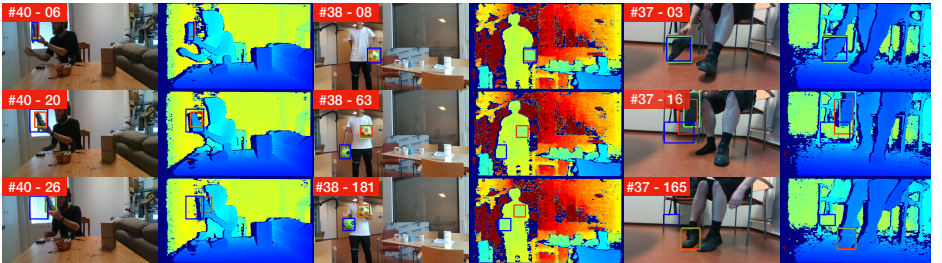


Figure 4: DepthTrack-ST test example sequences where the RGB DiMP tracker (red boxes) is clearly superior to the Depth-DiMP tracker (D^{\dagger} -CDTB-3×D) (blue boxes).

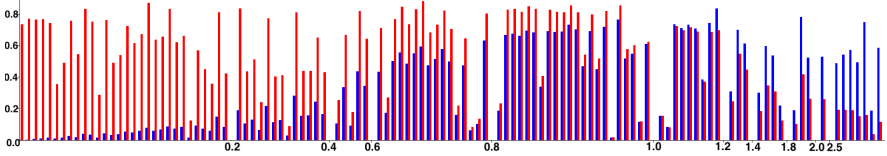


Figure 5: The plot of F-scores (Y-axis) and the ratios (X-axis) between the RGB DiMP tracker (red) and the Depth-DiMP tracker (D^{\dagger} -CDTB- $3\times D$)(blue) for each sequence in DepthTrack-ST test set.

<i>id</i>	<i>Seq. name</i>	Depth-DiMP			RGB DiMP		
		Pr	Re	F	Pr	Re	F
21	cube02_indoor_3	0.203	0.193	0.198	0.225	0.175	0.197
22	roller_indoor_3	0.839	0.817	0.828	0.033	0.009	0.014
23	pot_indoor_1	0.805	0.800	0.803	0.029	0.029	0.029
24	dumbbells01_indoor_2	0.641	0.556	0.595	0.340	0.352	0.346
25	bag02_indoor_3	0.520	0.473	0.495	0.034	0.033	0.033
26	notebook01_indoor_7	0.739	0.739	0.739	0.116	0.091	0.102
27	file01_indoor_1	0.490	0.493	0.492	0.151	0.151	0.151
28	toy02_indoor_1	0.594	0.499	0.542	0.198	0.188	0.193
29	bandlight_indoor_1	0.316	0.300	0.308	0.252	0.240	0.246
30	bottle04_indoor_1	0.534	0.536	0.535	0.305	0.310	0.308
31	adapter01_indoor_4	0.013	0.002	0.003	0.738	0.732	0.735
32	bag01_indoor_4	0.143	0.005	0.009	0.786	0.742	0.763
33	pigeon01_wild_1	0.011	0.011	0.011	0.767	0.767	0.767
34	cat01_indoor_3	0.016	0.016	0.016	0.746	0.740	0.743
35	ball10_wild_2	0.021	0.007	0.011	0.331	0.381	0.354
36	colacan03_indoor_5	0.678	0.014	0.027	0.833	0.690	0.755
37	shoes02_indoor_2	0.023	0.017	0.020	0.589	0.507	0.545
38	ball20_indoor_3	0.037	0.043	0.040	0.837	0.824	0.831
39	ball10_wild_1	0.091	0.093	0.092	0.737	0.465	0.570
40	developmentboard_indoor_2	0.033	0.036	0.034	0.699	0.376	0.489

Table 1: 10 best DepthTrack-ST test sequences for depth (D^{\dagger} -CDTB- $3\times D$) and color (RGB) trackers respectively.

3 RGBD tracking

We combined the optimized Depth-DiMP and RGB DiMP and evaluated on the CDTB-ST dataset. The results are in Table 2. Our proposed RGBD-DiMP trackers are pretrained on the generated RGBD datasets and finetuned on the DepthTrack training set.

Tracker	Pr	Re	F	Input
Ours	0.694	0.710	0.702	RGB+ColMap
Ours	0.681	0.690	0.685	RGB+ $3\times D$

Table 2: Evaluation results of our proposed RGBD-DiMP trackers on the CDTB-ST.